# Revisiting Ambivalent Sexism and the Ambivalent Sexism Inventory: Examining the Effects of Respondents' and Targets' Racial Group Membership on Endorsement

A Dissertation

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

BY

Molly K. Madzelan

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Eugene Borgida, Mark Snyder

June 2024

## Acknowledgements

It has been a long, difficult, even exhausting journey to finally reach the point of handing in my dissertation, and there are so many people without whom I never would have made it this far in my graduate career.

As they should know by now, I am extremely indebted and grateful to my advisors, Eugene Borgida and Mark Snyder, who have been excellent mentors and colleagues throughout my journey as a graduate student, helping me to not just hone my skills but encouraging me to move beyond my comfort zone and be more confident in myself and my work. The enormous amount of feedback provided on numerous dissertation drafts was incredibly helpful. Over the years, I have also been fortunate to work with Christopher Federico and Barbara Loken, and am tremendously thankful that they agreed to serve on my dissertation committee as well. I would also like to thank the rest of the Social Area faculty, who have been very supportive during my time at the University of Minnesota.

I also would not be here without the support of my friends, especially those I have met through this program. Thank you all so much – particularly my fellow Slackers – for listening to my numerous complaints, answering my questions no matter how ridiculous they were, and easing my anxieties about this entire process.

Finally, I thank my family – Mom, Dad, Annie, Kaitlin, Thom, and Kathy – for not only putting up with my decision to quit my job and go back to school but fully supporting it; talking me off the ledge when my anxiety would get the best of me; proofreading papers despite not having a clue what I was talking about; and being the

best cheerleaders I could ask for, to name just a few of the ways they have supported me since I arrived in Minneapolis. I could not have done any of this without you.

## Abstract

In recent years, social psychologists have increasingly acknowledged the importance of intersectionality in social-psychological research, especially in the domain of gender. In particular, the intersection between gender and race may be especially important to understanding gender prejudice, particularly ambivalent sexism and its two components, hostile sexism and benevolent sexism (Glick & Fiske, 1996). I argue that ambivalent sexism theory as well as its measure, the Ambivalent Sexism Inventory, are due for an intersectional re-examination to determine the role that race plays in endorsement and application of hostile and benevolent sexism. In Study 1, I investigated the influence of respondent race on ASI scores, finding that measurement invariance could not be established across the four race-by-gender groups of interest (i.e., Black men, Black women, White men, and White women) and therefore ASI scores could not be compared between these groups. This raises several important questions about whether or not the ASI is a valid measure of ambivalent sexism for Black Americans and if that construct properly reflects Black Americans' contemporary sexist attitudes. In Study 2, I investigated the influence of target race on ASI scores after first establishing measurement invariance for three of the four race-by-gender groups (the exception being Black Men). I found no evidence that target race affected hostile and benevolent sexism endorsement but did find differences in endorsement between Black women and White women, as well as between White men and White women. My research indicates that an intersectional perspective is critical to research using ambivalent sexism theory and the ASI, suggesting that, at the least, race must be considered in concert with gender when studying this particular form of sexism. Finally, and more generally, my research

highlights the need for social psychologists to better attend to proper measurement evaluation (including testing measurement invariance) to ensure that measures are functioning as intended across all groups of interest so as to avoid invalid or overly broad findings and conclusions.

# Table of Contents

# List of Tables

## Study 1 Tables

**Study 2 Tables**

# List of Figures

## Study 1 Figures



## Study 2 Figures

# Chapter 1

## Introduction and Overview of Studies

Although sexism, misogyny and the subordination of women are as old as human history itself, mainstream interest in these topics has re-emerged in recent years in the United States thanks to landmark events such as Hillary Clinton's Presidential campaign in 2016 and the rise of the #MeToo movement in the 2010s. Clinton's candidacy highlighted the sexism that even the most powerful, competent women continue to face in a country that is supposed to have true gender equality (Beinart, 2016). In a similar vein, the #MeToo movement has brought to light the pervasiveness of sexual abuse and sexual harassment of women and girls, empowering survivors to speak up about their experiences and seek justice (Langone, 2018).

This renewed attention has coincided with a cultural reckoning in which Americans have had to examine the role that systemic racism has played in this country's past, present and, potentially, its future ("America's Racial Reckoning," n.d.). Bringing these two prejudices together, particular attention is being paid to the unique concerns and experiences of women of color, covered by mainstream media for perhaps the first time. For example, news outlets such as *The Washington Post* featured articles about white feminism and the 2017 Women's March (Heaney, 2019) and in 2020 then-Senator (and current Vice President) Kamala Harris's identity as a woman of color was a central topic of discussion, with pundits analyzing the unique bias directed at her (similar to but also very different from the sexism faced by Clinton four years prior; Astor, 2020).

Although new to mainstream media and general public discourse, the idea that women of color experience a unique combination of racism and sexism has a long history

among feminist scholars, particularly Black feminists (McCall, 2005; Cole, 2009; Mays & Ghavami, 2018). This idea is at the heart of what has come to be known as *intersectionality*, a term coined by legal scholar and critical race theorist Kimberlé Crenshaw (1989). Today there is no single agreed-upon definition of intersectionality across the various social sciences – an issue in and of itself – but at its core, this concept emphasizes the fact that individuals belong to multiple social categories and that these categories have varying degrees of societal power (Else-Quest & Hyde, 2016a). Moreover, these identities are intertwined and function interdependently, meaning that their effects cannot be isolated. Thus, for example, this position would suggest that to more fully understand an individual you must account for their gender, race, age, sexual orientation, and so on, and you must do so simultaneously.

Despite being established by Crenshaw over thirty years ago and further developed in the ensuing decades (particularly by Black feminist scholars), intersectionality theory has been relatively neglected in mainstream psychology until relatively recently (Shields, 2008; Cole, 2009; Goff & Kahn, 2013; Rosenthal, 2016; McCormick-Huhn et al., 2019; Settles et al., 2020). More broadly, psychologists *are* getting better at acknowledging the biases in our field; for example, we have recognized that research tends to not only oversample from WEIRD societies (i.e., White, Educated, Industrialized, Rich and Democratic) but also overgeneralize results based on those samples (Henrich et al., 2010). Nevertheless, psychologists have been remiss at actually implementing research practices that examine psychological phenomena using an

intersectional framework (McCormick-Huhn et al., 2019; Settles et al., 2020).[1] The ratio of theory papers – or even just papers calling for more empirical work – to empirical papers is still too large. There are many explanations for why intersectional research has not taken a stronger hold in psychology. Perhaps most importantly, intersectionality is an extremely complex concept, which has serious implications for research that is based on it. Although that complexity is rich fodder for theorists, it may actually be off-putting for empiricists, as research that truly takes an intersectional lens becomes very convoluted very quickly (McCall, 2005). Additionally, such research, if done well, may take more time and resources than many researchers can afford. However, it is also important to keep in mind the inherently political origins of intersectionality theory (Cole, 2009), which can elicit backlash from those critical of feminism, critical race theory, and the acknowledgement of inequalities in modern society. But no matter the underlying causes, the result is that much of our understanding of the human mind and behavior remains incomplete, often reflecting the experiences of a select (and privileged) group of people (i.e., White men and women).[2]

And so, returning to the opening examples, one area of psychology that most certainly needs to center itself on intersectionality is the study of gender, particularly sexism and gender prejudice. Unfortunately, and perhaps surprisingly, research on sexism has been lacking in this regard and thus there is a need to re-examine our long-standing theories from an intersectional perspective, as many were developed without any

[1] There *are* some areas of psychology in which lines of research are intersectional in nature but not necessarily labelled as such. For example, the social cognition literature on multiple social categorizations fits neatly within intersectionality theory, but has only recently directly recognized this thinking (e.g., Kang & Bodenhausen, 2015; Nicolas et al., 2017).

[2] While the current paper is focused on intersectionality (or lack thereof) in psychology, similar issues play out in other fields, such as sociology, philosophy, and political science.

attention paid to the potential impact of social categories other than gender. Although categories such as age and sexual orientation, to name just two examples, are relevant to gender prejudice (e.g., Andreoletti et al., 2015; Cowie et al., 2019), there is reason to expect that race and/or ethnicity, especially for those in the U.S., is particularly relevant to gender (e.g., Bowleg, 2008; Goff & Kahn, 2013; Reid et al., 2014). Thus, the current research focuses on the intersection of these two social categories in the context of sexism. Furthermore, the current research homes in on one of the most influential theories of sexism, ambivalent sexism (Glick & Fiske, 1996), which seems especially in need of an intersectional re-examination.

In the remainder of this first chapter, I review the literature that informs the development of both Study 1 and Study 2. Both studies are grounded in an intersectional approach to ambivalent sexism theory, and so this literature review begins with this theory, as it currently stands, and then examines how an intersectional perspective changes how we might think about ambivalent sexism and its measure. The next section reviews ambivalent sexism theory and includes subsections on hostile sexism and benevolent sexism – the two dimensions of ambivalent sexism – as well as the role that subtyping plays in ambivalent sexism endorsement, the measure of ambivalent sexism, and claims that ambivalent sexism is a universal construct. Following this, intersectionality theory is reviewed and I argue that this theory has important implications for psychological research on gender. In the third section, I contend that an intersectional re-examination of ambivalent sexism suggests that race has an important role to play in both the endorsement and application of ambivalent sexism. Finally, the chapter ends with a brief overview of the two studies as well as an outline of the chapters to follow.

## Ambivalent Sexism

The early-to-mid-1990s saw renewed interest in studying sexism among psychologists (Glick & Fiske, 2011). Many researchers were looking to answer the same basic question: Why has sex discrimination persisted in the U.S. despite increasing endorsement of gender equality by the public? Researchers studying racism had been asking similar questions since the 1970s and during the 1990s some sexism researchers turned to modern racism theories (e.g., McConahay et al., 1981) for inspiration in understanding contemporary sexism.

Ambivalent sexism (AS) theory, in particular, arose directly from a comparison between racial intergroup relations and gender intergroup relations, specifically, the realization that they were different in critical ways (Glick & Fiske, 1996). Relations between men and women do not easily fit the mold of racial intergroup relations given the substantial interdependence that exists between the former and not the latter. While women may depend on men for protection, resources and the like, men depend on women for sexual reproduction, domestic help (e.g., child rearing) and, in many cases, psychological intimacy. Thus, although men typically wield structural power, women are not powerless; because men depend on them to meet certain needs, they wield dyadic power. It is this power that leads to men's conflicting (i.e., ambivalent) feelings towards women: They cherish women as wives and mothers, but resent the power that women hold and wish to control it themselves.

As such, it is easy to see how gender relations do not fit Gordon Allport's (1954) classic definition of prejudice as a uniform antipathy toward an outgroup, but instead are marked by men's deep ambivalence toward women. Thanks to sexual reproduction, men

and women are intimately connected in ways that no other groups are and thus reverence exists alongside antipathy. This dynamic would later be labelled the *central gender relations paradox*: "Men simultaneously seek dominance over but also intimate interdependence with women. Whereas ethnic, racial, national or religious groups may desire to live without or completely eliminate outgroups, heterosexual men 'can't live without' women" (Glick et al., 2018, p. 363).

As a result of this line of thinking, Glick and Fiske (1996) characterize sexism as multidimensional, including both outright hostility and apparent benevolence. However, despite their differences, both hostile sexism and benevolent sexism are rooted in the belief that women are inferior to men. Additionally, both forms of sexism are derived from three sources: paternalism, gender differentiation and heterosexuality.

### *Hostile Sexism*

The first dimension of AS is hostile sexism (HS), which consists of those aspects of sexism that align with Allport's classic definition of prejudice (Glick & Fiske, 1996). It is marked by explicit negative stereotypes and beliefs about women, and is what we typically think of when we think about sexism and gender prejudice. HS is overt and obvious; at least today, it is generally understood to be socially unacceptable by most people within the U.S. As an example, someone who believes that a woman is less intelligent than a man simply because she is a woman is a hostile sexist.

One way AS theory distinguishes itself from other contemporary theories of sexism is in its theoretical sources of antipathy. First, Glick and Fiske (1996) note that, in common discourse, 'sexism' and 'paternalism' are often used interchangeably; however, from the point of view of AS theory, paternalism is just one part of sexism. On the HS

side of AS there is *dominative paternalism*, which refers to the obviously negative belief that women are not fully competent adults – perhaps more akin to children – and need a male authority figure in order to function properly. This belief is a justification for patriarchy and male structural power.

Second, also underlying HS is *competitive gender differentiation*. This refers to the idea that only men naturally have the traits and abilities to effectively lead society, whereas women are fit only for subservient roles due to their inherent lack of such. As a result, men have a competitive drive to differentiate themselves from women, which leads them to boost their own self-esteem and image by further degrading and looking down on women. Competitive gender differentiation, like dominative paternalism, also sets the stage for patriarchy, as it casts women as naturally unfit to lead social institutions.

Finally, *heterosexual hostility* is a critical piece of HS. As previously noted, sexual reproduction has important implications for gender relations: It creates an unusual situation in which members of the dominant group (i.e., men) are inherently dependent on members of the subordinate group (i.e., women) for survival, which gives the latter unique dyadic power. Women are thus seen as gatekeepers of an incredibly important resource: sexual intercourse. As the theory goes, men are believed to resent this power – and the vulnerabilities it creates for them – and thus resent women for holding it, once again leading to a desire to dominate.

Taken together, dominative paternalism, competitive gender differentiation and heterosexual hostility set the stage for men to resent women and treat them as incompetent ninnies who cannot fend for themselves.

### *Benevolent Sexism*

The second prong of AS is benevolent sexism (BS; Glick & Fiske, 1996). BS is marked by seemingly positive beliefs and stereotypes about women – positive from the perspective of the attitude holder, but not necessarily from that of the target. These are nevertheless harmful to women because they are still rooted in ideas about female inferiority and male dominance. For example, a benevolent sexist might believe that all women are nurturing and warm, a positive belief; but this belief can restrict women to caretaker roles and limit their agency. Similarly, a man repeatedly offering unsolicited help to his female coworkers may truly believe he is just being nice, but it may also reflect an (unconscious) assumption that he is more competent than they are, especially if he fails to offer the same help to his male coworkers. Underneath the positive veneer, BS is centered on the idea that women are weak and in need of protection, and in placing women on a pedestal, it restricts them in harmful ways. As such, in reality BS helps reinforce the gender status quo (Barreto & Ellemers, 2005).

Unlike HS, which is explicit and obvious, BS is typically conceptualized as more subtle because of its positive facade. It is often not perceived as sexist, even among women (Dardenne et al., 2007). Even in situations in which women feel such attitudes and behaviors as sexist, they are often reluctant to call it out because the majority of others – both men *and* women – perceive BS as acceptable (Glick & Fiske, 2001; Barreto & Ellemers, 2005).

Just as with HS, Glick and Fiske (1996) argue for three sources of BS. First, instead of dominative paternalism, BS derives from *protective paternalism*. Women are still seen as weak and incompetent. However, because of their dependence on women as

wives, mothers and romantic partners, men may feel the need to protect women rather than dominate them, per se. From this view, women are to be cherished, protected, and provided for. This belief can still involve treating women as no more capable than children, but such treatment is thought to come from a place of love rather than contempt. In any case, the end result is still a justification for patriarchy.

Similarly, *complementary gender differentiation* justifies patriarchy just as competitive gender differentiation does, but in a nicer 'package'. It is a different approach to the same social division: male dominance and female submission. Once again, men are the ones who have the necessary traits and skills to lead, but instead of a competitive drive to differentiate, men's dependence on women leads to viewing their traits as complementary, necessary for society to properly function. But, in the end, women's traits and abilities still make them best suited for subordinate roles, leading to the traditional division of labor between men and women (i.e., men work outside the home and women within). Thus, women are still restricted in what they can do, but it is not necessarily framed as degrading; instead, women are in their proper place, complementing men and making up for what men lack (e.g., a nurturing nature necessary for child rearing).

Finally, like heterosexual hostility, *heterosexual intimacy* also comes directly from women's dyadic power. Although our most basic, fundamental motive is to reproduce and pass on our genes, humans also have a strong, genuine desire for psychological closeness with others. Especially in more contemporary times, men typically do not seek this from other men (as women do from other women, even outside of romantic relationships; Fehr, 1996); instead they look to women for this emotional

intimacy. Once again, women must be cherished and protected in order for that intimacy to be achieved, and achieving that intimacy has become a fundamental need.

Together, protective paternalism, complementary gender differentiation and heterosexual intimacy reflect how women's dyadic power creates a unique intergroup dynamic in which men maintain dominance over women, but do so out of a need to protect the subordinate group.

Because of its links to seemingly positive motives, beliefs and feelings, it can be difficult to see BS as a prejudice in the same way as HS. As previously noted, HS fits neatly into Allport's traditional definition of prejudice while BS does not. However, Allport (as discussed in Glick & Fiske, 2001) also argued that, in the end, the effect of prejudice is to place the target of that prejudice at a disadvantage; and such a disadvantage is then often perpetuated by social structural forces. This is what makes BS a prejudice, same as HS: Although BS has a positive front, its downstream effects lead to numerous disadvantages for women (see for example Dardenne et al., 2007; Becker & Wright, 2011; King et al., 2012), which contribute to the maintenance of the gender hierarchy openly desired by hostile sexists.

### *Ambivalent Sexism and Subtyping*

The last central component of AS theory is subtyping. A key finding of the original AS research – which has been frequently replicated over the years – is that, among men, HS and BS are often moderately correlated (*r*s = .40-.50; Glick & Fiske, 1996, 1997, 2011). This suggests that men often hold both attitudes simultaneously, despite the fact that they differ in valence toward the same attitude object (i.e., women). Traditionally, within psychological thinking, this kind of attitudinal ambivalence is

associated with confusion, conflict or tension (Glick et al., 1997). Per Katz (1981, 2014), people often develop strategies to cope with ambivalent attitudes so as to avoid such negative outcomes. Interestingly, however, ambivalent sexists do not seem to have such problems. How do they achieve this?

Research has revealed that men – especially sexist men – tend to stereotype women at the subtype or subgroup level, rather than at the more general level of 'women' (Glick et al., 1997). This focus on subgroups, rather than a monolithic whole, is what allows HS and BS to function complementarily rather than be mutually exclusive. Specifically, BS is directed toward women who reinforce conventional gender relations and chiefly occupy roles such as wives, mothers and romantic partners. Such women adhere to traditional gender roles and norms, fitting the traditional stereotype of women as communal, dependent, and warm (Broverman et al., 1972; Ellemers, 2018; Sczesny et al., 2018). Thus, a BS-response is most likely to be triggered by role-related cues (Bareket & Fiske, 2023). In contrast, HS is elicited by women who are viewed as directly challenging or trying to usurp men's power; in other words, women who are trying to disrupt the gender status quo rather than adhere to it. Such women defy traditional gender roles and norms, more closely fitting the traditional stereotype of men as agentic, competent, dominant, and assertive (Broverman et al., 1972; Ellemers, 2018; Sczesny et al., 2018). HS often targets feminists, career women and seductive women (or, women who control their own sexuality) because HS responses are most likely to be triggered by sexuality or power cues (Bareket & Fiske, 2023).

Therefore, another critical piece of the AS puzzle is the distinction between traditional and nontraditional women, or, more generally, between traditional and

nontraditional ideas about gender (which can also include roles, beliefs and stereotypes about men). Glick and Fiske have acknowledged that ambivalently sexist men may have difficulty when they encounter a woman who is difficult to categorize (e.g., a stay-at-home mother who also identifies as a feminist; Glick et al., 1997), but how men cope with this kind of situation has not been studied.

Because the two can act in concert with one another rather than produce conflict, HS and BS work together to promote gender inequality across multiple content domains (Bareket & Fiske, 2023) – they are "two sides of a sexist coin" (Glick & Fiske, 2001, p. 532). Based on prior work by Jackson (1994), BS and HS have been conceptualized as the 'carrot' and the 'stick," respectively, that keep women in line, forcing them into the traditional subtype by rewarding those who adhere to it (i.e., by treating them with BS) and punishing those who do not (i.e., by treating them with HS; Connor et al., 2016).

### *The Ambivalent Sexism Inventory*

In addition to laying out their theory, Glick and Fiske (1996) also developed and validated a scale to measure individuals' endorsement of both HS and BS. The Ambivalent Sexism Inventory (ASI) consists of 22 items, 11 each for HS and BS. Participants read a statement and indicate their agreement with that statement using a 6-point response scale, ranging from 1 (*disagree strongly*) to 6 (*agree strongly*) with no neutral response option (e.g., *neither agree nor disagree*). An example of a HS item is "Most women interpret innocent remarks or acts as being sexist" whereas an example of a BS item is "Women, compared to men, tend to have a superior moral sensibility." (See Figure 1 for the measurement model and the Appendix for the full list of items.)

Six items are reverse scored as appropriate and then averaged to form an HS score and a BS score. All 22 items can be averaged to form a single overall measure of sexism, but the measure is typically not used in this way because of the important theoretical differences between HS and BS. Initially, items were designed to tap the three sources (paternalism, gender differentiation and heterosexuality) for both HS and BS, but factor analyses determined that the three subfactors existed for BS but not for HS. Glick and Fiske (1997) later argued that although they believed the three HS domains do exist, they may be "too inextricably bound together to distinguish empirically" (p. 125).

The measure was created and validated across six samples (Glick & Fiske, 1996) and has essentially remained unchanged since its introduction.[3] Most modifications to the original scale seem to involve using a subset of items rather than the full scale. It is important to note that the six original samples included four undergraduate samples and two small community samples, all of which were collected in the Midwestern and Northeastern U.S. Demographic information was not collected for all samples, but Glick and Fiske reported that all samples were majority White, ranging from about 70-90% per sample. Thus, although there was some diversity within these samples regarding certain demographic features such as age, there was very little diversity regarding race and/or ethnicity. Although such sample demographics are not at all uncommon in psychological research, especially at the time that this research was conducted, this raises important

[3] The ASI *has* been modified in individual studies, usually by removing items to make it shorter and/or rewriting the reverse-scored items to be straightforward – some studies have even done both (e.g., Glick & Whitehead, 2010). Since its introduction, no new items have been added to the ASI. The implications of such modifications are further addressed in Chapter 4.

questions about the validity of the ASI across different racial and ethnic groups (Sue, 1999; McCormick-Huhn et al., 2019).[4]

Importantly, Glick and Fiske (1996) did perform subgroup analyses to determine if the ASI was a valid measure of sexism for both men and women. Although AS theory was explicitly developed from the male motivational perspective, the ASI factor structure held across the two groups, suggesting that it is a valid measure of sexism for women as well as men, which has been supported by later research (e.g., Fields et al., 2010). Glick and Fiske suggested that most women come to adopt or are at least aware of (if unconsciously) men's sexist beliefs about women, which are culturally transmitted to them. Thus, AS is not solely a matter for men. However, interesting endorsement patterns do emerge for each group, patterns that have been replicated rather consistently over the last two decades. Specifically, while men typically endorse HS to a much greater extent than women, the difference in BS endorsement between the groups is much smaller (but with men still ahead). These findings suggest that women do not resist BS as strongly as they do HS (Glick & Fiske, 1996), which squares with research demonstrating that BS is often not recognized as sexist by women (Dardenne et al., 2007) and therefore is more difficult to challenge in everyday life. Moreover, although HS and BS are associated with opposing evaluations of women (negative vs. positive) for men, for women both HS and BS are related to negative evaluations of women (Glick & Fiske, 1997). Rather than fostering benevolent attitudes toward other women, BS may lead women to view

[4] Other popular psychological measures have been questioned for similar reasons. For example, Davis et al. (2016) cited the predominantly White sample used to validate the Moral Foundations Questionnaire as one reason for investigating for Black participants.

traditional women as competitors (for men's attention) and nontraditional women as deviant – an intriguing theory, but one that has not received much empirical follow-up.

***Universality***

Despite some potential concerns regarding the ASI's validity, AS has essentially been treated as a universal construct and the ASI has been administered cross-culturally for over twenty years (Bareket & Fiske, 2023). Glick and Fiske have been careful not to explicitly claim universality (e.g., Glick et al., 2000), but they do argue that HS and BS "originate in social and biological factors common among human groups" (p. 763), the biological factors being human reproduction and the social factors being division of labor (and the consequences thereof). Even these social factors stem from a biological source, namely sexual dimorphism, which is at the heart of many evolutionary explanations for gender differences in attitudes and behaviors (e.g., parental investment theory as an explanation for gender differences in mate selection; Trivers, 1972; Buss & Schmitt, 1993). Furthermore, evidence that ambivalence towards women is not 'new' and is in fact a longstanding part of human history also suggests that AS may be common to all groups just as it has been common across time periods (Glick et al., 1997).

But what does the evidence actually say about universality? Most researchers point to a 2000 study by Glick and colleagues that examined HS and BS in 19 countries around the world, including nations such as Australia, Brazil, Germany, Japan and Nigeria. Results suggested that "HS and BS are pervasive across cultures" (p. 763), as well as "recognizable, coherent ideologies in a variety of nations" (p. 772). The authors did say it was premature to claim HS and BS are human universals – arguing that replications are needed, as well as more work in additional countries – but nevertheless

pointed to the "consistent replication of the complex factor structure of the ASI" (p. 772) across nationally diverse samples as promising evidence. However, there is more to validating a measure across groups than just confirming its factor structure; but, unfortunately, in psychological science, this is often the only aspect examined when cross-cultural validity is checked (Hussey & Hughes, 2020) and additional steps have not been taken by Glick and colleagues. Additionally, with cross-cultural research, care needs to be taken to ensure that translations are culturally accurate, as direct translations may not hold up because the same word can have different connotations in different cultures (Davidov et al., 2014). For example, the Positive and Negative Aspect Schedule (PANAS; Watson et al., 1988) includes shame and guilt as two indicators of negative affect. However, although these are generally considered negative emotions in individualistic cultures, they are seen somewhat positively in collectivistic cultures (Lee, 2018); thus, simply translating 'guilt' and 'shame' from English into Japanese, for example, fails to capture critical differences in how those concepts are construed. Glick and colleagues acknowledged that culturally specific measures of HS and BS might be necessary – suggesting the possibility of different construals of the same items – but felt that keeping the measures constant was more desirable in this case, an argument that perhaps made more sense at the time compared to today. Furthermore, the samples used in this study were once again limited and non-representative (e.g., many samples consisted of undergraduate students), which raises questions about generalizability within each of the countries studied. Glick et al. acknowledged this limitation in the paper, but did not give it much attention at the time. A more recent systematic review of AS

research more clearly addresses the limitations presented by participant samples, pushing researchers to prioritize diversifying their samples (Bareket & Fiske, 2023).

Therefore, although some evidence that the ASI is a valid measure across different cultures/nations, claims of universality – explicit or implied – may be exaggerated at this time. But even if we were to take Glick et al.'s (2000) work at face value as evidence for the cross-cultural universality of AS and the ASI, the present project points to intersectionality theory as suggesting that AS theory and its measure, among different groups within the United States, have not been rigorously tested by psychometric standards associated with measurement invariance (an argument also made by Davis et al., 2022). As a result, it is possible that AS theory and the ASI may be more limited to certain groups than previously recognized. It is critical to take a step back and determine whether the measure is valid for different groups *within* a culture before jumping to answer questions about its validity for different groups *between* cultures. To be clear, this is an analytic point that has not been pursued empirically in any rigorous way in the research literature on AS and the ASI.

**Intersectionality**

As previously mentioned, despite being introduced by Crenshaw (1989) over thirty years ago, there is still no agreed upon definition of *intersectionality*, even within just the field of psychology. However, most scholars agree on several defining aspects, as outlined by Else-Quest and Hyde (2016a). First, intersectionality recognizes that all people are characterized by multiple social categories and that these categories are interdependent. Second, embedded within each of these categories is an element of power or inequality: some categories are more privileged or marginalized than others. For

example, a White woman's race grants her privilege but her gender marginalizes her. And third, these categories are not static but fluid and dynamic because they reflect not only properties of the individual but also properties of the social context that the individual inhabits. For example, the social context can change over time, resulting in changes in the power/inequality associated with a certain social category (e.g., lesbians in the U.S. are still marginalized due to their sexual orientation, but have more guaranteed rights now than they did fifty years ago). Similarly, the importance of certain categories for an individual can change. For example, a Black woman's race can be most salient in one context but her gender most salient in another.

In sum, intersectionality recognizes that individuals simultaneously belong to multiple social categories that afford them different advantages and disadvantages, and these categories and their significance can change over time. Therefore, to fully understand an individual, all of their social identities must be taken into account. For example, to best understand Jane, a Black lesbian, we cannot just think about her gender identity or her ethnicity or her sexual orientation, and so on. None of these identities exists in a vacuum and they are all interacting with one another to create Jane's unique experience. To be fair, this approach does create methodological difficulties and challenges – how can a researcher account for so many variables at once? And unquestionably this is one reason why intersectional research has been lacking in psychology (McCall, 2005). Moreover, situational cues can make certain categories or identities more salient than others in a given moment. In a measurement context, certain categories might unintentionally be made more salient than others, affecting results in

unanticipated and perhaps unacknowledged ways – adding to an already lengthy list of methodological concerns.

Furthermore, as argued by Crenshaw (1989) and others (e.g., Bowleg, 2008), intersectionality suggests that intersections of identities are more than just the 'sum' of those identities – in other words, intersectionality is not additive. A Black woman's experience, for example, is not just 'the woman's experience' plus 'the Black experience' – it is a unique combination of the two (Crenshaw, 1989). This does not mean that her experience is in no way similar to White women's or Black men's, but it does mean that there are important differences. For example, Black women are more likely to experience 'invisibility' as a unique form of discrimination: Because they are non-prototypical in terms of both their gender and their race, Black women are often unnoticed or unheard (Sesko & Biernat, 2009). Historically, these kinds of differences have been overlooked, in both the legal system (Crenshaw, 1989) and in psychology (Bowleg, 2008).

Although typically referred to as a theory, most scholars view intersectionality as more akin to an approach or framework (Hancock, 2007; Shields, 2008; Cole, 2009), or as a critical theory rather than a falsifiable one (Else-Quest & Hyde, 2016a). Some feminist scholars have argued that essentially all psychological research should have a built-in intersectional perspective because research that does not tends to result in an incomplete understanding of psychological phenomena, in which group-specific findings are generalized to all (Cole, 2009; McCormick-Huhn et al., 2019). This conclusion has important implications for all psychological research, but may be especially critical to our understanding of gender and gender-related phenomena.

### *Implications*

Regarding gender, an intersectional framework argues that there is no universal woman or universal man. This is because gender does not exist in a vacuum and is instead shaped by other social categories, such as race/ethnicity and sexual orientation (Shields, 2008). We can easily see this in how different cultures have different gender norms and expectations as well as in how norms and expectations within the same culture have changed over time (Reid et al., 2014). Research has also shown that individuals with multiple marginalized identities often do not view themselves in terms of single identities but rather view themselves as a combined or linked whole. For example, Settles (2006) found that Black women place greater importance on a 'black-woman identity' than on the individual identities of 'black person' or 'woman.' Thus, generalizations such as "men do A whereas women do B" are inappropriate and unwarranted from an intersectional perspective, not only because people simply do not behave so uniformly but also because many people do not think about themselves this way. This is especially true in a multicultural society such as in the U.S.

Intersectional research supports this cautious approach to gendered generalizations, suggesting that gender differences are often more nuanced than they initially appear. For example, research has typically supported the stereotype that girls and women have lower self-esteem than boys and men, although meta-analytic effect sizes tend to be small (Kling et al., 1999). However, an intersectional approach to the analysis sheds a new light on this finding: In the same meta-analysis, Kling and colleagues (1999) report an effect size of $d = 0.20$ for White samples and $d = -0.04$ for Black samples. This suggests that the 'well-established' gender difference in self-esteem may only apply to White Americans and that the finding of no gender differences in self-

esteem for Black Americans may be obscured when males and females were examined as monolithic groups. Similarly, research on gender stereotypes of emotion reveal different stereotypes among different ethnic groups (Durik et al., 2006). Durik and colleagues (2006) found that European Americans' emotion stereotypes are more gender differentiated than those of African Americans, Asian Americans, and Hispanic Americans. For example, while European Americans in the study stereotyped men as more likely to express pride than women, there was no such gender stereotype among African Americans, who rated men and women equally as likely to express pride. Intersectional research, such as the studies highlighted here, indicates that we may need to seriously re-consider how we study gender, gender differences, and related constructs/theories, as a purely gender-focused lens may not be adequate to fully understanding the phenomenon of interest.

**An Intersectional Re-Examination of Ambivalent Sexism**

AS theory and the ASI were developed at a time when there was less mainstream awareness (in psychology, at least) of intersectionality theory, and thus it is not surprising that the original theory and measurement approach are deficient from an intersectional perspective. Now that we have a more thorough understanding of intersectionality and its application to and implications for gender, as well as better psychometric tools and methods (discussed in greater depth in a subsequent section), a re-examination of AS theory and the ASI is needed. In particular, intersectionality theory, along with certain aspects of AS theory and measurement development, suggest that AS needs to be re-examined to consider how race might affect both endorsement and application (i.e., treatment of a target) of HS and BS. Although other social categories may also be

important to consider – and intersectionality theory indicates that they *should* be explored – race is inextricably linked to gender, especially for women (Shields, 2008; Reid et al., 2014), and may be particularly important in the context of AS, as will be reviewed. Factors such as age (Hammond et al., 2014), religion (Glick et al., 2002; Burn & Busso, 2005), education (Glick et al., 2002) and conservatism (Christopher & Mull, 2006) have all been studied with regard to how they affect AS endorsement, but race has rarely been examined in such a way, often because non-White sample sizes published in the research literature are too small for analysis (e.g., Christopher & Mull, 2006). Just one pair of researchers appears to have directly examined such differences in AS endorsement (Hayes & Swim, 2013), leaving a significant gap in the literature.

### *Re-Thinking the Origins of Ambivalent Sexism*

With this overlooked piece in mind, recall that AS is theorized to originate in "social and biological factors common to human groups" (Glick et al., 2000, p. 763). Specifically, the critical biological factor is the motivation to reproduce and pass along one's genes; women are required for sexual reproduction and thus have dyadic power due to men's dependence on them for this process (Glick & Fiske, 1996, 1997). This aspect of gender relations has essentially been static since humans evolved: the two groups are inherently interdependent and must rely on one another to survive in a purely biological sense. Biological factors such as sexual dimorphism (e.g., on average men are larger and physically stronger than women) have also led to common social factors, such as the gendered division of labor. Although the biological factors have been fundamentally

stable since humans evolved,[5] the social factors have not been as stable or consistent as AS theory implies. Moreover, changes in social factors have not been uniform across groups, even within the same culture or society.

Within the United States, one variable that has strongly influenced the social side of gender – and thus the social factors that underlie AS – is race[6] (Reid et al., 2014). Specifically, cultural histories of racism in the U.S. may significantly influence how men and women of different racial groups view the traditional vs. nontraditional dichotomy that is at the heart of AS theory. As will be explained, it is not clear if this dichotomy is universally defined or if it is defined based on Whites' experiences alone. In Glick et al.'s (1997) studies establishing men's subtyping tendencies and the traditional/nontraditional dichotomy, the majority of participants were White college undergraduates, which could mean that results are not generalizable to other Americans. A similar problem underlies development of AS theory and the ASI. Regarding BS in particular, Glick and Fiske (1996, 2011) admit that it has a very Victorian feel, which suggests that it might reflect White, Eurocentric ideas about gender relations rather than universal ones (Shields, 2007).

Similarly, AS theory and the ASI are heavily grounded in stereotypes about women and men. But an intersectional approach again raises questions about whether

[5] This view does not account for sexual orientation, specifically non-heterosexual orientations. Glick and Fiske (1996) acknowledge that AS theory is heteronormative, but very little research has looked at AS within different sexual minorities.

[6] Although the terms are often used interchangeably, race is distinct from ethnicity. As defined by the American Psychological Association (2022), race is "a socially defined concept sometimes used to designate a portion, or 'subdivision,' of the human population with common physical characteristics, ancestry, or language"; in contrast, ethnicity refers to "social categorization based on an individual's membership in or identification with a particular cultural or ethnic group." Based on these definitions, "Black" is a race while "Caribbean" is an ethnicity, for example. The present research primarily focuses on race, but that is not to say that ethnicity is not important in the context of AS.

such stereotypes are truly about *all* women and men, or just certain kinds of women and men (i.e., White men and women). At least one study indicates that the latter is the truth: Ghavami and Peplau (2013) – the first, and still (to my knowledge) only, study to directly and systematically study gender-by-ethnicity stereotypes beyond the typical Black/White binary and from an explicitly intersectional perspective – found that stereotypes for men and women of different racial and ethnic groups *do* differ in important ways and that stereotypes for the general groups 'men' and 'women' are much closer to stereotypes of 'White men' and 'White women,' respectively, than they are to any other gender-by-ethnicity group.[7] Similarly, earlier research by Durik and colleagues (2006) demonstrated that different ethnic groups have different gendered stereotypes about emotion, again suggesting that when considering men and women as groups, individuals can vary in how they conceptualize them. Taking these findings into account raises critical questions: How applicable are AS theory and the ASI to non-White men and women? Does the theory and its measure only accurately reflect the gender experiences and sexism of White men and women?

In sum, there is reason to believe that AS theory and the ASI may be biased toward White Americans' thinking about gender, including norms, roles and expectations. Thus, intersectional research that investigates the role that race plays in AS is critical to this area of research, given that this is such a popular, widely applied theory of sexism. It is critical to know if the theory and its measure are applicable for all racial and ethnic groups. Even now, over twenty-five years removed from their original paper,

[7] The close link between 'whiteness' and 'femininity' is further discussed in Chapter 3.

Glick and Fiske themselves have acknowledged the need for more intersectional research concerning AS (Connor et al., 2016; Glick & Raberg, 2018; Bareket & Fiske, 2023).

### *Overview of Studies*

The current research attempts to answer two key questions about AS and race, approaching the issue from two angles. The first question concerns personal endorsement, or the 'respondent side':

> (1) Do individuals of different racial groups – specifically Black and White Americans – demonstrate different levels or patterns of AS endorsement?

The second question concerns the targets of HS and BS (the 'target side'):

> (2) Does endorsement differ depending on the race of the women (and men) respondents are thinking about when responding to the ASI?

By looking at AS theory and the ASI from these two angles, a more comprehensive examination of its theoretical underpinnings and validity can be undertaken.

The current research consists of two complementary studies, each addressing one of the two focal research questions. The project takes a multi-method approach, using the first of a three-wave a panel study to examine the 'respondent' question and an experiment to examine the 'target' question. As target demographics can be manipulated, the latter examines the causal effects of race on HS and BS endorsement whereas the former is primarily a measurement study. Although the two are independent, Study 2 builds on Study 1, providing additional but different support for some of the hypotheses examined in Study 1.

In the next two chapters of this paper, I report the method and results of Study 1 (Chapter 2) and Study 2 (Chapter 3). Both chapters also include a review of the literature

that informed each study's hypotheses. Additionally, Chapter 2 includes a review of measurement invariance – an important precursor to testing group differences on any measure and the distinctive contribution of the present research – and the process of testing it. Finally, in the General Discussion (Chapter 4), I examine the current state of ambivalent sexism theory and its measure by discussing in detail the results, implications and limitations of each study as well as directions for future research.

# Chapter 2

## Study 1: The Role of Respondent Race in Endorsement of HS and BS

Study 1 addressed the first research question, investigating how individuals' racial group membership affects their endorsement of HS and BS on the ASI. This study focused specifically on potential differences in endorsement between Black and White Americans based on their responses to the ASI. However, before group comparisons could be analyzed, the *measurement invariance* (MI) of the ASI across the four relevant groups (i.e., Black men, Black women, White men and White women) was examined. MI "assesses the (psychometric) equivalence of a construct across groups or measurement occasions and demonstrates that a construct has the same meaning to those groups or across repeated measurements" (Putnick & Bornstein, 2016, p. 72). MI must be established before comparing group means: If the construct or measure is not invariant, i.e., it does *not* have the same meaning across groups, then group comparisons are meaningless. Therefore, the first part of Study 1 concerns establishing the MI of the ASI across the four gender-by-race groups of interest.

### Measurement Invariance

The importance of measurement invariance has been known for more than 50 years (Putnick & Bornstein, 2016), with its significance for the validity of multi-group comparisons laid out in the early 1990s (Van De Schoot et al., 2015). However, it was not until the early 21st century, and the advent of more accessible statistical techniques (e.g., the structural equation modeling approach), that the importance of MI – and a better understanding of the consequences of noninvariance – started to take a strong hold in mainstream psychology (Van De Schoot et al., 2015; Putnick & Bornstein, 2016). At the

time Glick and Fiske (1996) were developing their theory and measure, MI was known but not given as much attention as needed.

Even today, although it is recommended that MI should be checked before conducting any group comparisons on any measure – some measurement experts would say MI *must* be checked – this still rarely happens (or, at the least, is usually not reported on when it *is* done; Putnick & Bornstein, 2016; Hussey & Hughes, 2020). Instead, the measure is often assumed to be invariant across the groups of interest. But no matter how reasonable such assumptions may seem, they can be unfounded and may lead to erroneous conclusions if left unchecked. For example, Perez and Hetherington (2014) found that an extremely common measure of authoritarianism, Stenner's (2005) four-item child rearing scale, was *not* a valid measure of that construct for Black Americans but was for White Americans. Because the cross-racial validity of this measure was assumed and not empirically tested for over ten years, an important phenomenon was grossly misunderstood for a particular group of people: Based on this measure, Black Americans could be considered more authoritarian than White Americans, despite other measures of authoritarianism suggesting the reverse pattern. Such a conclusion, while seemingly sound based on the statistical analysis, also did not align with our general understanding of authoritarianism itself. Perez and Hetherington's study has been an important wake-up call for researchers and inspired others to think more carefully about the invariance of popular psychological measures (e.g., Davis et al. [2016] examined whether the Moral Foundations Questionnaire is a valid measure for Black Americans).

Measurement noninvariance indicates that the construct or measure in question has a different structure or meaning for different groups, making comparisons between

the groups meaningless and invalid (Putnick & Bornstein, 2016). For example, MI may lead to group means that are artificially inflated or deflated, which, when compared, indicate group differences that are not actually true (Chen, 2008). Feminist and intersectional scholars (e.g., Else-Quest & Hyde, 2016b; McCormick-Huhn et al., 2019) have long argued that the psychometrics of psychological measures should be validated for non-White groups (and other non-majority groups) for precisely this reason. Too many group comparisons have been made between Whites and racial minorities based on measures that may have a different cultural meaning for the latter groups, resulting in conclusions that may actually be incorrect or lacking important context. Therefore, it is critical to ensure that valid comparisons can be made before moving on to testing the actual research question. Moreover, conclusions that unknowingly rely on noninvariant measures or constructs can have serious consequences for how we approach certain issues. For example, Putnick and Bornstein, in their 2016 review of MI, discuss the possibility that a measure that includes weight gain and frequency of crying as indicators of depression may underestimate the prevalence of depression among men because, while these symptoms may be indicators of depression in women, they typically are not in men. Such a problematic measure could lead to missed diagnoses among men, who then fail to receive the proper treatment they need.

As will be reviewed in the next section, evidence from a number of different disciplines (e.g., history, sociology, women's studies) suggests that we might expect different patterns of HS and BS endorsement among Black and White men and women. This is assuming that MI holds for all four of these groups, but that is not necessarily expected to be the case. It is possible that the justifications given for the current study's

hypotheses (reviewed in the next section) could instead disrupt the entire structure and meaning of the AS construct for Black women and/or men. Returning to Perez and Hetherington's (2014) work, they found that the essential metaphor underlying the child rearing scale of authoritarianism – namely, if a person favors conformity in a child, they favor conformity in social subordinates – breaks down for Black Americans due to their experiences with racism from authority figures. Black Americans encourage conformity in their children not because they themselves are authoritarians, but because they know that as racial minorities, their children need to obey (White) authority figures to survive in a society in which they are part of a subordinate group whose rule violations may be more severely punished. Similarly, the traditional-nontraditional dichotomy that underlies much of AS may not 'hold' for Black Americans as it does for White Americans. This could simply result in different endorsement patterns as measured by the ASI, but it could result in measurement noninvariance. MI cannot be assumed and therefore must be tested before other analyses can be considered. In the following section, the substantive rationale for examining the MI of the ASI in the context of race is laid out.

**Race, Gender, and Ambivalent Sexism**

As previously noted, research in numerous disciplines indicates that race has been a key factor in the construction of gender, including norms and stereotypes. For Black Americans, the influence of race on gender can be traced as far back as the 17$^{th}$ century, specifically to slavery (Johnson & Loscocco, 2015; Abrams et al., 2016). Although there *was* gendered labor division during this time, enslaved Black men and women often worked together (e.g., harvesting cotton in plantation fields) and, ever since, Black women have been expected to work *and* take care of their families (Settles et

al., 2008). This expectation is at least in part due to systemic racism (e.g., discriminatory hiring practices, disparities in incarceration rates): Since the aftermath of the Civil War (and continuing through today), Black men have had difficulty fulfilling the traditional (American) male role of provider because of limited economic opportunities (Johnson & Loscocco, 2015). Thus, Black women have had to take on additional, traditionally masculine roles – i.e., breadwinner and head of household in addition to wife and mother – for far longer than White women. As such, Black women have been less likely than White women to live in households headed by a married couple and are less likely to be economically dependent on men (Kane, 2000), As such, Black women might see themselves as naturally fitting the so-called nontraditional subtype, as women who are independent and take-charge. However, from their point of view, would this instead be the traditional subtype? Following this, compared to White women, Black women are also more likely to consider traits such as strength, resilience and independence as important aspects of femininity and womanhood (Settles et al., 2008; Belgrave et al., 2015). Such traits align more closely with the nontraditional rather than the traditional subtype and form the basis of the Strong Black Woman ideology frequently endorsed by Black women (Davis et al., 2018).

Glick and Fiske (2001, 2011) have argued that BS in particular is important for maintaining the gender status quo, as women can be convinced to buy into and endorse it themselves, something they usually do not do with HS. But once again, an intersectional perspective suggests that this acquiescence might not come as easily to Black women as it does to White women. Hurtado (1989) has argued more broadly that White women, due to their more frequent interpersonal (i.e., romantic) relationships with White men, have a

more vested interest in placating them than do women of color. Because of their race, White women have a measure of privilege that women of color do not; thus, although they are disadvantaged by the gender status quo, they are advantaged by the racial status quo. Therefore, by submitting to the gender hierarchy they can ensure that they at the least maintain their racial privilege. Women of color, in contrast, generally have less reason to placate White men in interpersonal relationships, as they will always be disadvantaged in terms of both their gender and their race. Although written before the development of AS theory, Hurtado's analysis is interesting in the context of BS in particular, as it suggests that women of color are less likely to benefit from BS than are White women, which in turn suggests that they may be less likely to endorse it themselves.

More broadly, research reveals that Black Americans, both women and men, tend to have less traditional gender attitudes than their White counterparts, again suggesting that their 'default' ideas about gender may not be in line with those that underlie AS thinking (if indeed AS is primarily based on Whites' gender thinking; Kane, 2000; Carter et al., 2009; Harnois, 2017). More specifically, such research suggests that Black women hold the least traditional gender role attitudes and White men the most traditional. Interestingly, some research suggests that Black Americans' experience with racial discrimination might make them more aware of gender inequality and might also make them more aware of the social origins of gender inequality (Kane, 2000; Harnois, 2017). Perhaps this experience with racial discrimination (which can also take paternalistic forms) makes Black women in particular more likely to identify BS as problematic, thus lowering their endorsement. Although not tested directly, this idea is tentatively

supported by a study from Settles and her colleagues (2008), who compared Black and White women's perceptions of womanhood. One intriguing finding was that Black women were more likely than White women to report feeling frustrated with experiences that could be labeled manifestations of BS. Relatedly, White women were more likely to describe perceived advantages of their gender that Glick and Fiske would likely label examples of BS (e.g., acts of chivalry directed towards them).

Taken together, this research suggests that Black women may have an aversion to BS not seen among White women. More egalitarian gender attitudes should also be associated with lower HS compared to White men and women. Additionally, Black Americans' experience with racism may make then more aware of the harms of gender inequality, in turn making them less likely to endorse an ideology such as HS (Kane, 2000).

However, there is an important caveat to the research on racial differences in gender attitudes as reviewed by Kane (2000). Specifically, the attitude domain may be important to consider, as some research suggests that while Black Americans may have more egalitarian attitudes towards maternal employment and women's family roles/obligations compared to White Americans, but may have less egalitarian attitudes regarding intimate and/or romantic relationships. If the latter are more important to BS endorsement than the former (as argued by Davis et al. [2022], for example), Black Americans may be expected to score higher on BS than White Americans.

Importantly, AS endorsement among different racial and ethnic groups appears to have been directly examined just twice in prior research. In the first study, Hayes and Swim (2013) tested the validity and internal reliability of four popular measures of

sexism: the Modern Sexism Scale (Swim et al., 1995), the Attitudes Towards Women Scale (Spence et al., 1973), and the two subscales of the ASI (Glick & Fiske, 1996). Using data collected from college students in the mid-to-late 1990s, Hayes and Swim concluded that the measures were generally suitable across four different groups: European Americans, African Americans, Latina/o Americans, and Asian Americans. However, the authors did suggest that researchers be cautious in using the BS subscale with African Americans or Latina/o Americans, due to low internal reliability on this subscale and inconsistencies in correlations with other sexism measures compared to the other two groups. Additionally, in contrast with previous research, HS and BS were not correlated among African American participants and there were no gender differences for this group, suggesting that HS and BS may not work together as expected among Black Americans. Aside from these concerns, Hayes and Swim found that African Americans endorsed BS to a greater degree than European Americans, but there were no group differences in HS scores.

This study, then, seems to contradict the pattern of findings that the literature generally suggests would occur. However, there are some important limitations to Hayes and Swim's work that suggest another look at this question is needed. First, as acknowledged by Hayes and Swim themselves, the data analyzed are well over twenty years old at this point and sample sizes for the non-White groups were quite small, so whether such results would generalize to Black men and women today is unknown. Moreover, Hayes and Swim's decision to use a factorial ANOVA (i.e., Race x Gender) to analyze their data is an approach that is now discouraged by intersectional researchers, who have argued that is best to examine an intersectional whole (e.g., Black woman)

rather than continue to treat the two identities as independent (e.g., Black and woman), even if just statistically (e.g., Warner et al., 2018; Warner & Shields, 2018). Finally, although Hayes and Swim did consider measurement validity (via Cronbach's alpha) and factor structure (via confirmatory factor analysis), they did not conduct a full MI analysis as recommended by psychometric experts today (e.g., Putnick & Bornstein, 2016). Therefore, the question of whether or not Hayes and Swim's group comparisons were valid was not fully answered.

The second study is more recent but reported similar findings as Hayes and Swim. Davis and her colleagues (2022) investigated race and gender group differences in BS only among Black and White undergraduate students. They also examined possible mediators of group differences in BS endorsement. Davis et al. found that Black men and women scored higher on the BS subscale[8] than White men and women; they also found no gender differences for Black participants, but did for White participants. Their results, too, conflict with what might be expected based on previous research, but there are again important limitations to consider, as acknowledged by the authors themselves. First, the numbers of Black men and Black women in the sample were quite small compared to the numbers of White men and White women, which could have influenced the results. Additionally, the sample consisted of undergraduate students, whose young ages and lack of life experience may lead them to respond differently than older adults. Specifically, Davis et al. argued that gender attitudes about romantic relationships may be more important to BS endorsement than gender attitudes about maternal employment and

[8] One item from the protective paternalism subscale was dropped due to its poor performance among all participants: "In a disaster, women need not be rescued before men."

women's family roles; because Black Americans appear to be more traditional regarding the former domain than the latter, they were expected to have higher BS scores than White participants. However, this emphasis on gender roles in romantic relationships may be more salient for younger adults than older adults, who have more experience in the workplace and heading a household – and thus gender attitudes in those domains may be more salient in their responding. In general, it is unknown whether or not these findings can generalize to older Black Americans in particular. Finally, Davis et al. acknowledged that their findings may be a measurement artifact rather than a true difference, as they did not test the MI of the ASI's BS subscale before comparing group means. They suggest that BS may manifest differently for Black Americans than for White Americans, something that would not be captured outside of direct tests of measurement invariance. Thus, additional research is needed that does take this necessary analytic step to determine if valid racial group comparisons can even be made using the ASI.

**Hypotheses**

The available literature suggests that race should affect endorsement of HS and BS, as outlined in the hypotheses below. However, it must be noted that rather than resulting in differential endorsement between Black and White participants, it is possible that the different gender norms, expectations, and beliefs held by these groups may instead disrupt the meaning of the AS construct as well as the structure of its measure, the ASI, for Black participants. In other words, it is possible that MI cannot be established across the four race-by-gender groups of interest and therefore specific hypotheses about endorsement cannot be validly tested.

However, if the MI of the ASI *can* be established across the groups of interest, the following hypotheses will be tested:

*Hypothesis 1.1a.* White men will score higher on HS than Black men.

*Hypothesis 1.1b.* White men will score higher on BS than Black men.

*Hypothesis 1.2a.* White men will score higher on HS than White women.

*Hypothesis 1.2b.* White men will score higher on BS than White women, although this difference will be smaller than the difference found for HS.

*Hypothesis 1.3a.* Black men will score higher on HS than Black women.

*Hypothesis 1.3b.* Black men will score higher on BS than Black women, although this difference will be smaller than the difference found for HS.

*Hypothesis 1.4a.* White women will score higher on HS than Black women.

*Hypothesis 1.4a.* White women will score higher on BS than Black women.

If MI can be confirmed for some groups but not others, only the appropriate hypotheses will be tested. For example, if MI were to be established for White men and women but not Black men and women, Hypotheses 1.2a and 1.2b would be tested but no others.

**Method**

Study 1 was part of a larger three-wave, multi-investigator panel study conducted by the Center for the Study of Political Psychology (CSPP) at the University of Minnesota (UMN) in the fall of 2020. Participants were recruited by the market research and data analytics firm YouGov, and matched to a sampling frame on gender, age, race, and education in order to produce a nationally representative sample. In addition, an oversample of Black Americans was also recruited. Analyses focused on responses at Wave 1 only, which was in the field from October 6 through October 14, 2020.

### *Participants*

Of the 2,615 participants in the panel study (recruited by YouGov), $N$ = 2,304 were eligible for this study based on gender and racial identity. Analyses included Wave 1 participants who responded to *all* ASI items at Wave 1. This final sample ($N$ = 2,174*)* included 247 Black men, 385 Black women, 731White men, and 811 White women. Additional demographic information is provided in Tables 1 through 5. As Table 1 indicates, White men and women in the sample were, on average, older than Black men and women, with Black women being the youngest group on average. Additionally, as shown in Table 2, over 60% of the sample has at least some college education and less than 5% had no high school education. Notably, the majorities of Black men and women identified as Democrats, whereas White men and women were more evenly split regarding their political party affiliation (Table 4). Furthermore, 38% of the overall sample considered themselves to be at least slightly liberal, 31% moderate, and 31% at least slightly conservative (Table 5). However, compared to White participants, more Black men and women considered themselves to be at least slightly liberal and, compared to Black men and women, more White men and women considered themselves to be at least slightly conservative.

### *Measures*[9]

**Wave 1 Measures.** Participants completed the full 22-item ASI within a single block of the survey (see Appendix A for full measure). Items such as "Most women interpret innocent remarks as sexist" represent HS whereas items such as "Women should

[9] Additional items were included at Waves 2 and 3 for the purposes of evaluating the discriminant and predictive validity of the ASI. However, due to the results of the MI analysis, these items were not analyzed and therefore are not discussed here.

be cherished and protected by men" represent BS. Participants indicated agreement with each item on a Likert scale, ranging from 1 (*Disagree strongly*) to 6 (*Agree strongly*), with six items reverse scored. Items representing HS were averaged for the HS score and items representing BS were averaged for the BS score. BS subscale scores were also created in this manner.

After completing the ASI, participants next answered a set of follow-up questions developed by the researcher. These questions were exploratory in nature, designed to tap into what kinds of men and women participants tend to think about when responding to the ASI items, which could give insight into potential endorsement differences or even lack of MI. Participants were directly asked "Now, when you were answering the last set of questions, what came to mind? Specifically, what images and thoughts came to mind when asked to think about "women" and "men" and their relationships in society?" Participants were then asked "To what extent were the women and men you had in mind [racial identity]?" six times for each of the following identities: White/Caucasian, Black/African-American, Asian/Asian-American, Hispanic/Latino American, Middle Eastern American, and Native American. Response options included the following: 1 (*Not at all*), 2 (*A little*), 3 (*Somewhat*) and 4 (*A great deal*).

Prior to responding to the ASI, participants were asked questions to determine the importance of their racial (Black or White) and gender (man or woman) identities, based on Huddy et al.'s (2015) measure of partisan identity . These items were included for the purposes of exploring potential sources of noninvariance, if MI cannot be confirmed (i.e., a source of noninvariance may well be differing racial identities that respondents report having in mind). However, if MI is established, these items could moderate the effect of

race-by-gender identity on AS endorsement. For each identity, participants were asked two questions: "How important is being [identity] to you?" [*importance* item] and "To what extent do you think of yourself as being a/an [identity]?" [*self-view* item]. Response options ranged from 1 (*Extremely important*) to 5 (*Not at all important*) for the first question and from 1 (*A great deal*) to 5 (*Not at all*) for the second; however, these were reverse-scored for analysis, so that higher scores indicate a stronger racial or gender identity. These two items were combined to form a racial identity centrality score and a gender identity centrality score.

**Demographics**. Demographic information, such as education, income, and religion, were provided by YouGov while questions concerning participants' gender, race, political ideology, and political party identity were asked at the beginning of the Wave 1 survey.

### *Procedure*

As noted, participants were recruited and compensated by YouGov to participate in a three-wave, multi-investigator panel study conducted by the UMN's CSPP. The survey was administered online via Qualtrics and included material from other investigator studies besides the one presented here. Wave 1 was in the field from October 6, 2020 through October 14, 2020; Wave 2 was in the field from October 23, 2020 through November 2, 2020; and Wave 3 was in the field from November 9, 2020 through November 16, 2020.

### *Analysis Plan*

All statistical analyses were conducted using R Statistical Software (version 4.2.2; R Core Team, 2022) and STATA (version 15.1; StataCorp, 2017). More specifically, the

*lavaan* R package (Rosseel, 2012) was used to conduct the MI analyses while the *stats* (R Core Team, 2022) and *rstatix* (Kassambara, 2023) packages were used to conduct mean comparison tests. Initial data cleaning and preparation were done in STATA.

**Measurement Invariance (MI)**. MI of the ASI across the four gender-by-race groups was tested using a structural equation modeling (SEM) approach, in which model parameters were increasingly constrained to test increasingly strict levels of invariance across groups (a 'bottom-up' approach). If scalar invariance – the third level of MI – is confirmed, subsequent analyses can continue as planned.[10] Due to the complexity of the ASI measurement model, which includes a second-order factor, the MI analysis followed the procedure outlined by Rudnev et al. (2018), which itself is based on traditional procedures as outlined in Putnick and Bornstein's (2016) review.

In determining whether or not MI is present, multiple pieces of evidence need to be considered at each level. Unfortunately, this evidence can be contradictory, suggesting both invariance and noninvariance rather than just one or the other. As such, it is important to be cautious in drawing conclusions about MI, as findings based on noninvariant measures can have severe consequences, as previously discussed.

***Initial Steps.*** Before starting the formal MI analysis, a few informal analyses were conducted to get a better sense of what the MI analysis might reveal and to provide insight into possible explanations if MI was not confirmed across the four groups, a preliminary step followed in some but not all MI investigations. First, internal reliability (Cronbach's alpha) was checked for each scale and subscale for each race-by-gender

[10] There is a fourth level of MI, known as residual (or strict) invariance, but there is general agreement that this level is not a prerequisite for testing mean differences (Putnick & Bornstein, 2016) and therefore it is not addressed in either Study 1 or 2.

group. Poor internal reliability for any of the groups would suggest that there may be a problem with the factor structure. Second, HS-BS correlations for each group were checked against the typical finding [originally reported by Glick and Fiske (1996)] of $r$ = .40-.50. Correlations much higher or lower than this standard could be a sign of potential MI issues. Third, for each of the four groups, three models were tested for fit, following Glick and Fiske's (1996) original approach: a one-factor model (AS only), a two-factor model (HS and BS), and the full model (including the second-order BS factor and the three BS subfactors). If any groups show better fit on the one- or two-factor model compared to the full model, this would suggest the presence of measurement noninvariance. Once these initial steps were completed, MI was formally tested.

***Configural Invariance*.** Configural invariance is the least strict level of MI. At this level, the researcher is checking whether or not the model form holds across the groups of interest (Putnick & Bornstein, 2016). In other words, the key question here is: Does the expected factor structure of the measure hold across groups? To answer this, we determine whether or not the same items load significantly on the same (expected) factors for each group. If the known factor structure does not adequately fit each group, then we have failed to confirm configural invariance.

Statistically, configural invariance is tested by simultaneously fitting the model for all groups via confirmatory factor analysis and then check the overall fit indices to determine if fit is adequate. This initial model is generally unconstrained; however, following Rudnev et al.'s (2018) recommendations for model identification, one loading per factor is fixed to 1 and one item intercept per factor is fixed to zero. Recent MI research (e.g., Rudnev et al., 2018; Pauls et al., 2019) has typically followed the

standards for good fit suggested by Chen et al. (2005), which were in turn based on those put forth by Cheung and Rensvold (2002). These guidelines consider three different fit indexes:

- The Comparative fit index (CFI) should be greater than .90 for adequate fit or greater than .95 for good fit.
- The root mean squared error of approximation (RMSEA) should be less than .08 for adequate fit or less than .05 for good fit.
- The standardized root mean square residual (SRMR) should be less than .08 for adequate fit or less than .05 for good fit.

If the indices together suggest good fit – or even adequate fit – we have established configural invariance. An additional check involves examining the factor loadings to determine if any are nonsignificant for one or more groups, but significant for other groups. If there are inconsistencies in the significance of factor loadings, this points to configural noninvariance. Moreover, modification indices can provide insight regarding whether items might load better on other factors for certain groups. If there are discrepancies in where items best load, this also points to configural noninvariance. The factor loadings and modification indices can be especially helpful checks if the fit indices suggest different conclusions regarding model fit.

***Metric Invariance.*** If configural invariance is established, we then move on to check metric (or weak) invariance. Metric invariance concerns the equivalence of the item loadings on the factors, asking whether or not each item contributes to the factor to a similar degree across groups (Putnick & Bornstein, 2016). If the factor loadings are equivalent across groups, metric invariance is established and we can meaningfully

compare unstandardized regression coefficients across groups. If factor loadings are not equal across groups, metric invariance is not established and group comparisons cannot be made, as unequal loadings may indicate that the construct is understood differently across groups (Davidov et al., 2014).

Metric invariance is tested by fitting a new model in which factor loadings are fixed to be equal across each group of interest. This new model (i.e., the metric model) is then compared to the configural (or unconstrained) model using the chi-square difference (likelihood ratio) test: If the test is significant, it suggests that the constraints on the metric model may be too strict, thus metric invariance is not confirmed (Chen et al., 2005). If the test is not significant, this suggests that both models fit the data equally well, which would confirm metric invariance (Pauls et al., 2019). Some researchers (e.g., Chen et al., 2005) have cautioned that large sample sizes[11] can affect the chi-square test such that even a very small difference can be flagged as significant. Therefore, researchers have recommended that changes in fit indices between the configural and metric models should also be examined. The recommended thresholds vary depending on other factors (e.g., overall sample size, sample sizes across groups), so both the more conservative and less conservative cutoffs are provided here (Chen, 2007):

- ΔCFI greater than .005 or .010
- ΔRMSEA greater than .010 or .015
- ΔSRMR greater than .025 or .030

[11] Per Chen (2007) and Cheung and Rensvold (2002), a large sample size is $N > 300$.

If the respective differences are greater than the cutoffs provided, it is evidence that metric invariance is not present. As with configural invariance, all of these factors – the chi-square test and the changes in fit indices – should be considered when drawing a conclusion about the presence or lack of metric invariance.

Because the ASI includes a second-order factor, metric invariance must be tested twice (Rudnev et al., 2018; Pauls et al., 2019). First, only the first-order factor loadings are fixed to be equivalent in the metric model (i.e., the first-order metric model). If first-order metric invariance is confirmed (by comparing the first-order metric model with the configural model), a second metric model in which both the first- *and* second-order factor loadings are fixed to be equivalent across groups is tested. This second-order metric model is compared to the first-order metric model to determine if fit decreased (disconfirming invariance) or remained the same (confirming invariance).

***Scalar Invariance.*** If both levels of metric invariance are confirmed, scalar (or strong) invariance is examined. Most experts agree that scalar invariance must be established in order to conduct meaningful group comparisons (Putnick & Bornstein, 2016). Scalar invariance concerns the equivalence of the item intercepts; if such equivalence exists, it means that the factors' scales are measured with the same units and have the same zero point for all groups, both of which are necessary for factor mean comparisons to be valid (Rudnev et al., 2018). The presence of unequal intercepts might indicate that groups use the measure in different ways, making group comparisons inappropriate (Davidov et al., 2014).

To test for scalar invariance, we once again fit a new model (i.e., the scalar model) to compare to the previously established metric model. In this case, we fix the

item intercepts such that they are equal across groups. Once again, due to the ASI's second-order factor, two scalar models must be tested sequentially. First, we fit a model in which the first-order item intercepts are set to be equal across groups and then compare this first-order scalar model to the second-order metric model to determine if the fit significantly changed with the new model. If no change is observed, we next fit a model in which both the first-order and second-order item intercepts are fixed to be equal across groups. This second-order scalar model is then compared to the first-order scalar model to determine if second-order scalar invariance is present.

While the cutoffs for changes in CFI and RMSEA remain the same when considering scalar invariance, Chen's (2007) simulation studies suggest that change in SRMR is more sensitive to metric invariance than to scalar invariance. Therefore, different thresholds (more conservative and less conservative) were recommended for change in SRMR at this level of testing:

- ΔSRMR greater than .005 or .010

If the change in SRMR exceeds the cutoff, it is evidence that scalar invariance is not present. As with other levels of AI, these multiple factors should be considered in concert when making a conclusion about the presence of scalar invariance.

***The Consequences of Noninvariance*.** If each level of invariance, from configural through second-order scalar, is confirmed, we can continue with our analyses and test the proposed hypotheses (as outlined below). However, if we do not achieve the prerequisite for mean comparisons – i.e., second-order scalar invariance – group analyses *cannot* continue as originally intended and the hypotheses must remain untested. Putnick and Bornstein (2016) review the options available at each level of noninvariance. One option

is to accept that the construct or measure is invariant and simply discontinue analyses. However, as another option, the researcher can further investigate the sources of the noninvariance and make potential adjustments to the construct and/or measure based on those sources. According to Dadinov et al. (2014), it is possible that invariance results simply from measurement problems unrelated to the actual construct – if the problems are corrected or accounted for, then a better test of MI can be undertaken. More concerning are the more complicated sources of invariance, which can include fundamental differences in the meaning of the construct across groups and/or different interpretations of measurement items across groups. In such cases, the researcher will need to investigate *why* these differences exist, perhaps by looking at societal/cultural or historical contexts. This may involve entirely new lines of research. If proper explanations can be identified and tested, the construct theory and/or measure can be adjusted to 'work' across all groups of interest.

Other solutions may also be available. First, if there are more than two groups of interest, it may be possible that MI exists between some groups but not the full set of groups (Davidov et al., 2014). Therefore, mean comparisons can be conducted with just the invariant groups. For example, MI may exist only between White men and White women; thus, mean differences can be analyzed for these two groups only, omitting Black participants from analyses. A second possibility is to simply remove the noninvariant items from the model and re-check MI. However, this is a data-driven approach (rather than a theory-driven one) and without knowing *why* those items are invariant (e.g., because of their content or their wording), we cannot know if their

removal from the measure changes the meaning of the construct that is being measured, which makes this a risky option to pursue (Putnick & Bornstein, 2016).

**Additional Validity Assessments.** If MI is established, additional validity checks would be conducted. Convergent validity would be assessed by correlating BS scores and SSI scores and then determining if correlations significantly differ among the four groups. Additionally, predictive validity would be assessed via OLS regression analyses to determine if HS and BS differentially predict the selected outcome variables (including policy attitudes and feeling thermometers) depending on group membership.

**Hypothesis Testing.** If MI is established, group comparisons can be carried out. Specifically, we can test for differences in group means for HS, BS and the three BS subfactors. The planned approach follows from Warner and Shields' (2018) suggestion to reconceive what would typically be treated as two factors (i.e., gender and race) instead as a single factor with four levels: Black man, Black woman, White man, and White woman. Although not perfect, this helps to preserve the intersectional locations of interest and does not treat the two identities (gender and race) as statistically independent, which would undermine the intersectional approach (Warner et al., 2018). Therefore, scores will initially be compared using a one-way analysis of variance (ANOVA), adjusted to account for unequal variances, if necessary; if the ANOVA is significant, appropriate post-hoc tests (e.g., Tukey's HSD or Games-Howell in the case of unequal variances) will be conducted to test the specific hypotheses.

**Exploratory Analyses in the Face of MI.** If MI is not established, the original hypotheses cannot be tested as planned. Instead, we can begin to investigate the sources of the noninvariance. It would not be possible to fully answer questions about the source

of noninvariance in this sample given the nature of the study, but potential explanations can be examined in a preliminary manner, with future research building on suggestions provided here.

***Identity Centrality Items.*** Measurement noninvariance may be due to differences in construct or item meanings across groups. One possible explanation (or hypothesis) for such differences is that gender is more central to identity for some groups than it is for others and this affects their perceptions and/or experiences of sexism; similarly, race might be more central to identity for some groups than it is for others, which could also affect their perceptions and/or experiences of sexism. Therefore, between-group differences in racial identity centrality and gender identity centrality can be tested (via one-way ANOVA and post-hoc tests, as appropriate) to determine if this is a viable explanation worth additional investigation in future research.

As previously noted, these items could also be tested as moderators in the case that MI is established across the four race-by-gender groups.

***ASI Follow-Up Questions*.** It is also possible that measurement noninvariance is the result of respondents thinking about different kinds of women and men while rating their agreement with the ASI items. If different men and women elicit different attitudes, stereotypes and so on, this could profoundly impact responding, and although the ASI instructs respondents to think about men and women generally – providing no description other than "men and women" – there is no guarantee that this is what respondents actually do.

While the initial purpose of these follow-up questions was to get a sense, going into Study 2, of the kinds of men and women respondents naturally think about when

reading the items on the ASI, the findings here can give some early insight into whether that explanation is possible. Using these questions, we can determine whether there are substantive group differences in how people think when responding to the ASI by comparing means on the Black and White follow-up items. We can also compare the frequency of certain responses across these items by group.

## Results

### *Measurement Invariance*

**Initial Steps.** Cronbach's alpha ($\alpha$) coefficients for each scale and subscale of the ASI are reported by group in Table 6. This table also reports the coefficients for the overall sample, as well as the average coefficients reported by Glick and Fiske in their original 1996 paper, for comparison purposes.[12] Although many of the group coefficients are similar to the averages reported by Glick and Fiske (1996) and meet the typical standard of acceptance ($\alpha \geq .70$; Tavakol & Dennick, 2011), a few suggest that caution is needed moving forward. In particular, for the heterosexual intimacy (INT) and protective paternalism (PAT) subscales, reliability coefficients are especially low for Black Men ($\alpha = .63$ and $\alpha = .53$, respectively) and Black Women ($\alpha = .51$ and $\alpha = .54$, respectively). In contrast, while lower than the coefficients for the HS and BS scales, the alpha coefficients for these subscales are above .70 for White men and White women. These discrepancies indicate that there may be a problem with one or more of the intimacy and paternalism items for Black men and Black women, but not for White men and White women.

[12] Glick and Fiske (1996) did not report internal reliability coefficients for the three BS subscales, possibly because the numbers of items per subscale are quite small. Nevertheless, it is still worthwhile to examine how these groups of items hang together and determine if any items do not seem to fit within their subscale.

Next, the HS-BS Pearson correlation coefficient (*r*) was checked for each group and compared to the typical range of *r* = .40-.50 reported in AS research (e.g., Glick & Fiske, 1996, 1997, 2011). The correlations for White men (*r* = .41) and White women (*r* = .52) were in the expected range. However, the correlations for Black men (*r* = .07) and Black women (*r* = .16) were extremely low, again suggesting that the ASI may not be working in the same way across all four groups. However, the different correlations could also simply reflect differences in endorsement, rather than a sign of measurement noninvariance. For example, perhaps Black men and women are less ambivalent in their sexist attitudes – i.e., they lean toward HS or BS, not both – than White men and women. Thus, while this is an important piece of the puzzle, it does not tell us everything we need to know about the MI of the ASI.

The final preliminary step was to check the fit of different models within each race-by-gender group. For each group, the full model proved to best fit the data, followed by the two-factor model, and finally the one-factor model.[13] But the fit indices for the full model varied among the groups, suggesting that the expected factor structure might hold for some groups but not others. As reported in Table 7, across all fit measures, the only group for which we can definitively say fit is adequate is White women, with the fit for White men being on the cusp of adequate fit. Fit measures are especially poor for Black men and, to a somewhat lesser extent, Black women. In line with the internal reliability and correlational analysis previously reported, these findings suggest potential problems

[13] Fit indices for each of the three models for each race-by-gender group are reported in the Supplemental Materials (Tables S1-S5).

with establishing configural invariance. However, a formal test must be conducted before drawing this conclusion.

**Configural Invariance.** The first step in the formal test of MI is to determine if there is configural invariance across all four groups. This is accomplished by simultaneously fitting the model for all groups and then checking the fit indices. The model was identified using the marker indicator approach as suggested by Rudnev et al. (2018). Following their strategy (see their Table 1 on p. 56), the configural model was fit as follows:

1. First-order factor loadings were free, except one per factor was fixed to 1
2. First-order item intercepts were free, but one per factor was fixed to 0
3. First-order latent means/intercepts were free
4. Second-order factor loadings were free, except one per factor was fixed to 1
5. Second-order latent means were fixed to 0

The indicator item for each factor was selected at random and changing the indicator item did not alter the final conclusion regarding configural invariance.

Taken together, the results suggest that configural invariance cannot been established. Two of the three goodness-of-fit indices clearly signaled less than adequate fit: The CFI was .869, failing to reach the cutoff of .90 for adequate fit, while the RMSEA was .082, above the adequate fit cutoff of .080.[14] The SRMR was .080, which matches the typical cut off for adequate fit. The chi-square goodness-of-fit test was

[14] Some researchers might argue that these indices are 'close enough' to the cutoffs for acceptable fit, especially if rounded. To be as consistent as possible, I adhered strictly to the recommended cutoffs to determine fit. Furthermore, the pattern of nonsignificant factor loadings (discussed in the next paragraph and displayed in Table 8) confirm that the factor structure does not hold for all groups.

significant, $\chi^2(820) = 3792.68$, $p < .001$, suggesting poor fit; however, this could be a reflection of the large sample size rather than poor model fit. Taken together, the fit indices and the chi-square test suggest that configural invariance cannot be confirmed. Additional evidence was also considered, which further supports this conclusion.

An examination of the factor loadings revealed that all factor loadings were significant, as expected, for White men and White women based on the standard significance level of $\alpha = .05$. Unexpectedly but in line with the configural invariance test, there were four nonsignificant factor loadings each for Black men and Black women, as reported in Table 8 (with the full wording of each item reported in Table 9). Three items (hs11, int2, pat2) had nonsignificant factor loadings in both groups, while hs4 was nonsignificant for Black men and hs10 nonsignificant for Black women.[15]

Although the evidence for the lack of configural invariance provided by the fit indices may be marginal at best, the existence of nonsignificant factor loadings for some groups but not others firmly supports the conclusion that configural invariance is not present across these four groups, as it indicates that the expected factor structure does not hold for all groups. Indeed, the analyses suggest that these items best load on a separate factor altogether, unrelated to either the HS factor or the BS subfactors.

Intriguingly, the five noninvariant items consist of five of the ASI's six reverse-scored items, which raises questions about whether the noninvariance is due to the content of the items or due to their wording. This issue is further addressed in the

[15] Across all five of these items, White women had the highest factor loadings (all significant). Why this is the case cannot be directly examined with this dataset, but if AS theory and the ASI are based on White women's experiences of sexism, the measure may be most relevant to them among the four race-by-gender groups and thus the measure works best for them.

Discussion section, but I was able to begin examining one of these hypotheses using the data from the current study.

**Initial Responses to Noninvariance.**

***Revised Model.*** One option in response to the lack of MI at any level is to drop the noninvariant items from the model and test this new model (Davidov et al., 2014). As noted previously, this can be risky, as dropping items may change the meaning of the construct itself and such changes in meaning might vary across groups. However, because the invariant items may simply be the result of poor wording rather than content (or construal of that content), per se, a revised model in which the five invariant items were dropped was tested to determine if this is a worthwhile hypothesis to pursue further. *The revised configural model (which included 17 items instead of the full set of 22) had substantially better fit than the original configural model.* Although the chi-square test was significant, $\chi^2(460) = 1368.43$, $p < .001$, this again may be the result of the large sample size rather than poor fit, especially when considered in concert with the CFI (.952), RMSEA (.060), and SRMR (.055) values, which all easily meet the requirements for adequate (or even good) fit. These numbers are a stark contrast to the fit indices for the original ASI configural model.

Continuing the MI analysis, a first-order metric model was fit using the revised set of items. In isolation, this model also showed adequate fit based on the three goodness-of-fit indices: CFI = .949, RMSEA = .060, and SRMR = .063. The change in these indices were smaller than even the strictest cutoffs provided by Chen (2007), indicating that the first-order metric model fits the data just as well as the configural model, establishing first-order metric invariance. Although the chi-square test was

significant, suggesting worse fit for the first-order metric model, this could also be due to the large sample size given the minor changes in the goodness-of-fit indices.

Overall, eliminating the five invariant items from the ASI does seem to improve MI across these four specific race-by-gender groups. Fit statistics are reported in Table 10 for all five models tested – from the configural model through the second-order scalar model. The CFI, RMSEA and SRMR numbers for all models suggest adequate-to-good fit. Nevertheless, for MI, what we are most interested in are the changes in fit indices between models, which are reported in Table 11. Based on Chen's (2007) recommendations (outlined in the Method section), the changes in fit indices are generally small enough so as to not disrupt MI. However, we do see that the CFI decreases by .010 between the second-order metric model and the first-order scalar model, and by .006 between the first-order and second-order scalar models. Per Chen (2007), the strict cutoff for change in CFI is .005 while the less strict cutoff is .010; therefore, we must be cautious in our conclusions about full MI for the revised model, despite other changes in fit statistics falling below the recommended cutoffs. The complexity of the model (e.g., the presence of a second-order factor) as well as the unequal sample sizes across the four race-by-gender groups make it particularly difficult to determine just how strictly we should adhere to the cutoffs proposed by Chen (2007).

The revised model is promising in terms of MI and these findings provide preliminary support for the hypothesis that the lack of configural invariance is due to item wording rather than content. However, the content hypothesis (i.e., item content is the source of measurement noninvariance) cannot be ruled out using the data in this study and thus we cannot move forward with the assumption that *only* item wording is a

problem. Additional research is needed to determine if simply dropping the reverse-scored items is an appropriate response to measurement noninvariance of the original ASI model. This possibility is further considered in the Discussion section.

***White Men vs. White Women.*** Another potential response to MI is to determine if invariance exists between a subset of groups (rather than between all groups at once) so that at least some group comparisons can be analyzed (Davidov et al., 2014; Rudnev et al., 2018). In this case, the initial fit measures for the full model for White men and White women (see Table 7) suggested adequate fit (or close to it), so it is possible that there may be MI for these two groups. Therefore, an additional MI analysis was undertaken – with the original 22-item ASI – to determine if mean comparisons could be conducted between just these two groups.[16]

Indices of model fit for all five models are reported in Table 12. In addition to a significant chi-square test, the CFI (.891), while close, is below the standard cutoff for adequate fit (.90); however, the RMSEA (.078) and SRMR (.068) suggest adequate fit. Because we have conflicting evidence, it is difficult to confidently conclude that configural invariance is present for these two groups. For exploratory purposes, metric and scalar invariance were also considered. Based on the changes in fit statistics (see Table 13), we may have reason to be concerned about first-order scalar invariance given the change in CFI as well as second-order scalar invariance given the change in SRMR. Moreover, the CFI for all of the models is below .90, the cutoff for adequate fit. Altogether, given the contradictory evidence across multiple levels of invariance, it is

[16] A separate analysis showed that configural invariance was not confirmed when Black men and Black women alone were tested. This finding is in line with the poor fit statistics for these groups, even for the full model, from the initial analyses.

best to not proceed with this subset of group comparisons for this sample. This is potentially *too* conservative an approach to MI. However, the literature on MI (e.g., Putnick & Bornstein, 2016) suggests that, in general, researchers have not been conservative *enough* when it comes to MI; thus, this approach seems appropriate here.

### *Hypothesis Testing*

The results of the MI analyses suggest that the ASI does not work equally well across the four-race-by gender groups of interest, supporting the possibility that the differences in gender norms, expectations, and beliefs between Black and White Americans disrupts the meaning of the AS construct as well as the structure of the ASI for the former group. As a result, the original hypotheses cannot be tested, as the lack of MI indicates that group comparisons are not appropriate with this sample.

### *Additional Analyses in the Face of Measurement Noninvariance*

Although hypothesis testing cannot move forward as originally planned, I initiated exploratory analyses of potential explanations for the measurement noninvariance by delving into the identity centrality items and the ASI follow-up questions, with the idea that either or both may shed light on the basis for measurement noninvariance.

**Identity Centrality.** In Table 14, Pearson correlation coefficients are reported by group for each pair of identity centrality items. Most correlations are strong and all are significant, suggesting that – while not necessarily ideal – the two-item identity composites are adequate for these constructs. However, it is important to note that the correlation coefficients for the two White Identity items are considerably lower than the

other correlation coefficients, suggesting caution when moving forward with racial centrality for White participants.

Group means and standard deviations for each centrality measure are reported in Table 15. A one-way between-subjects ANOVA was conducted to compare the effect of race-by-gender group on both racial identity centrality and gender identity centrality. Due to significant Levene's tests – indicating unequal variances across groups – this analysis was carried out using the *oneway.test* function in R, part of the *stats* package (R Core Team, 2022). This function compares the means of one or more groups via an *F*-test, but does not assume equal variances across groups, instead following Welch's (1951) approach. A significant *F*-test was followed by post-hoc tests, specifically the Games-Howell post-hoc test which, while similar to Tukey's test, does not assume equal variances or sample sizes (*games.howell* function in R; Schlegal, n.d.).

Regarding racial identity centrality, there was a significant effect of race-by-gender group, $F(3.00, 859.23) = 241.67$, $p < .001$. Post-hoc tests revealed no differences in racial identity centrality between Black men and Black women, nor between White men and White women. However, Black men's racial identity centrality ($M = 4.26$, $SD = 1.12$) was significantly greater than White men's ($M = 2.84$, $SD = 1.29$) and White women's ($M = 2.93$, $SD = 1.124$), $p$s $< .001$. Similarly, Black women's racial identity centrality ($M = 4.35$, $SD = 1.03$) was significantly greater than White men's and White women's, $p$s $< .001$. These results suggest that for Black participants, their racial (i.e., Black) identity is an important aspect of their sense of self, but this is not necessarily the case for White participants and their racial (i.e., White) identity.

Regarding gender identity centrality, there was also a significant effect of race-by-gender group, $F(3.00, 867.67) = 92.53$, $p < .001$. Post-hoc tests revealed no differences in gender identity centrality between Black men and Black women, but there was a significant difference between White men's ($M = 3.49$, $SD = 1.20$) and White women's ($M = 3.93$, $SD = 1.08$) gender identity centrality. Moreover, Black men's gender identity centrality ($M = 4.38$, $SD = 0.96$) was significantly greater than White men's and White women's, $p$s $< .001$. Similarly, Black women's gender identity centrality ($M = 4.47$, $SD = 0.89$) was significantly greater than White men's and White women's, $p$s $< .001$. These results suggest that, like their racial identity, their gender identity may be more important to the sense of self for Black participants than it is for White participants; but this is not to say that gender is not an important part of self for the latter group. The significant differences in racial identity centrality between White and Black participants are larger than the differences in gender identity centrality.

Interestingly, across all four groups, the gender identity centrality score was greater than the racial identity centrality score, although that difference is larger for White participants than for Black participants.

**ASI Follow-Up Questions.** This set of six items was analyzed in two ways, with a specific focus on the White/Caucasian and Black/African American versions of the basic question. First, means for each item by group were examined and are reported in Table 16. The higher the rating, the greater the extent to which participants pictured [racial group] men and women while responding to the ASI. For each of the four groups, participants indicated having their own race/ethnicity in mind to the greatest extent across

all six options. In other words, when asked to think about men and women in general, participants primarily had men and women of their own racial group in mind.

Regarding the White version of the question, there was a significant effect of race-by-gender group, $F(3.00, 847.35) = 10.77$, $p < .001$]. Games-Howell post-hoc comparisons revealed no differences between Black men ($M = 2.34$, $SD = 1.10$) and Black women ($M = 2.25$, $SD = 1.10$), White men ($M = 2.51$, $SD = 1.23$) and White women ($M = 2.62$, $SD = 1.19$), or Black men and White men in the extent to which the women and men they had in mind while responding to the ASI were White. However, there are significant differences in this regard between Black men and White women, Black women and White men, and Black women and White women. Interestingly, all group means were above the midpoint of the response scale (i.e., 2 "A little"), suggesting that even for Black participants, there may be a tendency to default to White – at least to some degree – when considering women and men in general. This tendency is further discussed in Chapter 3.

This conclusion is supported by the breakdown of response option frequency in each group. Figure 2 displays the frequency of each of the four response options for each race-by-gender group. For example, when asked "To what extent were the women and men you had in mind White/Caucasian?", 30.4% of Black Men selected "Not at all," 23.5% selected "A little," 27.5% selected "Somewhat" and 18.6% selected "A great deal." This pattern is similar for Black women, suggesting that nearly one-fifth of Black participants thought of White men and women to a "great" extent when asked to consider men and women in general. This is in stark contrast to how White participants responded:

For both White men and White women, over 30% answered “A great deal” to this question.

When considering the Black version of the question – “To what extent were the women and men you had in mind Black/African American?” – we see some interesting contrasts which, in conjunction with the previous iteration’s results, warrant further investigation. As with the White version of the question, we see a significant effect of race-by-gender group. $F(3.00, 783.72) = 113.94$, $p < .001$. Games-Howell post-hoc tests revealed no differences between Black men and Black women, nor between White men and White women. The means for Black men ($M = 2.64$, $SD = 1.13$) and Black women ($M = 2.65$, $SD = 1.18$) were each greater than the means for White men ($M = 1.67$, $SD = 0.92$) and White women ($M = 1.73$, $SD = 0.90$). Interestingly, while means for Black men and women were above 2 (“A little”) for the White version of this question, means for White men and women were below 2 for the Black version of the question, suggesting that White participants may be less likely to imagine general men and women as being from outside their own racial group; in contrast, Black participants seem willing to consider more than just men and women of their own racial group as indicated by their responses to the White version of the question.

This possibility is supported by the response option frequencies for each group for this version of the follow-up question, as seen in Figure 3. Just as over 30% of White men and White women answered “A great deal” when asked to consider the extent to which they had *White* men and women in mind while responding to the ASI, over 30% of Black men and Black women had the same response when asked to consider the extent to which they had *Black* men and women in mind while responding to the ASI. Both

findings once again indicate that participants primarily thought of women and women from their own racial group. But a substantial difference emerges when you consider these groups' "Not at all" responses. For the White version of the question, all groups were in the 27-33% range, but for the Black version of the question, 58.8% of White men and 54.4% of White women answered "Not at all" while just 21.9% of Black men and 24.2% of Black women answered as such. Furthermore, we also see a substantial difference in the frequency of "A great deal" responses: For the White version of the question, 18.6% of Black men and 18.4% of Black women gave this response, but for the Black version of this question, just 6% of White men and 4% of White women selected this response. Taken together, these results support the conclusion that Black participants *do* think about White men and women, not just Black men and women, when asked to consider men and women in general; in contrast, White participants generally *do not* think about Black men and women at all when asked to consider men and women generally.[17] These findings might have important implications for the measurement noninvariance in this sample, considered next in the Discussion section.

**Discussion**

Study 1 was originally intended to answer the first key research question: Do Black and White Americans demonstrate different levels of HS and BS endorsement? However, the focus of the analysis shifted when the configural invariance of the ASI could not clearly be confirmed across the four race-by-gender groups. As such, the study's hypotheses could not be validly tested and, instead, hypotheses about possible

[17] Additional mean and response frequency analyses examining the other four versions of the follow-up question (i.e., Asian, Latinx, Middle Eastern, and Native American) support the conclusion that White participants generally did not consider non-White racial groups when responding to the ASI.

sources of the noninvariance were explored but not directly tested. Specifically, five items were identified as problematic (Table 10), especially for Black participants, and these five items happened to be five of the six reverse-scored items included in the ASI. It is unknown if these items 'misbehaved' due to their wording or their content. Previous research on other psychological measures indicates that research participants tend to have difficulty with reverse-scored items, which would suggest wording is the source of the noninvariance rather than the content. However, this prediction could not be tested with the data collected in Study 1. Detailed discussion of the findings, implications, limitations, and future research directions suggested by this study is taken up in Chapter 4.

Moving on to Study 2, it was originally intended to focus on the second primary research question: Does endorsement differ depending on the race of the women (and men) respondents are thinking about when responding to the ASI? However, due to the results of Study 1, an additional step needed to be taken in Study 2: a replication of the MI analyses using a new sample. If MI is once again absent, the primary research question may go unanswered, or answered but in a limited way (e.g., only for White participants). Replication of the Study 1 MI findings would provide more support for the conclusion that the ASI is not a valid measure of sexism for Black Americans. However, if the Study 1 results are *not* replicated – i.e., MI is confirmed across the four race-by-gender groups – it would further emphasize the need for additional research to fully flesh out AS theory and the ASI for Black Americans.

# Chapter 3

## Study 2: The Role of Target Race in Endorsement of HS and BS

As in Study 1, the role of race in AS was once again examined in Study 2, but from a different angle: Rather than the respondent's own racial group membership, Study 2 focused on the influence of the target's racial group membership on HS and BS endorsement. In other words, Study 2 addressed the 'target side' research question articulated in Chapter 1. To my knowledge, this has been directly studied just once before in the AS literature by McMahon and Kahn (2016), whose study is further discussed later in this section. Unlike in Study 1, the research question in Study 2 was addressed via a priming experiment, which will enable stronger causal inferences about race and the activation of HS and BS responding.

How and why might target race factor into HS and BS endorsement? In Study 1, it was hypothesized that the respondent's own racial identity may impact their HS and BS endorsement due to the different gender roles, beliefs, and norms that have developed among different racial groups in the United States. Similarly, in Study 2, it was expected that the race of the target may impact the respondent's HS and BS endorsement due to different cultural stereotypes about women of diverse racial groups. These stereotypes may lead to respondents to automatically and differentially subtype potential targets (i.e., women) as either traditional or nontraditional depending on their race. This, in turn, would lead to differential application of HS and BS to these varying targets. Previous research, while limited, has examined how other subtypes can essentially serve as proxies for the traditional and nontraditional subtypes at the heart of AS theory, thus triggering differential application of HS and BS. Sibley and Wilson (2004) had male participants

read a vignette about a sexual encounter between a man and a woman, the latter of whom was either chaste (positive sexual subtype) or promiscuous (negative sexual subtype). A shortened 12-item ASI (six items each for HS and BS) was altered to specifically reference "women like" the woman in the vignette, rather than "women" in general. Participants who read about the chaste woman scored lower on HS and higher on BS compared to participants who read about the promiscuous woman, supporting the argument that sexual subtypes should trigger differential evaluations of women consistent with the traditional and nontraditional subtypes underlying AS theory. In a similar study, Fowers and Fowers (2010) found that women also tend to direct HS toward promiscuous targets and BS toward chaste targets.

I argue here that race should similarly trigger differential evaluations of women regarding HS and BS applications. Specific stereotypes about racial minority women provide insight into how they might be typically subtyped (i.e., as either traditional or nontraditional) and thus how HS and BS might be applied to them compared to White women or even women in general. More specifically, stereotypes linked to power and sexuality should provide particular insight into HS responding whereas stereotypes linked to gender roles should provide particular insight into BS responding (Bareket & Fiske, 2023).

As previously mentioned, only one study has directly examined the influence of target race on HS and BS endorsement. McMahon and Kahn (2016) hypothesized that lacking any other information about the target woman, her race should cue specific HS and BS evaluations of her. Participants (male and female) were asked to evaluate a White (Emily) or Black (Lakisha) woman based only on her name and a facial image; thus, the

only information they had about the woman was her race. They then responded to a shortened version of the ASI directly referencing either Emily or Lakisha, similar to how Sibley and Wilson (2004) modified the ASI in their study. Participants also rated the woman on eight different positive personality traits assessing warmth and competence. McMahon and Kahn (2016) found that, consistent with their expectations, more BS was expressed toward the White woman than the Black woman, and the former was assigned more positive traits than the latter. However, there were no differences in the amount of HS expressed towards the two women. The authors concluded that a woman's race (in this case White or Black) can be enough to trigger a BS response; in other words, it does not necessarily require specific, known traits or behaviors (Glick et al., 1997) to subtype a woman. In a second study, McMahon and Kahn found that when sexual information, positive or negative, is also provided about the target woman (in addition to her race), BS evaluations change: more BS is expressed toward a chaste Black woman than a chaste White woman. This suggests that the 'typical' Black woman is subtyped negatively in terms of sexual behavior and this subtyping only changes with information that directly counters the negative/nontraditional stereotype.

McMahon and Kahn's (2016) two studies are critical steps towards understanding the role that target race plays in HS and BS endorsement, with their findings suggesting that race is a crucial variable that cannot be overlooked as it has been for so long. However, there are limitations to their work that the present study addresses, allowing for further insights about the relationship between race and AS. First, the present study went beyond the Black/White racial dichotomy by including two additional racial target groups: Asian Americans and Latinx Americans. Second, a more general racial prime

was used to determine if race triggers the expected subtyping at the group level rather than individual level. Third, the original, complete ASI referencing a group of women was used rather than a shortened, modified version referencing a specific individual target. Fourth, the two participant samples used in McMahon and Kahn's research were predominantly White (72% in first study and 63% in the second study), and thus differences in how participants of different races perceived the various target women may have been obscured. The present study rectified this to a small extent by recruiting roughly equal numbers of Black and White participants. Finally, the present study also included a test of measurement invariance (MI) to determine if the ASI works equally well for all participants before conducting group comparisons, serving as a replication of Study 1.

Briefly, in the experiment, participants were instructed to respond to the ASI while thinking of specific groups of men and women: men and women in general (control), White American men and women, Black American men and women, Asian White American men and women, or Latinx American men and women. These instructions were accompanied by images of representatives of the group (or no images in the control condition). Thus, participants were directly primed to consider specific racial groups while responding to the ASI. HS and BS scores were compared across conditions to determine whether the racial primes differentially affected HS and BS endorsement. In secondary analyses, the interaction between the racial prime and participants' own racial and gender identity was examined.

The next section of this chapter reviews the relevant research on stereotypes about women in general, about specific racial groups, and about women of specific racial

groups (i.e., intersectional stereotypes based on gender and race). All three types of stereotypes inform our understanding (and hypotheses) about how women of different races may be differentially subtyped and thus elicit differential applications of both HS and BS.

### *Gender, Race, and Gender-by-Race Stereotypes*

Much of the stereotype literature focuses on a single axis, i.e., a single grouping variable such as gender, race, or class (Voyles & Nadler, 2020). However, intersectionality theory suggests that this is a poor approach because rarely does such a singular group exist in isolation. As a result, an individual target has the potential to be stereotyped on multiple different fronts simultaneously. Although limited, stereotype research that does consider more than one grouping variable suggests that stereotyping is indeed not as straightforward as most research would lead us to believe. Bayton and his colleagues (1956) produced perhaps the first attempt to study intersectional stereotypes, though they were not called such at the time. Bayton et al. argued that participants in stereotype research may not be stereotyping on the intended or expected variable, as certain variables are conflated with one another, the most common example being race and social class. In their study, Black and White college students considered four groups: upper-class White Americans, upper-class Black Americans, lower-class White Americans, and lower-class Black Americans. For each group, participants selected from a provided list the traits that best described each group. Bayton and his colleagues found that class was the more powerful influence on stereotypes, as upper-class White and Black Americans were stereotyped similarly, as were lower-class White and Black Americans. Moreover, the upper-class stereotypes were similar to traditional positive

stereotypes of Whites whereas the lower-class stereotypes were similar to traditional negative stereotypes of Blacks. Bayton et al. suggested that when asked to stereotype generic 'Whites' and 'Blacks,' participants assume that most White Americans are upper-class while most Black Americans are lower-class. Thus 'Whites' and 'Blacks' are not simple, single-axis groups but multidimensional ones – an assertion that has been supported repeatedly by later research (Voyles & Nadler, 2020). Bayton and colleagues' study (1956) was an early indication that it is not enough to examine stereotypes, particularly racial stereotypes, in isolation.

This conclusion easily extends to gender stereotypes, as demonstrated by Landrine (1985). Building on Bayton et al.'s (1956) work, Landrine (1985) examined how race and class interact to inform stereotypes about women. She, too, found that class and race were conflated, with stereotypes about Black women being most similar to those about lower-class women, and stereotypes about White women being most similar to those about middle-class women. She concluded that when asked to consider 'a woman' in general, we have a tendency to also ascribe to her a race and social class, among other attributes, which in turn influences how we stereotype that woman in particular. This tendency extends to targets from all kinds of groups and, unless otherwise specified, most targets are assumed to hold additional identities that reflect power and status (e.g., White, male, able-bodied; Voyles & Nadler, 2020).

Study 2 builds on the idea that race informs stereotypes about gender, women in particular. Just as race has shaped how different racial groups think about gender (as discussed in the Chapter 2 – *Race, Gender, and Ambivalent Sexism* section), it has also shaped how groups stereotype one another on the basis of gender. This is perhaps most

clearly demonstrated in work by Niemann et al. (1994) and Ghavami and Peplau (2013), each research team demonstrating that gender stereotypes in the United States vary depending on the racial group under consideration. Because women are subject to different stereotypes depending on their (perceived) race, they may also be subject to differential application of HS and BS based on that race. McMahon and Kahn's (2016) research indicates that Black and White targets do trigger differential sexual subtyping, at least, and a further examination of gender-by-race stereotypes can illuminate how Asian American and Latina American women, in addition to White American and Black American women, are most likely to be subtyped according to the traditional vs. nontraditional dichotomy underlying AS. Given that traits and behaviors are not only gendered but also racialized (Pyke & Johnson, 2003), it seems likely that there is overlap between the stereotypes representing the two subtypes outlined in AS theory and stereotypes about specific groups of American women. The degree of overlap should then indicate how each racial group of women should be subtyped.

**White American Women**. There is research that examines stereotypes of White American women specifically, but it is limited; most often, stereotypes of White women are studied in comparison to stereotypes about other groups of women, rather than in isolation. Perhaps most important to understanding stereotypes of White women, however, is recognizing the fact that, by default, 'women' and 'femininity' have been and continue to be associated with Whiteness (Crenshaw, 1989; Cole & Zucker, 2007; Bowleg, 2008; Deliovsky, 2008). It is precisely this association – present both among lay people and psychologists (Goff & Kahn, 2013) – that creates the potential problems underlying AS theory and the ASI discussed in Chapter 1 of this paper. Specifically, as a

result of this association, stereotypes about women in general might actually apply to White women specifically, not women of color, an assertion that is supported by stereotype research.

Although Bayton and Muldrow (1968) were not interested in sex-role stereotypes specifically, in their research on racial stereotypes they did find that stereotypes of White women matched traditional female sex-role stereotypes of the time, which portrayed women as interpersonally sensitive, warm, and expressive, as well as less competent, independent, and logical than men (Broverman et al., 1972). Landrine's (1985) findings were similar: Participants in her study stereotyped White women as dependent, passive, emotional, and warm, closely matching the traditional general stereotype of women. More recent work by Ghavami and Peplau (2013) – directly informed by an intersectional perspective – confirmed that stereotypes about White women specifically are closely aligned with stereotypes about the more general group of 'women.' These authors asked participants to generate attributes for 17 different groups: five ethnic groups (i.e., Asian Americans, Blacks, Latinos, Middle Easterners, and Whites); men and women; and ten gender-by-ethnic groups (e.g., Black women, Middle Eastern men). In comparing the stereotypes generated for these different groups, Ghavami and Peplau found that the stereotypes generated for the target group 'women' were most similar to those generated for the target group 'White women,' supporting the idea that Americans tend to think of White women when thinking of women in general. Traits attributed to both groups included feminine, submissive, emotional and attractive. An earlier study by Niemann et al. (1994) did not examine stereotypes about women in general, but did find that

stereotypes about White women specifically included traits such attractive, pleasant, and sociable, once again suggesting a close link with general stereotypes about women.

Furthermore, research that directly compares stereotypes and perceptions of White women to those of Black women also suggests that White women more closely fit the general stereotype of women. Decades of cross-cultural and U.S.-based research alike indicate that stereotypes about men fall under the umbrella of 'agency' (e.g., competent, active, competitive, assertive, rational) whereas stereotypes about women fall under the umbrella of 'communality' or 'communalism' (e.g., family-oriented, warm, socially sensitive, compassionate, subordinate; Ellemers, 2018; Sczesny et al., 2018; Clarke, 2020; Gann-Bociek & Harvey, 2020). Intersectional scholars, however, have found that this communal stereotype is more applicable to White women than to Black women, who are stereotyped in ways more aligned with the agentic umbrella of traits (Donovan, 2011; Andreoletti et al., 2015).

More negatively valenced stereotypes *are* applied to White women, but typically only when additional information is known about them aside from their race. For example, Masters and colleagues (2014) found that White women are stereotyped negatively (e.g., lazy, promiscuous) when they are known to be poor and/or recipients of welfare; otherwise, White women are assumed to be middle- or upper-class and stereotyped accordingly (i.e., more positively; Landrine, 1985). McMahon and Kahn (2016) also found that when a White woman is known to be promiscuous, positive evaluations (e.g., BS) of her decrease. In the present experiment, such negative stereotypes should not be applied under the White prime condition because nothing is known about the target women other than their race.

Based on this evidence, the following hypotheses were put forth:

*Hypothesis 2.1a.* The White prime condition and the control condition should not significantly differ on HS.

*Hypothesis 2.1b.* The White prime condition and the control condition should not significantly differ on BS.

As context for additional hypotheses discussed in the following three subsections, it is also important to note that traits such as fragility, domesticity and purity – traits that closely align with the traditional subtype – have historically been more closely associated with White women (particularly upper-class White women) than women of color (McMahon & Kahn, 2018). Furthermore, recent research also supports the idea that protective paternalism in particular may be more likely to be expressed toward White women than women of color (McMahon and Kahn, 2018). Additionally, recent research by Brown-Iannuzzi and colleagues (2022) suggests that BS and femininity are more likely to be associated with White women than with Black women. Together, this evidence suggests that BS is more likely to be expressed toward White women than women of color, an implication that informs other hypotheses.

**Black American Women.** Regarding racial minority women, most research has examined stereotypes and images of Black women. This is not surprising given the origins of intersectionality theory, which in its initial development focused primarily on the prejudice faced by Black women (Crenshaw, 1989). A review of this literature confirms McMahon and Kahn's (2016) argument that, overall, Black women are negatively stereotyped, strongly suggesting that Black women are more likely to be subtyped as nontraditional women than as traditional women, and thus should elicit more

HS and less BS, especially compared to White women or women in general (who are often assumed to be White).

Black feminist scholars agree that Black women are the subject of four broad stereotypes or images, each consisting of several specific stereotyped traits and behaviors: Mammy, Jezebel (or Sexual Siren), Sapphire (or Matriarch), and Welfare Queen (Welfare Mother; West, 1995, 2008; Thomas et al., 2004; Woodard & Mastin, 2005; Butler-Barnes & Allen, 2020; Collins, 2022).[18] The first three images are quite old, dating back to the era of American slavery (West, 2008; Collins, 2022). The Welfare Queen image, in contrast, is relatively recent, dating to the 1960s; however, it also has roots in American slavery and the Jim Crow era (Woodard & Mastin, 2005; Littlefield, 2008; Collins, 2022). Although each is unique, these stereotypes are not mutually exclusive (West, 2005).

The first of these images, the Mammy, portrays Black women as loyal, happy, and devoted domestic servants to White people (West, 2005, 2008). She is subordinate, nurturing and self-sacrificing, prioritizing her White family (i.e., master or employer) over her own (Woodard & Mastin, 2005). In the 19th century, this imagery was used to counter narratives arguing that enslaved Black women were abused and miserable; if these women were actually happy as slaves, slavery could be seen as a humane institution that should continue indefinitely (West, 2008). More recently, it has been used to justify

[18] A fifth image, that of the Superwoman or Strong Black Woman, is a more contemporary combination of the Mammy and Sapphire images: she is caring and communal, but also strong and resilient (Rosenthal & Lobel, 2016). Because it is a combination of two stereotypes discussed in detail in this section, it is not discussed separately. Additionally, it seems more likely to be applied to educated, upper- or middle-class women (e.g., Michelle Obama, Oprah Winfrey; Rosenthal & Lobel, 2016) and therefore is unlikely to be relevant in this particular experimental setting, in which the only information provided about the targets is their race. Moreover, the Strong Black Woman stereotype may be more likely to be internalized by Black women themselves, rather than applied to them by others (Harrington et al., 2010; Davis et al., 2018).

the overrepresentation of Black women in domestic service, especially in the postbellum South (Woodard & Mastin, 2005). Although Mammy has many traits that would be considered proper for a traditional woman – such as her strong links with maternal caregiving – and match the general stereotype for women, it nevertheless sets Black women apart from White women and sets the former as inferior to the latter. Additionally, as shown by the Welfare Queen stereotype, although motherhood and related traits are generally viewed positively, this is not the case when the mother is Black (e.g., Rosenthal & Lobel, 2016).

The second image is that of the Jezebel or the Sexual Siren, which is particularly common in contemporary media (Littlefield, 2008). This name comes from the Biblical figure of Jezebel, whose name has become synonymous with a sexually immoral, deceitful woman (Butler-Barnes & Allen, 2020). Unlike Mammy, who is often presented as lacking sexuality, Jezebel is hypersexual and "cares for nothing but her own sexual satisfaction" (Woodard & Mastin, 2005, p. 272). She is promiscuous and sexually irresponsible, as well as seductive, manipulative, and sexually dominant, using sex to get whatever she wants from the men in her life (West, 2005, 2008; Cheeseborough et al., 2020). The Jezebel stereotype, too, emerged in the era of American slavery, during which Black women's bodies were seen as epitomizing uncontrolled sexuality (Davis & Tucker-Brown, 2013) and Black women themselves were said to constantly want and seek sex (West, 2008). As a result, White men – slaveowners in particular – could justify their sexual abuse and exploitation of enslaved Black women (Woodard & Mastin, 2005; West, 2008). This kind of justification continues today, even among Black men and women (Cheeseborough et al., 2020). The Jezebel image portrays a woman who easily

fits the negative sexual subtype (Sibley & Wilson, 2004; McMahon & Kahn, 2016) and thus would be subtyped as a nontraditional woman who violates feminine norms of chastity and propriety.

The third image is that of the Matriarch, also known as Sapphire.[19] This name is taken from Sapphire Stevens, a character from the long-running radio-turned-television show *Amos 'n' Andy*; however, the image itself once again dates back to the era of American slavery (West, 2008). Sapphire Stevens was a hostile, nagging wife to her good-for-nothing husband – traits which have come to embody the Sapphire/Matriarch stereotype. Sapphire is angry, combative, aggressive, and overbearing; in other words, she is masculine and thus the exact opposite of a White woman, especially in antebellum America (West, 2008). Unlike Mammy, who is nurturing and maternal, Sapphire is controlling and emasculating, dominating not only her children but also her husband; as such, she is not as maternal as she should be (Woodard & Mastin, 2005). As noted, this image emerged during the period of slavery in the United States. As West (2008) has argued, Black women during this time had every right to be angry: Not only were they themselves enslaved, but their children and husbands were often sold away, lost to them forever. The Sapphire image takes this justifiable anger and presents it as dangerous or even funny, as evidenced by characters like Sapphire Stevens (whose anger towards her husband was also justified). The traits most closely associated with the Sapphire

[19] Although the terms "Sapphire" and "Matriarch" appear to be more common in the academic literature, the term "Angry Black Woman" is also used, particularly in informal settings (Ashley, 2014). However, some researchers, such as Ashley (2014), have argued that the Sapphire/Matriarch image is a precursor to the more contemporary Angry Black Woman image, rather than the two being the same. Nevertheless, the overlap between the two, as described in the literature, is substantial; thus, only the former is specifically discussed in this section. Given the overlap, the Angry Black Woman stereotype – like the Sapphire/Matriarch stereotype – should lead to Black women being subtyped as nontraditional women.

stereotype are usually classified as masculine, painting Black women as not feminine enough and thus challenging the gender hierarchy – in other words, they are nontraditional.

Finally, the fourth image is that of the Welfare Queen, or Welfare Mother. This image is relatively recent – the term is closely associated with Ronald Reagan (Littlefield, 2008) – and emerged explicitly in the 1960s when welfare became racialized (Masters et al., 2014). Nevertheless, like the other three images, it does have roots dating in slavery, when Black women had far more children than White women, often not of their own volition (Woodard & Mastin, 2005; Collins, 2020). The Welfare Queen has no desire to work despite being a single mother; instead, she is content to live off of the state and takes advantage of the social safety net (Woodard & Mastin, 2005). Like Jezebel, she is sexually permissive and scheming (Masters et al., 2014). Although she is a mother, she is not a good one; her children are pawns in her schemes to avoid work and responsibility. As a result of this stereotype, Black women's motherhood is viewed unfavorably. Rosenthal and Lobel (2016) have found that compared to a pregnant White woman, a pregnant Black woman is more likely to be perceived as a single mother receiving public/government assistance, less likely to be perceived as using birth control, and assumed to already have at least one other child. As a result, rather than being celebrated, her impending motherhood is seen as a burden on society.[20] Thus, despite the Welfare

[20] Chavez (2004), in his examination of the discourse surrounding Latina reproduction, refers to "stratified reproduction," a concept that delineates "how for some groups women's reproduction is characterized positively, while for that of other women is 'disempowered'" (p. 174). Within the United States, White women's reproduction is highly valued while Black and Latina women's reproduction is devalued (as evidenced by racial disparities in, for example, infertility and receipt of medical services related to pregnancy [Greil et al., 2011]), thus creating a positive association between White women and motherhood, but a negative association between Black and Latina women and motherhood.

Queen being a mother, she fits the nontraditional subtype due to her promiscuous, scheming, and irresponsible nature.

Taken together, these four common stereotypes suggest that Black women are both hypersexual and masculine, as well as unlikely to properly fulfill the roles of wife and mother. Therefore, Black women should be more likely to be subtyped as nontraditional women rather than traditional women; thus, they should be more likely to be subject to HS and less likely to be subject to BS, especially in comparison to White women. Research that has directly compared stereotypes of Black women to those of White women supports these hypotheses. Research has shown that, in line with general racial stereotypes linking Blackness with maleness, Black women are rated more masculine than their White counterparts (Goff et al., 2008). Similarly, Andreoletti and colleagues (2015) found that Black women are less likely than White women to be perceived in gender stereotypic ways throughout the life span; in particular, they are perceived as just as competitive as men and are viewed as less communal than White women, even in old age.

Furthermore, in their study of intersectional stereotypes, Ghavami and Peplau (2013) found no overlap in attributes generated for the 'women' group and those generated for the 'Black women' group. Indeed, those generated for the latter were generally more negative than those generated for the former, and included such traits as promiscuous, not feminine, and aggressive. This finding replicated earlier work by Weitz and Gordon (1993), who found little overlap between traits most commonly ascribed to Black women and traits most commonly ascribed to American women in general. The former included more negative traits such as loud, aggressive, argumentative, stubborn,

bitchy and 'too many children.' Additionally, Ghavami and Peplau (2013) found no overlap in the top 15 attributes listed for Black women and those listed for White women; once again, the traits ascribed to Black women were, overall, more negative than those ascribed to White women, although there were negative and positive traits listed for both groups. These findings replicated those reported by Niemann et al. (1994) two decades earlier: Although participants in their study provided both negative and positive adjectives to describe Anglo American women and Black American women, the latter were ascribed more negative adjectives than the former, including traits such as antagonistic, unmannerly, and lower class.

In sum, prevalent stereotypes about Black women suggest that they are more likely to be categorized as nontraditional women than traditional women, especially compared to White women and women of an unspecified race (who are often assumed to be White). Therefore, the following hypotheses were put forward:

*Hypothesis 2.2a.* Compared to the White prime condition and the control condition, the Black prime condition should elicit greater HS scores.

*Hypothesis 2.2b.* Compared to the White prime condition and the control condition, the Black prime condition should elicit lower BS scores.

**Asian American Women.**[21] Although there has been a substantial amount of research on Asian stereotypes and women stereotypes, there is little research examining

[21] An important question to consider when asking research participants about their stereotypes of Asians and/or Asian Americans is, who 'counts' as Asian (Lee & Ramakrishnan, 2020)? Asian Americans are a pan-ethnic group, consisting of individuals from over 20 national origin groups, and includes East Asians (e.g., Chinese, Japanese, Korean), Southeast Asians (e.g., Vietnamese, Cambodian, Thai), and South Asians (e.g., Indian, Pakistani, Bangladeshi; Lee, 2019). Despite this diversity, most Americans have a limited view of this group. In one recent study, researchers found that Americans are more likely to consider East and Southeast Asians as prototypically Asian, rather than South Asians (Goh & McCue, 2021). Additional research supports the claim that the default for Asian is East Asian for most Americans

stereotypes about Asian American women specifically (Li, 2014). Often the latter are teased out by asking Asian American women themselves about how they are stereotyped (e.g., Mukkamala & Suyemoto, 2018). Nevertheless, consideration of general racial stereotypes about Asian Americans can provide insight into how Asian American women might be subtyped in terms of traditional or nontraditional.

Researchers agree that there are two overarching images of Asian Americans: the perpetual foreigner and the model minority (Zou & Cheryan, 2017; Lee & Hong, 2020; Bu & Borgida, 2021). The perpetual foreigner image has evolved from the 'yellow peril' stereotype that emerged as Asian workers (i.e., Asian men) began immigrating to the United States in the 19th century (Lee & Hong, 2020). These immigrants were seen as "cultural, economic, and moral threats to the White nation" who refused to assimilate to the American way of life (Lee & Hong, 2020, p. 166). This image was reinforced during World War II, when East Asian Americans and immigrants were persecuted in the aftermath of the Japanese attack on Pearl Harbor. Today, Asian Americans continue to be stereotyped according to this image: Recent research has shown that Asian Americans are frequently ascribed traits such as having accents, not speaking English well, and sticking to their own culture/race (Zou & Cheryan, 2017). As a result, Asian Americans are

(Lee & Ramakrishnan, 2020; Yamashita, 2021). Therefore, it is possible that when research participants are asked about their stereotypes of Asian Americans and Asian American women specifically, they are not thinking about South Asians at all. As a result, this body of literature has a potentially severe limitation in that Asian American stereotypes may not apply to all Asian Americans. However, the hypotheses based on this literature are nevertheless likely to be valid for this particular study because the faces used to prime "Asian American" in the experiment appear to be of East and/or Southeast Asian origin rather than South Asian origin. This is likely because the faces from the Chicago Face Database were selected based on racial prototypicality as rated by an American sample (Ma et al., 2015). Thus, there should be no disconnect between the stereotypes associated with Asian Americans in general (and Asian American women in particular) and those associated with the faces presented in the Asian prime condition.

generally stereotyped as low in warmth and unsociable, especially towards the dominant group (i.e., White Americans; Lin et al., 2005).

The model minority image contradicts the perpetual foreigner image to a certain degree. As the model racial minority group in the United States, Asians and Asian Americans are touted as a success story: They have high educational achievement, are economically successful, have assimilated into mainstream American (i.e., White) culture, and no longer are subject to racial discrimination (Li, 2014; Lee & Hong, 2020). Asian Americans are thus stereotyped as competent, hardworking, intelligent, ambitious, and self-disciplined (Zou & Cheryan, 2017; Bu & Borgida, 2021). It is easy to see that this image is in stark contrast to the stereotypes associated with Black Americans; in fact, the model minority stereotype emerged during the Civil Rights Movement, providing a direct contrast to Black Americans who, according to this stereotype, could be just as successful as Asian Americans if only they stopped complaining about discrimination and worked harder (Li, 2014; Lee & Hong, 2020).

These two images of Asians and Asian Americans suggest that Asian American women might be subtyped as nontraditional, given both their foreignness and lack of social skills as well as their competence and achievement. However, just as research has illuminated the association people have with *Black* and *masculinity*, it has also demonstrated that people tend to associate *Asian* with *femininity* (Johnson et al., 2011; Galinsky et al., 2013). In particular, the general Asian stereotype and the general woman stereotype do share traits such as shy, family-oriented, and soft-spoken (Johnson et al., 2011), suggesting Asian women may actually be seen as traditional rather than nontraditional. Furthermore, researchers have suggested that the two major images

previously discussed can play out differently depending on gender (e.g., their perceived foreignness results in Asian women being exoticized; Lee & Hong, 2020) and that these broad images fail to capture sexual stereotypes that are disproportionately directed at Asian American women (Yoo et al., 2010). It is these sexual stereotypes, in addition to role-related stereotypes, that may hold the key to understanding how Asian American women are subtyped according to the traditional/nontraditional dichotomy underlying AS.

Based on stereotypes specific to their group, both sexual and otherwise, Asian American women seem especially likely to be categorized as traditional women and thus may be especially likely to elicit more BS and less HS compared to other groups of women. Regarding sexual stereotypes, Asian women have historically been stereotyped in two ways: as either the Dragon Lady or the Lotus Blossom (also known as the Geisha Girl or China Doll; Prasso, 2005; Li, 2014). Both images emerged during the 19th century, when many of the Asian women who were allowed to immigrate to the United States were prostitutes. The Dragon Lady is similar to the Jezebel stereotype of Black women: In terms of her sexuality, she is dominant, aggressive, manipulative, and predatory – a clear threat to men (Li, 2014). This image became popular during the early days of Hollywood and is still seen in media today (Prasso, 2005), but more common now is the Lotus Blossom image. The Lotus Blossom is an ideal wife or partner, docile and obedient (Prasso, 2005). She is the epitome of Western male sexual fantasy: ultra-feminine, passive, domesticated, meek, innocent, and naïve (Li, 2014). Interestingly, this image solidified during the mid-20th century as second-wave feminism took hold in the United States and White women were seen as overstepping their bounds, becoming more

independent and career-focused while abandoning their roles as wives and mothers (Li, 2014). Just as Asian Americans were seen as a contrast to Black Americans during the Civil Rights Movement, Asian American women have been seen as a contrast to White women during feminist movements. Asian American women – or, more specifically, the stereotyped image of them – came to represent the ideal of what a woman should be: Deriving joy and meaning from serving the men in her life (Li, 2014; Mukkamala and Suyemoto, 2018). If the Lotus Blossom is indeed the dominant image of Asian American women today, they should be subtyped as traditional women and elicit more BS and less HS than even White women.

Research that has examined the stereotyped traits associated with Asian American women supports this argument. Niemann et al. (1994) found that Asian American women were frequently ascribed traits such as shy, passive, pleasant, and speak softly. According to Ghavami and Peplau's (2013) work on intersectional stereotypes, attributes for the 'Asian American women' group shared some overlap with attributes for the general 'women,' group, although the overlap was not great (about 11%). However, one of these shared attributes – submissive – may be especially likely to cue a BS response. Regarding unique attributes for Asian American women, these too seem likely to elicit BS and depress HS due to their links to the traditional subtype and include traits such as shy, family-oriented and quiet.

Furthermore, research that has examined Asian American women's own experiences with gendered racism and racialized sexism indicates that they are stereotyped in ways that align with the traditional subtype. For example, Pyke and Johnson (2003) interviewed Asian American women, who reported facing expectations

that they should be quiet, passive, and yielding, especially to the men in their lives. Similarly, Mukkamala and Suyemoto (2018) found that Asian American women frequently reported dealing with assumptions that they are submissive and passive, and are often treated as invisible by others.

In sum, despite general stereotypes of Asian Americans suggesting that Asian American women might be seen as nontraditional, specific stereotypes about this group – especially sexual stereotypes – indicate that they are instead more likely to be seen as traditional women. As a result, they are likely to elicit more BS and less HS than White, Black and Latina women. Based on the evidence reviewed here, the following hypotheses are put forward.

*Hypothesis 2.3a.* Compared to the White prime condition, Black prime condition, Latinx prime condition, and the control condition, the Asian prime condition should elicit lower HS scores.

*Hypothesis 2.3b.* Compared to the White prime condition, Black prime condition, Latinx prime condition, and the control condition, the Asian prime condition should elicit higher BS scores.

**Latinx American Women.**[22] The final group examined in Study 2 was Latinx Americans. As with Asian Americans, Latinx Americans are an incredibly diverse, pan-ethnic group but are nevertheless treated as monolithic by many in the United States and subject to similar stereotypes despite their varying ethnic backgrounds (Cauce &

[22] There is debate about the use of the term 'Latinx' in place of 'Latinos/as' to refer to Latin Americans of all genders. This paper follows Harris et al.'s (2020) use of the term 'Latinx' in this gender-neutral sense, the use of the term 'Latinos' to refer to men specifically, and the use of the term 'Latinas' to refer to women specifically.

Domenech-Rodriguez, 2002). An additional complication for research on stereotypes of Latinx Americans is the fact that, especially today, Latinx Americans are conflated with immigrants, especially unauthorized immigrants (Roman, 2000; Ghavami & Peplau, 2013; Zou & Cheryan, 2017; Harris et al., 2020). Thus, Latinx Americans today – regardless of their immigration status – are stereotyped similarly to major immigrant groups of the past: as "uncivilized, uneducated, poor, dirty, immoral, prone to crime, and un-American" (Bellovary et al., 2020, p. 149). These links between Latinx Americans, foreignness, and low status (Zou & Cheryan, 2017; Harris et al., 2020) have important implications for how Latina women in particular may be subtyped according to the traditional/nontraditional dichotomy.

As with Asian American women, sexual stereotypes about Latina Americans are contradictory, but one set of stereotypes seems more likely to be endorsed by others and thus influence their application of HS and BS to Latinas. Underlying sexual stereotypes of Latinas is a "virgin-whore" dichotomy (Arrizón, 2008; Harris et al., 2020): Latinas are either chaste and innocent or hypersexualized. The latter image is more common in contemporary media and is also called the "hot Latina" stereotype (Merskin, 2007; Harris et al., 2020). The hot Latina is hot-tempered and sexually aggressive, even predatory, as well as promiscuous (Merskin, 2007). Indeed, this image of Latina women is very similar to both the Dragon Lady and the Jezebel, directed at Asian women and Black women, respectively. She is also irresponsible and lazy, using her sex appeal to manipulate men to do what she wants (Harris et al., 2020).

Also similar to Black women, Latina Americans' motherhood is not celebrated. Instead, due to their association with immigration and the decline of the White majority

in the United States, Latinas and their reproduction are seen as a threat because they are thought to be highly fertile and have too many children (Chavez, 2004). Thus, even a Latina woman who might otherwise follow the virgin image and fulfill her role as a devoted wife and mother would be viewed negatively due to the threat context surrounding her motherhood. Like Black women, Latinas are often assumed to be single mothers who started having children at young ages and by multiple fathers (Rosenthal et al., 2020). Latina mothers are also assumed to be receiving public assistance and manipulating the welfare system, similar to the Welfare Queen image of Black women (Rosenthal et al., 2020). Moreover, Latina mothers are stereotyped as having children for the 'wrong' reasons, i.e., to gain material benefits such as welfare or even citizenship by giving birth to so-called anchor babies (Lòpez & Chesney-Lind, 2014; Villarreal, 2020). Thus, a Latina mother is not truly a mother in the traditional sense, but rather a schemer who uses her motherhood for nefarious purposes.

Additional research on Latina stereotypes further supports these negative images that are likely to elicit nontraditional subtyping than traditional subtyping. Niemann and colleagues (1994) reported that participants in their study frequently ascribed traits including lower class, promiscuous, baby makers, and family-oriented to Mexican American women. Per Ghavami and Peplau's (2013) intersectional stereotype research, there was very little overlap in attributes generated for 'Latina women' and attributes generated for 'women.' Specifically, only one attribute was shared between the two groups: attractive. Several of the unique traits generated for the Latina group seem to be more in line with the nontraditional subtype than the traditional subtype, including traits such as loud, promiscuous, and sexy. Furthermore, two unique attributes for this group

were ‘early motherhood’ and ‘have many children’ – again giving the impression that Latina motherhood is not viewed favorably, but is instead linked to promiscuity.

In sum, while there are some positive stereotypes of Latina Americans, the predominant imagery paints them in a negative light, most notably regarding their sexuality as well as their motherhood. As such, they should be subtyped similarly to Black women: as nontraditional women. Therefore, the following hypotheses were put forward:

*Hypothesis 2.4a.* Compared to the White prime condition and the control condition, the Latinx prime condition should elicit greater HS scores.

*Hypothesis 2.4b.* Compared to the White prime condition and the control condition, the Latinx prime condition should elicit lower BS scores.

***Respondent Race-by-Gender Identity and Target Race***

Unlike in McMahon and Kahn’s (2016) studies on how target race may affect applications of HS and BS, the present study had large enough sample sizes of White as well as Black participants to test how respondent race-by-gender group and target race might interact to affect application of HS and BS. No specific hypotheses were set regarding this potential interaction, as the dearth of literature on this topic hinders the ability to make informed hypotheses. Although, as reviewed in the previous section, women in general are stereotyped as if they are White women, this does not necessarily mean that all individuals view a generic woman as White. Instead, because of the lack of racial diversity in most past research, it may be more accurate to claim that White people will view a generic woman as White and it is unknown how people of color will view a generic woman. Weitz and Gordon (1993) have suggested that given a race-neutral

stimulus, most individuals will think of a target that matches their own racial identity. If true, we should then see similarities between the Black prime condition and the control condition in terms of BS and HS endorsement among Black participants. However, Weitz and Gordon's claim has not been sufficiently investigated. It is possible that even Black participants still consider the default woman to be White because they are aware of or perhaps even internalize the cultural links between White and femininity and Black and masculinity.

Furthermore, perceiver race – in addition to gender – may be important to consider if there are indeed significant baseline differences in HS and BS endorsement between Black and White participants. Once again, however, due to the lack of research on racial differences in HS and BS endorsement, specific hypotheses were put forth for this study. (Study 1 was unable to provide insight into this possibility due to the lack of MI for the ASI among the four race-by-gender groups in the study sample.)

**Method**

Study 2 consisted of two phases, each completed by the same sample of participants. Each phase of the study was published on Prolific, an online research platform, four times, once for each participant group of interest: Black men, Black women, White men and White women. In this way, relatively equal numbers of participants could be recruited for each group (i.e., through quota-sampling, discussed below), resolving one of the major limitations of Study 1 (i.e., large differences in *N*s across race-by-gender groups). In Phase 1, participants responded to the ASI and other individual-difference measures. In Phase 2, an experiment was conducted. Each of the four race-by-gender-groups responded to the same survey at both Phase 1 and Phase 2.

The survey and data collection strategy for both phases were approved by the Institutional Review Board of the University of Minnesota (Study ID: 00017824). At the beginning of each phase, prior to starting the survey, all respondents provided informed consent.

### *Recruitment and Final Sample*

Participants were recruited and compensated via Prolific (www.prolific.com), which is generally regarded as providing higher-quality data compared to other online research platforms, including Amazon's Mechanical Turk (e.g., Peer et al., 2022; Albert & Smilek, 2023; Douglas et al., 2023).

In order to ensure relatively equal numbers of Black and White participants, quota-sampling was used (a technique that has been recommended for researchers who want to do intersectional research; Else-Quest & Hyde, 2016b). Prolific allows researchers to pre-screen participants based on certain demographic features, including race, so that certain demographics can essentially be oversampled. For Phase 1, four versions of the survey were published on Prolific, each with different eligibility criteria based on race[23] and sex.[24] Participation in each survey was open to all eligible Prolific users, based on screening information provided to Prolific.

---

[23] Prolific has two screeners for ethnicity and none for race. However, the "Ethnicity (Simplified)" screener, used for Study 2 recruitment, corresponds to the American Psychological Association's (2022) conception of race and includes the following response options: *White*, *Black*, *Asian*, *Mixed*, and *Other*. In contrast, the "Ethnicity" screener corresponds to the APA's conception of ethnicity, noting that "peoples' ethnicity describes their feeling of belonging and attachment to a distinct group of a larger population that shares their ancestry, color, language or religion." Response options for this latter screener are numerous, including categories such as *African*, *Caribbean*, *East Asian*, *South Asian* and so on. Thus, for the purposes of this study, the "Ethnicity (Simplified)" screener is seen as referring to race rather than ethnicity.

[24] For the "Sex" screener, participants were asked "What is your sex, as recorded on legal/official documents?" Response options included *Male* and *Female*.

It can be difficult to determine appropriate sample sizes when conducting MI testing because statistical power depends on not only sample sizes but also model complexity and other model characteristics; thus, there is no general 'rule' regarding how many participants should be included in this kind of analysis (Kline, 2016; Meade & Bauer, 2007). Simulation studies suggest that sample sizes below 200 rarely have adequate power for detecting MI, while sample sizes over 400 tend to have high power (Meade & Bauer, 2007). The conventional rule of thumb for CFA/SEM analysis is 20 cases per measured indicator (Kyriazos, 2018) – for the ASI, which has 22 items, this suggests a sample size of 440 per group. Based on all of these considerations, we aimed to recruit 500 participants per race-by-gender group at Phase 1 to account for expected attrition at Phase 2. This goal was nearly met for three of the four groups, but only 464 Black men were able to be recruited at Phase 1.

The final sample consisted of participants who completed both Phase 1 and Phase 2 of Study 2, and responded to all ASI items at both times. Retention rates from Phase 1 to Phase 2 were satisfactory, but did vary across groups: 89.4% for Black men, 83.9% for Black women, 94.9% for White men, and 94.4% for White women. Thus, group *N*s were not equal in the final sample, but the differences were much smaller than they were in Study 1 and, critically, the numbers of Black men and Black women in Study 2 were much larger than in Study 1.

Additionally, participants who could not be confirmed as cisgender[25] were eliminated from the final sample. AS theory is inherently based on a biological/binary

[25] Per the American Psychological Association (2018), the term *cisgender* is "used to describe an individual whose gender identity and gender expression align with the sex assigned at birth" – for example, an individual who is assigned female at birth and identifies as a woman is cisgender. In contrast, the term

view of sex and does not distinguish between sex and gender (this is especially true of the early research, conducted during a time when mainstream understanding of the distinction was less widespread than it is today[26]). It is unclear how transgender or non-binary individuals 'fit' into the framework, particularly given the role that sexual reproduction plays in the dynamics outlined by AS theory. Therefore, it is appropriate, at this time, to focus on cisgender participants only, as additional research is needed to fully flesh out how non-cisgender identities intersect with AS rather than assume non-cisgender individuals think, behave, and feel the same way as cisgender individuals. At the end of Phase 1, participants were asked "What is your gender?", with response options including *Man*, *Woman*, *Prefer to self-describe*, and *Do not wish to specify*. Participants were considered cisgender if their reported gender matched their reported sex (i.e., Man and Male, Woman and Female). Participants who indicated they were transgender (i.e., their sex and gender did not match [e.g., Man and Female, Woman and Male] or self-described as transgender) were excluded from the final sample, as well as participants who could not be confirmed as cisgender (i.e., they did not report their gender). Overall, 33 participants were eliminated from the final sample based on these criteria.[27] It is possible that, even with this screening process, cisgender participants were

---

*transgender* is "an umbrella term encompassing those whose gender identities or gender roles differ from those typically associated with the sex they were assigned at birth" – for example, an individual who is assigned male at birth but identifies as a woman is transgender.

[26] There is still great inconsistency in how many psychological researchers use the terms *sex* and *gender*. Within the domain of research on women and gender, however, it is now generally agreed that *sex* is based on biological characteristics (e.g., chromosomes, genitalia) whereas *gender* is based on social and cultural characteristics (e.g., attitudes, behaviors). Nevertheless, despite being distinct concepts, the two are often closely intertwined (Hyde et al., 2019).

[27] Of these 33 participants, 3 were from the Black men group, 11 from the Black women group, 1 from the White men group, and 18 from the White women group.

excluded and/or non-cisgender participants were included, but the number for either group should be quite small and not influence results.

The final sample consisted of 1,733 participants, including 412 Black men, 401 Black women, 467 men, and 453 White women. The final *N*s for Black men and Black women were below the initial goal of 440 participants, but still much larger than those in Study 1. Additional demographic information is provided in Tables 17 through 21. As reported in Table 17, White women were, on average, the oldest participants in the sample whereas Black men were the youngest. As in Study 1, a majority of participants in Study 2 (85%) had at least some college education and very few (1%) had no high school education (Table 18). In terms of political party affiliation, as expected large majorities of Black men and Black women identified as Democrats (Table 20). Compared to Study 1, in Study 2 there was less diversity in political party affiliation among White men and White women. Similarly, unlike in Study 1, a clear majority of the Study 2 sample considered themselves to be at least slightly liberal (57.4%), with just 22.2% considering themselves moderate and 20.4% considering themselves to be at least slightly conservative.

### *Phase 1: Measures and Procedure*

The Phase 1 survey was published on Prolific on May 2, 2023. Phase 1 remained open until the quota of 500 participants was met or it was determined that no more participants could be recruited. The surveys for White men and White women closed the same day they opened, on May 2, 2023; the survey for Black women was closed on May 10, 2023; and the survey for Black men was closed on May 13, 2023. The survey

typically took no more than 10 minutes to complete and participants were paid $1.50 upon completion of the survey.

Participants who were eligible based on demographic screening were alerted to the study's availability and, after reading a brief summary of the study (which indicated they would be invited back for a second phase in approximately two weeks), could click on a link taking them directly to the survey on Qualtrics. On the first page of the survey, participants had to indicate that they consented to participate in the study in order to move on with the survey. Phase 1 remained open until the quota of 500 participants was met or it was determined that no more eligible participants were likely to take the survey.

**Prolific ID and Screening Questions.** Participants were next prompted to provide their Prolific ID number, which would enable researchers to link their Phase 1 data with their Phase 2 data. Although this was presented as a question, the Qualtrics survey was programmed such that the ID number would autofill based on the participant's URL; thus, there was no risk that participants would provide erroneous ID numbers. Next, participants responded to three questions to ensure that their nationality, sex, and race – as reported to Prolific – did make them eligible for that specific version of the study. "Incorrect" answers to any of these questions resulted in the participant being exited from the survey. For example, in order to move on to the next section of the survey, participants in the Black Men group had to select *United States*, *Male*, and *Black*; any other response meant the participant was ineligible to continue and therefore they were exited from the survey.

**Attention Check.** After confirming their eligibility, participants responded to an attention check question. They were instructed to respond to a question asking about their

favorite color by selecting the option *Purple*. All participants in the final sample provided the correct answer and no participants were excluded from analyses based on this question.

**ASI.** Next, participants were presented with the instructions for the ASI, presented in black text centered on a white background. This format matched the control condition from the experiment in Phase 2 of the study. The instruction screen is presented in full in Figure 4.

After reading the instructions, participants moved to the next screen which included the full 22-item ASI (see Study 1 – Method for a description and Appendix A for the full measure). Items were presented to each participant in a randomized order to account for potential order effects.

**Additional Measures.** Several additional individual-difference measures were included in the Phase 1 survey to serve as potential controls or covariates in the final analysis. However, for the purposes of the current analysis, only the feeling thermometers will be discussed.

***Feeling Thermometers*.** After the ASI, participants responded to a series of feeling thermometers, which are used to measure attitudes about (or feelings towards) specific groups of people. In this case, we were particularly interested in participants' attitudes towards different racial groups (i.e., Black, White, Asian and Latino/a Americans), as racial attitudes may interact with target race to influence HS and BS endorsement.[28] However, other groups based on gender, political party membership, and

[28] In their first study, McMahon and Kahn (2016) found a marginal effect for modern racism, such that participants who scored higher on the modern racism measure expressed (marginally) more HS toward the Black faces than participants who scored lower on the modern racism measure. Feeling thermometers were

age were also included in this section so as to not suggest that the researchers were specifically interested in racial attitudes. Participants first read the following instructions:

> Next, we'd like to get your feelings toward some people and groups in the news these days. On the next several pages, we will give you the name of a person or group, and we'd like you to rate that person using something we call the feeling thermometer.
>
> Ratings between 50 degrees and 100 degrees mean that you feel favorable toward the person or group.
>
> Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person or group and that you don't care too much for that person or group.
>
> You would rate the person or group at the 50 degree mark if you don't feel particularly favorable or unfavorable toward the person or group.
>
> If we come to a person or group whose name you don't recognize, you can decline to rate your feelings.

Then, participants moved through a series of 13 screens, each showing a different group to rate. Groups were presented to participants in a randomized order so as to counter potential order effects.

**Demographics.** In the final section of the Phase 1 survey, participants were asked several demographic questions. Most notably, participants were asked to report their gender – responses to this question were used in concert with the sex information from Prolific to determine if participants were cisgender. Participants were also asked about the highest level of education they have achieved, their annual household income, their political ideology, and their political party affiliation.

---

included in this study because the modern racism scale taps into attitudes about Blacks specifically and cannot directly inform about attitudes about other racial groups. Specific hypotheses were not put forth about potential interactions between respondents' racial attitudes and the racial primes.

Once these questions were answered, participants were thanked for their participation and reminded that they would be invited back for a second survey in approximately one to two weeks.

### *Phase 2: Measures and Procedure*

The Phase 2 survey was published on Prolific on May 16, 2023. Only Prolific users who completed the first survey were invited to take part in this second study phase. On May 30, 2023, messages were sent on Prolific to eligible participants who had not yet completed Phase 2. The survey closed on June 1, 2023 for White men; June 5 for White women; and June 6 for Black men and women. The overall retention rate from Phase 1 to Phase 2 was approximately 90%. As previously discussed, retention did vary across groups – 89.4% for Black men, 83.9% for Black women, 94.9% for White men, and 94.4% for White women – but all rates were acceptable. The survey typically took no more than five minutes to complete and participants were paid $1.00 upon completion of the survey.

As with the Phase 1 study, eligible participants were alerted to the study's availability on Prolific. After reading a brief summary of Phase 2, eligible participants could click on a link taking them directly to the survey on Qualtrics. Once again, on the first page of the survey, participants had to provide consent in order to move on with the survey. On the next screen, participants' Prolific ID numbers were recorded based on their URL, allowing their Phase 1 and Phase 2 data to be linked.

**Attention Check.** Participants next responded to an attention check question. They were instructed to respond to a question asking about their favorite food by

selecting the option *Apple*. All participants in the final sample provided the correct answer, and no participants were excluded from analyses based on this question.

**Experiment.** After the attention check question, Qualtrics randomly assigned participants to one of five experimental conditions (see Table 22 for number of participants per condition). The experimental manipulation involved altering the ASI instructions screen so that participants would think about men and women of a certain racial group while responding to the ASI. In other words, participants were exposed to either a control condition (no racial prime) or one of four racial primes. After reading the instructions, participants then responded to the ASI.

***Experimental Manipulations.*** For the control condition (see Figure 4), the original instructions were shown on the screen. For the remaining four conditions (i.e., the four racial primes), the original instructions were altered by adding the following statement: "As you read through these statements, we'd like you to specifically consider [*racial group*] American men and [*racial group*] American women, such as those pictured here." For each racial group manipulation – White, Black, Latino/a, and Asian – six images (three women's faces and three men's faces) were added to the instruction screen (see Figure 5) to reinforce the specific instructions.

***Photo Selection.*** All images are from the Chicago Face Database (Version 3.0) and were selected based on norming data provided by the developers (Ma et al., 2015). The database includes 597 unique individuals, all of whom provided at least one image in which they had a neutral expression. These individuals self-reported their gender and racial identities. The database developers (Ma et al., 2015) provided norming data by having independent raters rate the images on a number of dimensions. For our purposes

in choosing images for the experimental manipulation, we relied on five key norming variables: perception of gender, perception of race, feminine, masculine, and racial prototypicality. Regarding gender, we looked at the percentage of raters who categorized each image (correctly) as male or female, eliminating any images from consideration for which gender ambiguity existed. Similarly, regarding race, we looked at the percentage of raters who categorized each image (correctly) as White, Black, Asian, Latino/a and so on, and eliminated from consideration any images for which racial ambiguity existed. However, for the Latino/a images, none were rated as such by all of the raters; therefore, we selected the images with the highest percentage of agreement that the individuals were Latino/a. For female images, we considered high feminine ratings and low masculine ratings; for male images, we considered low feminine ratings and high masculine ratings. Finally, we considered each image's racial prototypicality score, selecting images with higher ratings so as to reduce the racial ambiguity of the primed images. All of these factors were considered together in choosing the final six images for each racial manipulation. (See Footnote 21 for additional context regarding the selection of the Asian faces.)

***Manipulation Check*.** After completing the ASI for the final time, participants answered a manipulation check question:

> At the beginning of this survey, you were asked to think about men and women and their relationships in contemporary society, and then respond to a series of statements.
>
> Which group of men and women were you asked to think about while reading these statements?

There were five response options: *White Americans*, *Black Americans*, *Asian Americans*, *Latino/a Americans*, and *I was not asked to think about a specific group of men and women*. Responses to this question were compared to participants' actual experimental condition: participants whose manipulation check response matched their experimental condition passed the manipulation check while those whose responses did not match failed the manipulation check.

**Debriefing.** After answering the manipulation check question, participants were thanked for their participation and fully debriefed about both phases of the study.

### *Analysis Plan*

As in Study 1**,** all statistical analyses were conducted using R Statistical Software (version 4.2.2; R Core Team, 2022) and STATA (version 15.1; StataCorp, 2017). More specifically, the *lavaan* R package (Rosseel, 2012) was used to conduct the MI analyses while the *stats* (R Core Team, 2022), *car* (Fox & Weisburg, 2019), and *rstatix* (Kassambara, 2023) packages were used to conduct mean comparison tests. Initial data cleaning and preparation were done in STATA.

**Measurement Invariance.** Before testing the Study 2 hypotheses, MI of the ASI at Time 1 was checked following the procedures outlined in Study 1 (i.e., preliminary checks followed by the formal MI analyses). Although MI should usually be checked at each time point of analysis, it was unclear if the experimental manipulation at Time 2 would affect MI at Time 2. Nevertheless, the conclusions regarding MI at Time 2 generally matched those at Time 1, and thus only the Time 1 results are reported here. Additionally, because there are varying cutoff criteria for when changes in fit statistics signal measurement variance, the approach here was twofold: a more conservative

approach which adheres to the stricter cutoffs outlined in Study 1, and a less conservative approach which adheres to the less strict cutoffs outlined in Study 1, with some additional leeway granted. Finally, if MI could not be established across all four race-by-gender groups, subsets of groups would be tested to determine if certain mean comparisons could still be made (e.g., White men and White women) and hypotheses tested.

**Study 1 Hypothesis Testing.** As will be reviewed in more detail in the Results section, MI analyses confirmed that mean comparisons could be made between two sets of groups: between Black and White women, and between White men and women. Therefore, four of the eight Study 1 hypotheses were tested (i.e., Hypotheses 1.2a, 1.2b, 1.4a, and 1.4b). The analysis plan laid out in Study 1 was followed, based on suggestions by Warner and Shields (2018). Because only two groups could be compared at once (rather than all three simultaneously), HS and BS scores were compared using independent-samples t-tests and Cohen's *d* effect sizes were also reported (each adjusted as needed to account for unequal variances).

**Study 2 Hypothesis Testing.** The eight Study 2 hypotheses were tested for each eligible group, as determined by the MI analysis (i.e., Black women, White women, and White men). Within each eligible group, means for each experimental condition were compared via one-way ANOVAs, in which experimental condition served as the grouping variable. Analyses were adjusted as needed to account for any violations of the equal variances assumption. Further, if an ANOVA was significant, post-hoc tests (e.g., Tukey's HSD or Games-Howell in the case of unequal variances) were conducted as appropriate to determine exactly which means significantly differed from one another.

Taken together, the results from analyses for each of the three groups determined whether or not the Study 2 hypotheses were supported.

**Exploratory Analyses.** Exploratory analyses focused on two additional avenues of inquiry. First, the potential interaction between respondents' race and the racial prime was tested for each eligible group comparison. A series of two-way ANOVAs was conducted – using the *Anova* function in R's '*car*' package and accounting for unequal variances as necessary – to determine if race-by-gender group and racial prime had any effects on HS and BS scores, as well as whether an interaction between the two existed at Time 2.

Second, the role of racial attitudes, as measured by feeling thermometers, was also examined. Means scores on each of the four racial group feeling thermometers were compared across the eligible groups to determine if certain participants held more positive or negative attitudes about specific racial groups compared to other participants. If all eligible participant groups hold similar attitudes about the four racial groups, there would be little expectation that these attitudes would differentially affect HS and BS scores. Multiple approaches were used to examine the potential main effects of specific racial attitudes as well as potential interactions between these attitudes and experimental condition, and primarily relied on OLS regression analyses.

## Results

### *Manipulation Check*

Participants' responses to the manipulation check item were matched with their assigned experimental condition at Time 2 to determine whether they failed or passed the check. Of the original 1,733 participants (i.e., all race-by-gender groups), 291 (16.8%)

failed the manipulation check. Of these 291 participants, 36.4% were Black men (see Table 23 for full breakdown of manipulation check failures by condition and race-by-gender group). Moreover, 25.7% of Black men in the sample failed the manipulation check. These numbers are in sharp contrast to those for the other three race-by-gender groups, as reported in Table 23. Additionally, 16.2% of Black women, 13.3% of White men, and 12.8% of White women failed the manipulation check. It is not known why these discrepancies exist, but Black men were excluded from the primary and exploratory analyses due to the failure to establish MI for this group (see next section for additional details). As such, the discrepancies in manipulation check failures were not explored further in this analysis.

Regarding experimental condition, participants in the control condition were the least likely to fail the manipulation check, making up just 6.5% of manipulation check failures. Thus, participants who did not see a face were very unlikely to report that they did see one. Participants in the White condition made up 29.9% of the manipulation check failures, followed by those in the Latinx condition (25.8%), those in the Asian condition (23.4%), and finally by those in the Black condition (14.4%).

With the exception of the MI analyses – which used Time 1 data only – all analyses were conducted twice: first with the full sample and again with only those participants who passed the manipulation check. While there were some differences in the results – for example, an effect's *p*-value changing from marginally significant to non-significant – the overall conclusions (e.g., whether or not a hypothesis was supported) did not change when the sample was reduced. Therefore, the analyses reported here used the full sample of participants, including those who failed the manipulation

check. Results with the reduced sample (i.e., participants who passed the manipulation check) are noted in footnotes only if they substantially differed from results with the full sample.

***Measurement Invariance***

**Initial Steps.** Cronbach's alpha ($\alpha$) coefficients for each scale and subscale of the ASI at Time 1[29] are reported by group in Table 24, as well as the average coefficients reported by Glick and Fiske (1996) in their scale development paper for the AS, HS, And BS scales (see Footnote 12 for additional information). The findings here are similar to those in Study 1: Many of the coefficients are similar to those reported by Glick and Fiske and meet typical levels of acceptance (i.e., $\alpha \geq .70$); however, we still see a few coefficients that fall below this threshold. In particular, the protective paternalism subscale has an alpha of .65 for Black men and .64 for Black women, suggesting that there may be a problem with these items for these groups, as there was in Study 1. In contrast to Study 1, however, the heterosexual intimacy subscale performs much better for these groups, meeting the acceptable cutoff of $\alpha \geq .70$. Although there were some signs that caution is needed moving forward, any potential issues seem like they might stem from a single BS subscale rather than two as in Study 1.

Next, the HS-BS Pearson correlation coefficient ($r$) was checked for each group and compared to the range of $r$ = .40-.50 commonly reported in AS research (e.g., Glick & Fiske, 1996, 1997, 2011). At both Time and Time 2, the correlations for Black women ($r$ = .42 and $r$ = .43) and White women ($r$ = .57 and $r$ = .52) fell within the expected

[29] Coefficients for the (sub)scales at Time 2 are similar to those at Time 1 and thus are not reported here. The similarities are not surprising given that the Time 1 – Time 2 correlations for both HS and BS were quite high (i.e., all $\alpha$s $\geq .85$), suggesting substantial consistency between ASI responses at each timepoint.

range, with the latter group's correlations a bit higher than expected. The correlations for White men ($r$ = .34 and $r$ = .31) and Black men ($r$ = .23 and $r$ = .19) fell outside of this range, and were especially small for the latter group compared to the other three. These findings differ slightly from those in Study 1. In Study 1, the HS-BS correlation for Black women was just .16, much lower than either correlation in Study 2. In contrast, the correlations for White men in Study 2 were lower than the correlation in Study 1, which fell within the typical range ($r$ = .41). Nevertheless, the low(er) HS-BS correlations for Black participants in both studies are consistent with the findings reported by Hayes and Swim (2013) for HS and BS. As noted in Study 1, unexpectedly low (or high) HS-BS correlations may be a sign of measurement noninvariance, but may also simply reflect different endorsement patterns among these groups.

The final step of the preliminary checks was to check the fit of different models (i.e., one-factor, two-factor, and full model) for each of the race-by-gender groups. *As in Study 1, the full model proved to best fit the data for each group, followed by the two-factor model and then the one-factor model.* There was some variation in fit statistics for the full model for each race-by-gender group (Table 25), but not nearly as much as there was in Study 1 (Table 7). The only group for which fit statistics were not adequate across the board was Black men: The CFI for this group was .871, below the acceptable cutoff of .90. Nevertheless, this was much better than the CFI for this group in Study 1, which was .757, and the Study 2 RMSEA and SRMR are within the range of acceptable fit.

Taken altogether, the reliability, correlational, and full model fit analyses do not signal possible measurement noninvariance to the same extent as they did in Study 1.

While there may still be some cause for concern, overall these preliminary analyses suggest that MI may hold across the four groups within this particular sample.

**Formal MI Analysis.** Formal MI analyses were conducted on the ASI at both Time 1 and Time 2. However, as previously noted, the conclusions drawn are substantively the same at both timepoints; therefore, only the Time 1 results are reported here, with the corresponding tables for the Time 2 results reported in the Supplemental Materials (Tables S6-S23).

Following the procedure outlined in Study 1, in Study 2, five models were fit and compared to determine if MI of the ASI exists across the four race-by-gender groups of interest. The fit indices for each model – configural (M1), first-order metric (M2), second-order metric (M3), first-order scalar (M4), and second-order scalar (M5) – are reported in Table 26. What is immediately clear is that configural invariance *can* be confirmed with this sample: the fit statistics for the configural model all reach acceptable levels, suggesting good fit for this model. This is in stark contrast to the Study 1 findings, which clearly indicated a lack of configural invariance using the original 22-item ASI. Additionally, in Study 1, Black men and Black women each had four items fail to load on the assigned factor (i.e., they were nonsignificant); in Study 2, no items were nonsignificant for any groups at Time 1, and just one item was nonsignificant for Black men at Time 2 (hs10; see Table 9 for full wording of this item). This suggests that the reverse-scored items generally were *not* as problematic for this sample as they were for the Study 1 sample, an issue that will be examined further in Chapter 4.

Continuing the MI analysis, the chi-square test for each model was significant (see Table 26); however, this may be the result of the large sample size rather than poor

fit, especially considering the fit statistics for the first three models, all of which signal acceptable-to-good fit. Starting with the first-order scalar model, model fit declines from acceptable based on the CFI (less than .90) and the SRMR (greater than .80; additional details on these cutoffs are provided in the Method section of Chapter 2). Model fit further deteriorates for the final model, the second-order scalar model. These numbers suggest that we cannot confirm first-order scalar invariance, a conclusion supported by an examination of the changes in fit statistics across the five models.

Table 27 reports the changes in fit statistics between the first and second model, second and third model, and so on, as well as the results of the chi-square test comparing the two models. Again, all chi-square tests were significant, but this may be an artifact of the large sample size. Using the more conservative cutoffs for changes in fit (see Table 28; additional details are offered in the Method section of Study 1), the change in CFI between the configural and first-order metric model is unacceptable at .008. However, this change *is* acceptable under the less conservative cutoff, and thus, given the acceptable (by both standards) changes in RMSEA and SRMR, we can infer that M2 fits just as well as M1. Moving on, in line with the model fit statistics reported in Table 27, the less conservative approach does indicate that the changes in CFI from M3 to M4 and from M4 to M5 are unacceptable, as well as the change in SRMR from M4 to M5. Therefore, we must conclude that, even with a less conservative approach, we cannot confirm second-order scalar MI of the ASI across the four-race-by-gender groups, meaning four-way group comparisons cannot be made.

**Initial Responses to Noninvariance.** As noted in the Method section, in the face of noninvariance between the four race-by-gender groups of interest, subsets of groups

would be checked for MI to determine if certain mean comparisons could be made and certain hypotheses tested. A revised model, in which invariant items are dropped, was not tested here due to the issues discussed in Chapter 2 and again in Chapter 4 – i.e., the inability to empirically determine whether the items are invariant due to wording or content.

***Black Men.*** Given that, of the four race-by-gender groups, Black men had the poorest fit statistics for the full model, it seemed possible that this group was responsible for the lack of scalar invariance in the initial analysis. MI analyses were conducted for three comparisons involving this group: Black men and Black women (Tables 29 and 30); Black men and White men (Tables 31 and 32); and Black men and White women (Tables 33 and 34). For each of these comparisons, second-order scalar invariance cannot be confirmed, thus making mean comparisons with the Black men group inappropriate for this analysis. In general, the fit statistics for the Black men vs. White men comparison are better than those for the other two comparisons, but they still fail to confirm full MI.

***Black Women vs. White Men vs. White Women.*** Next, MI was tested to determine if three-way comparisons could be made between the remaining three race-by-gender groups: Black women, White men, and White women. The fit statistics (Table 35) are generally acceptable for four of the five invariance models, but poor fit emerges with M5, the second-order metric model. Although the CFI and SRMR for this model are outside of the range for acceptable fit, the numbers are close – but how close to the already conservative cutoff (e.g., the less conservative cutoff for acceptable CFI is .90, the more conservative cutoff is .95) is acceptable? The lack of guidance on this subject – especially for a complex measurement model such as the ASI has – is an issue addressed

further in the Discussion section. The changes in fit statistics (Table 36) between invariance models suggest that full MI cannot be confidently confirmed, as the CFI change between M3 and M4 and the SRMR change between M4 and M5 are above the less conservative cutoffs for acceptability. Thus, we next looked at pairs of groups that might have MI between them.

***Black Women vs. White Women.*** If Black women and White women are similar in their approach to the ASI due to their shared gender, MI between these two groups is a possibility. Each of the five invariance models was simultaneously fit for these two groups. The fit statistics for each invariance model are reported in Table 37. Aside from the significant chi-square tests (which are likely the result of the large sample size), the only signs of poor fit are the CFIs for M4 and M5; however, these are just slightly below the cutoff for acceptable fit (.90) and the RMSEA and SRMR indicate no problems with fit. Additionally, the changes in fit statistics (Table 38) are not particularly alarming, as just one – the change in CFI from M3 to M4 – is unacceptable by the more conservative cutoff standard. Considering both sets of findings together, and using a less conservative approach, full MI of the ASI for Black women and White women is confirmed in this sample. However, it is important to note that a stricter approach to the process (e.g., using Chen's (2007) more conservative cutoffs), would suggest halting analyses here.

***White Men vs. White Women.*** The next pair of groups considered was White men and White men, who may share a similar approach to the ASI due to their race. As before, the five invariance models were simultaneously fit for these two groups. The chi-square test results and fit statistics for each group are reported in Table 39. All chi-square

tests are significant, but this is likely the result of large sample size rather than noninvariance. Furthermore, all test statistics for each of the five invariance models suggest acceptable-to-good fit. Additionally, the changes in these statistics from model to model are generally acceptable (Table 40). Two exceptions exist: the change in CFI from M3 to M4 is unacceptable by the more conservative range, whereas the change in SRMR from M4 to M5 is unacceptable by the less conservative range. Because the SRMR itself for M5 is acceptable (although right at the cutoff for acceptable fit), it may be appropriate to tentatively confirm full MI for White men and women in this sample. However, it must be noted that a more conservative approach would likely have discontinued analyses at this point.

***Black Women vs. White Men.*** The final pair of groups to check are Black women and White men. These groups have neither gender nor race in common, thus they may approach the ASI differently to the point of disrupting MI. As reported in Table 41, the fit statistics generally indicate acceptable fit up until the fifth model, i.e., the second-order scalar model, for which the CFI and SRMR are unacceptable albeit close to the traditional cutoffs for acceptable fit. There are also some changes in fit statistics (Table 42) that suggest caution regarding confirmation of MI: the change in CFI is unacceptable, even by the less conservative cutoff, between M3 and M4; additionally, the changes in CFI and SRMR between M4 and M5 are both unacceptable based on the more conservative cutoffs. Taken altogether, these findings suggest caution regarding MI between Black women and White men in this sample, and it cannot be confidently affirmed. Thus, Black women and White men are not directly compared in subsequent analyses.

***Study 1 Hypothesis Testing***

Based on the MI analyses, mean comparisons were made between two pairs of race-by-gender groups: between White men and White women, and between Black women and White women. Therefore, four of the eight hypotheses from Study 1 were tested using Study 2 (Time 1) data. Time 2 data were not considered because of the potential effect(s) that the experimental manipulations may have had on participants' responses – as issue examined separately in the Study 2 Hypothesis Testing section. Even so, for each of the three groups the correlations between Time 1 HS and Time 2 HS, as well as the correlations between Time 1 BS and Time 2 BS, are all quite high (i.e., all *r*s $\geq .85$), suggesting that results at Time 2 would be similar to those at Time 1 because of the consistency in scores.

Because just two means were being compared at a time, the *t.test* function in R's *stats* package was used to test each hypothesis. This function assumes that the variances of the two groups are unequal, but this assumption was also tested via Levene's test (using the *leveneTest* function in R's *car* package) and the *t.test* syntax was adjusted to indicate equal variances if appropriate. The function *cohens_d* (found in R's *rstatix* package) was used to calculate the effect size of a significant difference in means, with syntax adjusted as necessary to account for unequal variances. Group means and standard deviations for each group on HS and BS at Time 1 are reported in Table 43 and depicted visually in Figure 6.

**Study 1 – Hypothesis 1.2a.** Hypothesis 1.2a from Study 1 predicted that White men would score higher than White women on HS. Levene's test indicated that the variances of the two groups on this measure were unequal ($p < .001$). A one-tailed t-test

indicated that White men ($M$ = 2.77, $SD$ = 1.27) scored significantly higher than White women ($M$ = 2.22, $SD$ = 1.02) on HS at Time 1, $t$(886.14) = 7.16, $p$ < .001. Thus, Hypothesis 1.2a was supported. According to Cohen's (1988) standards, this was a moderate effect, $d$ = .47. However, it is important to note that each group mean is below the midpoint of the scale (3.5), indicating that while a moderate difference does exist, both groups, on average, are not hostile sexists, as they tend to disagree rather than agree with the HS statements. Thus, the statistical difference is not necessarily a meaningful one.

**Study 1 – Hypothesis 1.2b.** Hypothesis 1.2b from Study 1 predicted that White men would score higher than White women on BS, but that this would be a smaller difference than that for HS. Levene's test indicated that the variances of the two groups on this measure were equal ($p$ = .628). A one-tailed t-test indicated that White men ($M$ = 3.20, $SD$ = 1.07) scored significantly higher than White women ($M$ = 2.81, $SD$ = 1.04) on BS at Time 1, $t$(918) = 5.69, $p$ < .001. Additionally, the difference between the two groups was smaller for BS (0.39) than for HS (0.55). Thus, both parts of Hypothesis 1.2b were supported. Once again, this difference was moderate according to Cohen's (1988) standards, with an effect size of $d$ = .38. Furthermore, as with the difference between these two groups on HS, each group mean for BS is below the midpoint of the scale (3.5), indicating that while a difference does exist, both groups, on average, are not benevolent sexists, as they tend to disagree rather than agree with the BS statements.

**Study 1 – Hypothesis 1.4a.** Hypothesis 1.4a from Study 1 predicted that White women would score higher than Black women on HS. Levene's test indicated that the variances of the two groups on this measure were equal ($p$ = .878). A one-tailed t-test

indicated that White women ($M$ = 2.22, $SD$ = 1.02) did not score significantly higher than Black women ($M$ = 2.38, $SD$ = 1.01) on HS at Time 1, $t(852) = 2.30$, $p = .989$. Thus, Hypothesis 1.4a was not supported. In fact, according to a two-tailed t-test, the difference in means *is* significant, but in the opposite direction than predicted: Black women scored significantly higher on HS than White women, $t(852) = 2.30$, $p = .021$.[30] However, this difference is small (Cohen's $d = .16$) and neither group should be considered hostile sexists, given their low their mean scores.

**Study 1 – Hypothesis 1.4b.** Hypothesis 1.4b from Study 1 predicted that White women would score higher than Black women on BS. Levene's test indicated that the variances of the two groups on this measure were not equal ($p = .013$). A one-tailed t-test indicated that White women ($M$ = 2.81, $SD$ = 1.04) did not score significantly higher than Black women ($M$ = 3.55, $SD$ = 0.95) on HS at Time 1, $t(850.93) = 10.91$, $p = .013$. Thus, Hypothesis 1.4b was not supported. As with the difference in HS means, a two-tailed t-test indicated that the difference in BS means *is* significant, but in the opposite direction than predicted: Black women scored significantly higher on BS than White women, $t(850.93) = 10.91$, $p < .001$. According to Cohen's standards, this is a large effect ($d$ = 0.74). Additionally, given that the group mean for Black women is above the scale midpoint of 3.5 they might be considered benevolently sexist whereas White women clearly are not, at least on average. However, the mean for Black women is likely too close to the scale midpoint to confidently draw this conclusion; instead, replications would be needed to further elucidate this finding.

[30] When participants who failed the manipulation check at Time 2 were excluded from analysis, the $p$-value drops from significant to marginally significant ($p = .096$).

**Summary.** In sum, in the Study 2 sample, two Study 1 hypotheses were supported (Hypotheses 1.2a and 1.2b) whereas two were not supported (Hypotheses 1.4a and 1.4b). Regarding the latter two hypotheses, the findings actually supported the opposite conclusions: that Black women scored higher on both HS and BS than White women. Overall, group means were all below the scale midpoints (with one exception, discussed above), suggesting that while differences did exist, they were not particularly meaningful as this sample generally was not sexist, in either a hostile or benevolent sense.

### *Study 2 Hypothesis Testing*

The primary hypotheses for Study 2 concerned the effect of the experimental prime on HS and BS scores. As no predictions were made about the interaction between participants' own racial identity and the racial primes, each hypothesis was tested independently within each of the three eligible groups: Black women, White women, and White men. That potential interaction is explored in a later section in the Results.

For each group, one-way ANOVAs were conducted to determine if experimental condition (i.e., racial prime) had any effect on the group's HS and BS. Levene's tests for all comparisons were non-significant, thus no adjustments were made for unequal variances. As in Study 1, the *oneway.test* function (found in R's *rstatix* package) was used to conduct analyses.

Means and standard deviations for each group on HS and BS by experimental condition (at Time 2) are reported in Tables 44 and 45 as well as depicted visually in Figures 7 and 8.

**Black Women.** Among Black women, there was no significant main effect of racial prime on HS, $F(4,396) = 0.78$, $p = .540$. In other words, there were no differences in HS scores between any of the five conditions for Black women. Additionally, there was no significant main effect of racial prime on BS, $F(4,396) = 2.21$, $p = .067$. This latter effect might be considered marginally significant, but Table 45 and Figure 8 suggest very little, if any, meaningful differences between BS scores across experimental conditions, given that all five means are between 3.50 and 4.00. Taken together, these results support Hypotheses 2.1a and 2.1b – which posited no differences between the White prime and control conditions on either HS or BS – but do not support any of the remaining six hypotheses, which all predicted differences between certain conditions.

**White Women.** Among White women, there was no significant main effect of racial prime on HS, $F(4,448) = 1.16$, $p = .330$. In other words, there were no differences in HS scores between any of the five conditions for White women. Additionally, there was no significant main effect of racial prime on BS, $F(4,448) = 0.96$, $p = .322$. The lack of significant differences for both HS and BS across racial primes is reflected in Tables 44 and 45, as well as in Figures 7 and 8: For HS, all experimental conditions had means between 2.00 and 2.50, whereas for BS, all experimental conditions had means between 2.50 and 3.00, suggesting an overall lack of endorsement for either form of sexism. Taken together, these results support Hypotheses 2.1a and 2.1b – which posited no differences between the White prime and control conditions on either HS or BS – but do not support any of the remaining six hypotheses, which all predicted differences between certain conditions.

**White Men.** Finally, among White men, there was no significant main effect of racial prime on HS, $F(4,462) = 0.53$, $p = .711$. In other words, there were no differences in HS scores between any of the five conditions for White men. Additionally, there was no significant main effect of racial prime on BS, $F(4,462) = 0.67$, $p = .613$. The lack of significant differences for both HS and BS across racial primes is reflected in Tables 44 and 45, as well as in Figures 7 and 8: For HS, all experimental conditions had means between 2.50 and 3.00, whereas for BS, all experimental conditions had means between 2.50 and 3.00, suggesting an overall lack of endorsement for either form of sexism. Taken together, these results support Hypotheses 2.1a and 2.1b – which posited no differences between the White prime and control conditions on either HS or BS – but do not support any of the remaining six hypotheses, which all predicted differences between certain conditions.

**Summary.** Table 46 shows whether or not each hypothesis was supported within each eligible race-by-gender group. Hypotheses 2.1a and 2.1b were consistently supported whereas the remaining six hypotheses were not supported. Overall, these findings suggest that the race of the target of HS and/or BS may not be as impactful as an individual's own racial identity regarding their endorsement of either form of sexism.

### *Exploratory Analyses*

**Potential Self-by-Target Racial Identity Interaction.** Although there was no evidence that the race of the sexism target affected participants' HS and BS scores at Time 2, it is possible that participants' own racial identity continued to play a role. Therefore, HS and BS responses may stem from an interaction between self and target racial and/or gender identities. However, no specific hypotheses were put forward

regarding this potential interaction, given the lack of previous research regarding race and AS. As such, these analyses were exploratory in nature.

***Black Women vs. White Women.*** Among Black women and White women in this sample, their race-by-gender group had what could be considered a marginally significant main effect on HS scores, $F(1,844) = 3.39$, $p = .066$.[31] However, there was clearly no main effect for racial prime [$F(4,844) = 0.78$, $p = 0.536$] nor an interaction between the two predictors [$F(4,844) = 0.27$, $p = 0.898$]. These findings are not unexpected given previously reported results. Specifically, at both Time 1 and Time 2, Black women scored significantly higher than White women on HS. Additionally, tests of the Study 2 hypotheses revealed that the experimental prime had no effect on Time 2 HS scores for either Black women or White women. The two-way ANOVA reinforces the conclusion that although race-by-gender identity of the participant plays a role, albeit a small one, in HS endorsement, racial identity of the target may not matter at all.

Similar results were found for BS scores at Time 2. Race-by-gender group had a significant main effect on BS scores [$F(1,844) = 24.70$, $p < .001$]; racial prime had no effect on BS scores [$F(4,844) = 2.08$, $p = .081$]; and there was no interaction between the two predictors [$F(4,844) = 0.98$, $p = .416$]. Once again, these findings are not unexpected given previously reported results. As with HS, Black women scored significantly higher than White women on BS at both Time 1 and Time 2. Furthermore, the tests of the Study 2 hypotheses revealed no effect for the experimental prime on Time 2 BS scores for either group of participants. Both sets of findings reinforce the conclusion drawn from the

[31] This effect was statistically significant when only participants who passed the Time 2 manipulation check were included in the analyses, $F(1,721) = 5.26$, $p = .022$.

two-way ANOVA that while the respondent's own racial identity may affect their BS endorsement, the racial identity of the target may not play any kind of role in BS endorsement.

***White Men vs. White Women.*** Among White men and White women in this sample, their race-by-gender group had a significant main effect on HS scores, $F(1,910) = 12.04$, $p = .001$. There was no main effect for racial prime [$F(4,910) = 0.65$, $p = .627$] nor an interaction between the two predictors [$F(4,910) = 0.36$, $p = .834$]. These findings are not unexpected given previously reported results. At both Time 1 and Time 2, White men did score significantly higher than White women on HS, reflecting the main effect for race-by-gender group found in this analysis. Additionally, tests of the Study 2 hypotheses revealed that the experimental prime had no effect on Time 2 HS scores for either White men or White women. Once again, the two-way ANOVA supports the conclusion that although race-by-gender identity of the participant plays a role, albeit a small one, in HS endorsement, racial identity of the target may not matter at all.

Similar results were found for BS scores at Time 2. Race-by-gender group had what could be considered a marginally significant effect on BS scores [$F(1,910) = 3.50$, $p = .062$][32]; racial prime had no effect on BS scores [$F(4,910) = 0.70$, $p = .593$]; and there was no interaction between the two predictors [$F(4,910) = 0.19$, $p = .943$]. As with the results for HS, these findings are not unexpected given previous findings. White men scored significantly higher than White women on BS at both Time 1 and Time 2. Furthermore, the tests of the Study 2 hypotheses revealed no effect for the experimental

[32] This effect was statistically significant when only participants who passed the Time 2 manipulation check were included in the analyses, $F(1,790) = 5.34$, $p = .021$.

prime on Time 2 BS scores for either group of participants. Both sets of findings reinforce the conclusion drawn from the two-way ANOVA that while the respondent's own racial identity may affect their BS endorsement, the racial identity of the target may not play any kind of role in BS endorsement.

***Summary.*** The exploratory results presented here suggest that – at least for Black women and White women, and for White men and White women – their own racial (and gender) identity may influence their HS and BS endorsement, but the racial identity of the target does not play a role in either kind of endorsement. As these analyses were exploratory and answering the question of how respondent and target racial identities might interact to influence AS endorsement was not the primary purpose of Study 2, additional research is needed to better examine this possibility.

**Racial Attitudes (Feeling Thermometers).** The effect of respondents' racial attitudes, as measured by feeling thermometers (FTs), was also explored, with no specific hypotheses to test. First, group means for each racial FT were examined and compared to determine if participants had significantly different views of these racial groups, especially in terms of valence (i.e., positive or negative). These means and standard deviations are presented in Table 47. All FT means were positive, i.e., above 50 degrees, which indicated complete neutrality toward the group. The lowest mean FT score was Black women's mean rating of White Americans, which was 55.21, suggesting a slightly positive attitude. All other FT means were above 65 degrees, suggesting positive attitudes towards these four groups. Although many of these means were significantly different from one another – for example, White women's mean FT rating for White Americans was significantly higher than Black women's mean FT rating for this group – these

differences may not be practically significant because there were no differences in the overall valence of these attitudes. These findings suggest that, overall, participants in this study were not prejudicial towards any of the four racial groups included in the experimental primes. This conclusion is supported by the low HS and BS scores at both time points, which indicate that these participants endorse neither HS nor BS. Racism and sexism are often correlated (along with other prejudicial attitudes; McMahon & Kahn, 2018; Bareket & Fiske, 2023), and so the low scores on the sexism measures and relatively high scores on the racial FTs suggest a participant group that is more egalitarian than not.[33] Taken together, all of these findings suggest that racial FTs are unlikely to have a significant influence on HS and BS endorsement in the present experiment.

This conclusion is supported by additional exploratory analyses taking several different approaches. The effects of racial FTs on HS and BS endorsement in this experiment were minimal. Most analyses revealed nonsignificant effects of the FTs on HS and BS endorsement, even when taking into account participant race-by-gender group as well as the racial prime. The few significant effects were small in magnitude, thus having little practical meaning. Therefore, these results are reported in the Supplemental Materials and not discussed here.

**Discussion**

[33] The group means for the remaining FTs (i.e., gender, political party, and age groups) were also examined. All group means were positive (i.e., greater than 50) for all additional groups, with one exception. Republicans were rated negatively by all three participant groups, with means ranging between 29 and 37. This is not surprising given that nearly 57% of participants identified as Democrats while just under 14% identified as Republicans.

Study 2 was originally intended to focus on the second key research question: Does endorsement differ depending on the race of the women (and men) respondents are thinking about when responding to the ASI? Due to the results of Study 1, Study 2 was modified to also include a replication of the MI analyses, albeit with a different sample. The MI analysis indicated that group comparisons could validly be made between Black and White women, and between White men and women. Therefore, four of the Study 1 hypotheses were also tested in Study 2, in addition to the eight Study 2 hypotheses. Additional exploratory analyses were conducted to examine the role of racial attitudes in HS and BS endorsement.

Regarding the Study 1 hypotheses, White men scored higher than White women on both HS and BS, although the difference was smaller for the latter than for the former; thus, Hypotheses 1.2a and 1.2b were supported. These findings clearly line up with past research that has reliably found these same gender differences. In contrast, regarding the differences between Black women and White women, Hypotheses 1.4a and 1.4b were not supported. In fact, the analyses supported the exact opposite predictions as suggested by Hayes and Swim's (2013) as well as Davis et al.'s (2022) research: Black women scored higher than White women on both HS and BS. Although the difference on HS was quite small, the difference on BS was large.

Regarding the Study 2 hypotheses, for each of the three eligible participant groups, there were no main effects for experimental condition on either HS or BS. In other words, for Black women, White women, and White men alike, there were no differences in HS or BS between the five experimental conditions. Thus, just two of the hypotheses were supported: 2.1a and 2.1b, which predicted no differences between the

White prime condition and the control condition on either HS or BS. Similarly, the exploratory analyses did not reveal any interactions between participant race-by-gender group and target race (i.e., experimental condition) and racial attitudes did not appear to have any effect on HS and BS endorsement in the experiment.

More detailed discussion of Study 2's findings, implications, limitations, and future research directions suggested by this study is taken up in Chapter 4.

# Chapter 4

## General Discussion

Individually, the results of Studies 1 and 2 have important implications for AS theory, measurement, and research. In this section, I discuss in detail the findings of each study, as well as their implications, limitations, and suggest directions for future research. Taken together, the present research paves the way for future research to consider a more in-depth re-examination of AS theory and the ASI from an intersectional perspective and support the more general argument that psychological theories and measures could benefit from an intersectional re-consideration.

### Study 1: Findings, Implications, Limitations, and Future Directions

The key finding of Study 1 was the ASI's lack of MI across the four race-by-gender groups of interest. As a result, the Study 1 hypotheses could not be (validly) tested, and instead analyses focused on potential explanations for this lack of MI. In this section, I discuss potential sources of the noninvariance, the limitations of the study, and directions future research can take to further investigate the MI of the ASI.

#### *Lack of Measurement Invariance*

Despite the inability to answer the original research question, the lack of MI is a critical finding and, as such, this study makes important contributions to the literature, both in terms of research on gender prejudice as well as intersectional research more broadly.

Perhaps most significantly, this is the first study to directly and formally test the MI of the ASI across different racial groups, beyond just checking factor structure and internal reliability. While the latter two are important aspects of psychometrics, and were

examined in the current study, they do not constitute a true test of MI. The current research reinforces what psychometric experts have been saying for years: We need to be more careful about the measures we use, as we may be unknowingly relying on flawed measurement, leading to flawed conclusions (Putnick & Bornstein, 2016). Moreover, just because a measure has been validated once for a particular group does not mean it is applicable to all groups at all times – both measurement theory and intersectionality theory argue that assuming such is inappropriate.

Although we cannot (and should not) throw out the ASI for Black Americans based on a single sample – replications, such as the one conducted as part of Study 2, are needed – the lack of MI in Study 1 does signal that past work with the ASI and Black samples (e.g., Frasure-Yorkley, 2018) may need to be re-evaluated as potentially lacking internal validity, meaning any claims put forth may not be justified or, at the very least, may be highly constrained in generalizability . Additionally, the results indicate a need for MI testing among other non-White (or non-majority) groups (e.g., Latino/a Americans, Asian Americans) who were unrepresented in the original measurement construction and validation. Although from a psychometric standpoint this should be carried out any time a measure is being used with a 'new' group (Putnick & Bornstein, 2016), confirmed measurement noninvariance in one racial minority may indicate that it is more likely to be found in others, as it could indicate that a racial group's unique history in the U.S. is playing a role in the construal of the measure.

**Possible Sources of Invariance.** An additional contribution that this study makes is to provide pathways for news lines of research that delve more deeply into why the ASI may not be a valid measure of sexism for Black Americans. Although definitive

conclusions about the sources of the noninvariance cannot be drawn based on the Study 1 data, exploratory analyses suggest multiple avenues future research in this area can pursue.

***Differential Item Construal.*** One possibility for the lack of configural invariance is that these groups understand and/or interpret certain items to such differing degrees that the items are tapping into different latent phenomena for these groups. As a result, the factor structures are not equivalent across groups, with substantial differences between Black participants and White participants. Different item construals might explain the measurement noninvariance, but the question then becomes: Why do these different construals exist in the first place? Individual, societal, or even historical factors may account for such differences (Davidov et al., 2014). While unable to concretely answer the question, the current study offers insight into two factors that may lead to differential construal of the ASI items between Black and White Americans.

First, as outlined by intersectionality theory, gender is only one part of an individual's social identity and its importance to the overall sense of self can vary from person to person. People for whom gender is not a central part of their identity or everyday experience may think differently about the ASI items than those for whom it is a central part of their identity. And even if gender is important to identity for an individual, that does not mean it is the *most* important part of identity for that individual. For example, racial identity could be the most important part of a person's sense of self, outweighing gender identity and having a greater effect on their construal of the world. The results from the exploratory analysis of the gender and racial identity centrality items suggest that there may be important differences in how central these identities are to

Black and White Americans. Both racial identity and gender identity were more central to Black participants' overall identity compared to White participants, although the difference was greater for the former than the latter. This heightened centrality of marginalized identities may lead Black participants to construe the ASI items differently from White participants. As noted earlier, Black men and women may have more awareness of social inequalities and discrimination due to their lived experiences as racial minorities (Kane, 2000; Harnois, 2017), and it is possible that their strong racial identity centrality contributes to this. It is also possible that racial and gender identity are closely tied for Black Americans – i.e., they think of themselves as not just Black and man/woman, but as Black man or Black woman – but not for White Americans. Furthermore, even if gender is central to the identities of both White and Black Americans, as it seems to be, these groups have different ideas about what it means to be a woman or a man (Settles et al., 2008; Belgrave et al., 2015) and thus different kinds of roles, attitudes, behaviors and so on make up their construction of their gender. This, too, could affect how each group construes items designed to tap into specific kinds of sexism. Nevertheless, conclusions cannot be drawn based on the current study; but these racial differences in racial and gender identity centrality should be further explored as a possible source of MI.

Second, item construal could also be affected by the kinds of men and women that participants have in mind when responding to the ASI. This question – how target race influences HS and BS endorsement – is addressed more directly in Study 2, but the ASI follow-up questions in Study 1 give the impression that when asked, participants do report thinking about different kinds of men and women, at least in terms of race. White

participants primarily considered men and women of their own racial group when responding to the ASI, and less than half considered any racial outgroup to any extent. Similarly, Black participants primarily considered men and women of their own racial group, but they nevertheless considered White men and women to a degree: nearly 60% of both Black men and Black women thought of White men and women to at least "a little" extent. We know that men and women of different racial and ethnic groups elicit different kinds of attitudes and stereotypes (Ghavami & Peplau, 2013), which in turn could affect responding to the ASI depending on the race of the men and women considered. There may be more consistency in how White men and women respond because they mostly have White men and women in mind while responding to the ASI and thus may be relying on a single set of gender attitudes, stereotypes and so on. Black men and women, in contrast, may be less consistent in their responding if they have conflicting gender attitudes and stereotypes in mind thanks to their consideration of both Black and White men and women – this could lead to less reliability in how they respond, resulting in poor factor loadings, for example, thus setting the stage for MI. But even if Black respondents are consistently thinking of only Black men and women while answering the ASI, their understanding and conception of gender may not fit with the (White) understanding and conception underlying AS and the ASI, again leading to MI.

Study 1 suggests two possible explanations for the ASI's lack of MI across the four race-by-gender groups of interest. Neither explanation can be confirmed with the current data, but further investigation is warranted, potentially leading to an adjustment of AS theory to better accommodate all women's experiences of contemporary sexism, or a new theory (and measure) altogether specifically for Black Americans.

***Reverse-Scored Items.*** Despite these promising and intriguing findings, it is possible that the measurement noninvariance has a much simpler source: item wording, rather than item content. As noted in Chapter 2 , the five noninvariant items – hs4, hs10, hs11, in2 and pat1 (see Table 8) – consist of five of the ASI's six reverse-coded items (the full wording of each noninvariant item is reported in Table 9). The exception to this trend is int4, "Men are complete without women," which loads as expected for all groups. Given that the Revised Model tested in this study – in which the five noninvariant items were dropped from the ASI – produced evidence that was generally supportive of MI, the hypothesis that the wording itself is the problem with the original ASI seems a fruitful one that warrants further study.

Furthermore, this pattern (and the results with the Revised Model) is particularly interesting in light of past research that has demonstrated that, in general, reverse-scored (or opposite-scored) items often 'misbehave' (Clifton, 2020, p. 265). For example, Rodebaugh et al. (2011) examined two measures of anxiety, concluding that reverse-scored items lowered validity, but only for older and/or less educated participants. Similarly, Carlson et al.'s (2011) psychometric evaluation of a depression measure also revealed that older participants tend to have difficulties with reverse-scored items, which lowers the measure's internal reliability. Carlson et al. suggested that reverse-scored items increase cognitive processing demands, which is more likely to affect the responding of older rather than younger participants. These are just two of many examples of studies showing that reverse-scored items, while intended to improve the measure's validity by combating acquiescent responding (Rodebaugh et al., 2011), may actually hinder it as well as the measure's reliability.

This begs the question: Have the reverse-scored items been a problem for the ASI in the past? In their original validation work, Glick and Fiske (1996) acknowledge that the reverse-scored items "typically did not fare as well" as the straightforward items in terms of factor loadings, but that these "loadings were generally acceptable" and did not affect the factor structure (p. 498). This suggests that the reverse-scored items were at least somewhat questionable even at the very beginning of AS research. Moreover, a deeper dive into the later literature suggests that reverse-scored items *have* been a problem for the ASI for quite some time, but not necessarily an acknowledged or discussed one. In their landmark 2000 study on the cross-cultural validity of AS and the ASI, Glick et al. reported that the six reverse-coded items from the original English version of the ASI typically did not translate well into other languages and had poor loadings in factor analyses. In some cases, the non-English versions of the ASI re-worded these items to be straightforward, but because there were discrepancies across samples regarding whether these six items were reverse-scored or forward-scored, they were eliminated from all analyses. Thus, Glick et al. (2000) did not validate the original ASI across countries, but rather validated a short-form version of it via factor structure confirmation. This overlooking of the potential problems with the reverse-scored items – i.e., simply getting rid of them rather than addressing the issue more directly – seems to be a trend in AS research, one that was followed, for exploratory purposes, in this study (see Chapter 2 – *Results* section. However, as previously noted, while the revised model tested here is promising, we cannot, at this stage, use it to properly test any hypotheses about the source of the ASI's measurement invariance.

In a study comparing ambivalent attitudes towards men to ambivalent attitudes towards women, Glick and Whitehead (2010) used a short-form version of the ASI consistent of the 12 highest loading items (six items each for the HS and BS factors) from the large cross-cultural study (Glick et al., 2000). This truncated scale was reliable based on measures of internal consistency (although the BS subscales were not checked). What is more notable now in the context of the current study's MI findings is that this short-form ASI did not include any reverse-coded items. Instead, three items that were initially reverse-scored were re-written to be forward-scoring instead. For example, "Feminists are making entirely reasonable demands of men" became "Feminists are making entirely unreasonable demands of men," both intended to tap HS. Glick and Whitehead do not acknowledge these changes from the original ASI nor how they might affect potential response bias. A similar approach was adopted by Rollero et al. (2014), who examined the psychometric properties of this specific short version of the ASI, as they too failed to acknowledge the altered wording for three of the 12 items. It seems likely that the decision to drop the reverse-scored items was based on the problems faced by Glick et al. (2000) in translating those items into languages other than English. It is understandable to want a more easily translated measure so that cross-cultural research can take place and Glick et al. were clearly focused on the cross-cultural nature of their work rather than the psychometrics. However, the reverse-scored items were initially included in the ASI to combat acquiescent responding and it is unknown how removing or altering them may affect response bias.

In sum, it seems that while the reverse-scored items have been known to be a problem, there has not been a systematic study of why that is the case. Clifton (2020)

argues that most reverse-scored items "misbehave" because they are improperly constructed; specifically, they are *derivatively* created. A derivative reverse-scored item, per Clifton, is one in which the item is originally written to be forward-scored, but is edited – by adding antonyms or negators – to be reverse-scored. This approach to reverse-scored items is most likely responsible for their misbehavior: Clifton reasons that such items are more susceptible to respondent carelessness, as it is easy to overlook or misread a single word (i.e., the necessary antonym or negator); more difficult compared to straightforward items; and often concern the wrong latent phenomena. Regarding the latter problem, this is because language is complicated and even a simple negator such as 'not' can have a number of different meanings: the absence of a thing entirely, the partial absence of a thing, the presence of an opposite, and so on. Clifton advises that far more care needs to be taken when constructing reverse-scored items – including writing them as reverse-scored from the start – because they *are* an important tool for minimizing biased responding.

Although not addressed explicitly, it seems that Glick and Fiske (1996) took the derivative approach when creating the reverse-scored items for the ASI, as all of the initial generated items were forward-scored, with some eventually being "reversed" (p. 498). This explains why the reverse-coded items might be a problem – e.g., they might increase cognitive load and/or elicit more respondent carelessness. However, it does not explain *why* these items failed to load properly for Black participants but loaded properly for White participants, a question that cannot be answered by the current study. It is important to note here that although these items did load significantly for White men and women, they were among the poorest performing items for these groups; as Glick and

Fiske (1996) put it, they did not perform as well as the straightforward items but the factor loadings were at least acceptable. Because they are on the margin for these groups, we might see variation in other samples, with the items performing better or worse depending on the study.

The Revised Model supports the item wording hypothesis. Dropping these items generally brought the results into alignment with past research, including Glick et al.'s (2000) cross-cultural study specifically: The revised factor structure held across groups and items loaded on the factors as expected based on the original ASI. However, once again, additional research, including replications, is needed before we can confidently conclude that the measurement noninvariance is the result of item wording rather than anything else (i.e., item content and/or construal). Such research could include studying other factors in concert with race, such as age and education, variables that were highlighted by Rodebaugh et al. (2011) and Carlson et al. (2011) as negatively impacting the performance of reverse-coded items.

Additionally, it must be noted that simply dropping noninvariant or reverse-scored items without considering their content may actually weaken the measure, particularly in terms of content validity. For example, the HS item "Feminists are not seeking for women to have more power than men" represents the perception that women are a threat to the gender hierarchy and intend to take men's position at the top. This is a fundamental aspect of HS and failing to capture it in the measure (e.g., as in the Revised Model) may mean a critical element of HS is not being assessed. Therefore, even if future research confirms the item-wording hypothesis, care must be taken to ensure that removing the items does not weaken the measure's content validity.

***Limitations***

As previously acknowledged, a final decision on the validity of the ASI for Black Americans cannot be made based solely on a single sample and thus replications are required (and one is undertaken in Study 2). Not only is this because samples can behave differently for a number of reasons, but because there are limitations to Study 1. First, the disparities in sample sizes across the four race-by-gender groups may have affected the MI analysis. It can be difficult to determine appropriate sample sizes when conducting invariance testing because statistical power depends on not only sample sizes but also model complexity and other model characteristics; thus, there is no standard 'rule' regarding how many participants should be included in this kind of analysis (Kline, 2016; Meade & Bauer, 2007). Simulation studies suggest that sample sizes below 200 rarely have adequate power for detecting MI, while sample sizes over 400 tend to have high power (Meade & Bauer, 2007). The conventional rule of thumb for CFA/SEM analysis is 20 cases per measured indicator (Kyriazos, 2018) – for the ASI, which has 22 items, this suggests a sample size of 440 per group. As reported in Table 1, even with an oversample of Black Americans, the *N*s for both Black men ($N = 247$) and Black women ($N = 385$) fell well-below the suggested sample size. This limitation was addressed in Study 2, in which quota sampling was used to recruit a relatively equal number of Black men, Black women, White men, and White women.

Second, the ASI was included as a module in a much larger survey conducted by a multi-investigator group. The other modules in the survey were unrelated to the ASI and, theoretically, are not believed to have influenced responding on the ASI, but this possibility cannot be completely ruled out. Moreover, the ASI module was placed

towards the end of the survey and therefore it is possible that participants may have been less engaged, more tired, etc., by the time they reached the ASI, which again could have affected responding for the worse. With both of these limitations in mind, a more 'isolated' test of MI (as seen in Study 2) is advisable.

Third, regarding the examination of possible sources of MI, the measures included in Study 1 were imperfect. The racial and gender centrality items were based on Huddy et al.'s (2015) measure of partisan identity, which includes four items. For this survey, however, just two items were included per construct, which may have reduced validity. Internal consistency *was* low for the two White identity items, suggesting it may not have been measuring what it was intended to measure. Future research delving into this possible explanation for MI should include the full measures of identity centrality, or perhaps use different measures altogether. Similarly, the ASI follow-up questions were not ideal for fully tapping into participants' mindsets while responding to the ASI. Future research should more thoroughly investigate this piece of the puzzle with more comprehensive methods and measures.

Finally, Black participants and White participants did differ on some demographic variables other than race, and these variables may have influenced HS and BS responding but were unaccounted for (the possibility is present but the *how* is unknown). Specifically, political party identity seemed to distinguish Black participants from White participants, just as race did. As noted in Chapter 2 (see Table 4), in Study 1, the majority of Black participants identified as Democrats, including nearly 60% of Black men and 70% of Black women. Furthermore, less than 8% of Black men and less than 5% of Black women identified as Republicans in Study 1. Among White participants in Study 1,

the number of Democrats and Republicans were more balanced: About 31% of White men and 37% of White women identified as Democrats, whereas about 29% of White men and 31% of White women identified as Republicans. The lack of political party diversity among Black participants may lead to the conflation of race and political party affiliation and the possibility that the latter is driving responding rather than the former. Future research must build on Bayton et al. (1956) to identify which social identity variables, such as political party identity, are most likely to be conflated and how researchers can disentangle those variables that are so closely related that they obscure one another's effects.

### *Moving Forward*

Despite these limitations, the current study is important to the future of AS research. It provides several avenues for future research to explore, including more thoroughly investigating the role that identity centrality and gender construals play in ASI responding as well as diving further into the problem of the reverse-scored items. A more qualitative approach may be useful to better understanding how participants of different racial groups construe the ASI items. For example, interviews or focus groups could be employed to have participants talk through their understanding and interpretation of the items. If groups have unique understandings or interpretations, this could account for the MI. Once patterns and themes in understanding and interpretation are identified, additional quantitative work could be undertaken to test the hypothesized sources of noninvariance. Regarding the item wording explanation, one possibility is to administer a version of the full ASI that has no reverse-scored items (i.e., as with the short version from Glick & Whitehead [2010], reverse-coded items would be re-written to be forward-

scored). This version could then be tested for MI across the four race-by-gender groups. Another option is to administer the original ASI and this revised version to a single sample and directly compare responses. For example, it could be determined if the responses to the original reverse-scored items and their straightforward forms match one another or change

More generally, Study 1 reinforces the intersectional argument that MI should never be assumed, especially for non-majority groups who likely did not make up a significant part of the participant samples in initial measure construction and validation (Else-Quest & Hyde, 2016b). In this study, race was highlighted as an important influence on gender, but it is not the only other social identity that could affect AS theory and measurement. For example, sexual orientation could play an important role, as AS theory, rooted in biology as it is, is explicitly heteronormative. Additionally, AS theory is cisgender-centered and thus tends to conflate sex and gender. This is not unexpected given the timeframe when AS theory was developed, but with our current – and better – understanding of the distinction between the two, as well as the recognition of other gender identities, transgender and nonbinary perspectives on sexism need to be considered. A recent systematic review of the AS literature agrees that future research needs to include samples that are diverse in terms of gender identity and sexual orientation (Bareket & Fiske, 2023). Other group variables have already been examined in the context of AS endorsement, including age (Hammond et al., 2014), religion (Glick et al., 2002; Burn & Busso, 2005; Mikołajczak & Pietrzak, 2014), education (Glick et al., 2002) and conservatism (Christopher & Mull, 2006), but this previous work did not

consider the ASI's MI across these different groups and therefore it is unclear how valid the findings related to these variables are.

Furthermore, the current study strongly suggests that other constructs aside from AS may need to re-examined through an intersectional lens, including testing the MI of their measures across different groups of interest. In many ways, Study 1, with AS theory and the ASI, serves as a case study for how to apply intersectionality theory to existing theory and measurement.

**Study 2: Findings, Implications, Limitations, and Future Directions**

Study 2 failed to replicate the MI findings of Study 1: It was determined that while valid mean comparisons could not be made between all four race-by-gender groups, comparisons could be made between two sets of groups, i.e., between Black and White women, and between White men and women. As a result, four of the eight Study 1 hypotheses were tested in addition to the eight Study 2 hypotheses. Altogether, the results of Study 2 suggest that although respondent race appears to be a critical variable in HS and BS endorsement, the race of the target may not be particularly relevant. The first finding supports previous research by Hayes and Swim (2013) and Davis et al. (2022), whereas the latter contradicts the findings of McMahon and Kahn (2016). In this section, I discuss potential explanations for the findings of Study 2 and their implications, as well how the limitations of this study can be addressed in future research.

***Measurement Invariance***

Using the Time 1 data, the MI of the ASI was once again tested among the four race-by-gender groups of interest: Black men, Black women, White men, and White women. The Study 2 MI analyses failed to replicate the Study 1 findings. In Study 1, MI

could not be confirmed for any of the four groups, and it seemed likely that at least one source of variance was the ASI's reverse-scored items, which performed especially poorly for Black participants in the Study 1 sample. Therefore, it was somewhat surprising to see much improved MI in Study 2: The ASI 'worked' for all groups except Black men, although only two sets of group comparisons could be confidently made (i.e., (1) between Black and White women, and (2) between White men and women).

What is behind these different results? Most notably, the reverse-scored items performed much better in Study 2 than in Study 1, even among Black men. Similarly – and likely relatedly – one of the problematic BS subscales from Study 1, heterosexual intimacy, had much better internal consistency in Study 2, reaching acceptable levels (i.e., $\alpha \geq .70$) for all four participant groups. The other problematic BS subscale, protective paternalism, once again did not reach acceptable levels of internal consistency for Black participants; however, the Cronbach's alpha values were better in Study 2 than in Study 1. Furthermore, the number of Black participants in Study 2 was much higher than in Study 1, and each race-by-gender group included at least 400 participants. The greater *N*s in Study 2 may mean that a better test of MI could be undertaken; if so, then this would indicate that the Study 2 results are more likely to replicate than the Study 1 results, at least in appropriately sized samples.

An additional question regarding the MI results is: Why were Black men the only group for whom scalar invariance could not be confirmed? It is difficult to answer this question due to the nature of Study 2, which was not primarily focused on MI analysis. Nevertheless, one survey question might provide some insight. As noted in the Chapter 3 *Results* section, nearly 26% of Black men failed the manipulation check question;

moreover, 36% of the total manipulation check failures were Black men. If the manipulation check can serve as a proxy for attention, perhaps Black men were not paying as close attention as Black women, White men, and White women, which affected their responding to the ASI resulting in measurement noninvariance for this group. However, all participants in the final sample, including Black men, passed the attention check question administered prior to the ASI, suggesting that attention (at least as assessed by the one item used in Study 2) was not an issue for participants. Additional research is needed to uncover the role that attention might play in MI testing.

Whatever the reasons behind these inconsistent MI results, they clearly indicate that more MI work is needed to truly understand its validity or lack thereof for Black Americans (as well as other groups). Just as we should be cautious in making broad claims based on a single study's results, we should not draw definitive conclusions about MI based on a single study and instead should synthesize the results of multiple replications. Additionally, these results emphasize the fact that, even if it has been established for one sample, the MI of a measure cannot be assumed to hold for a different sample and should be tested any time group comparisons are made (Putnick & Bornstein, 2016).

### *Study 1 Hypotheses*

The MI analyses indicated that valid group comparisons on the ASI could be made between two pairs of participant groups: Black and White women, as well as White men and women. Therefore, four of the Study 1 hypotheses were tested using the Study 2 Time 1 data: Hypotheses 1.2a, 1.2b, 1.4a, and 1.4b. As expected, White men scored significantly higher on both HS and BS than White women, with the difference in BS

smaller than the difference in HS. These findings are in line with past research that has consistently found these exact gender differences in HS and BS endorsement (Bareket & Fiske, 2023). However, the means for each group on both HS and BS were below the midpoints of response scales (3.5), indicating that despite this difference, neither group was, on average, sexist.

Regarding Black women and White women, Hypotheses 1.4a and 1.4b were not confirmed. In fact, it was found that Black women scored significantly higher than White women on both HS and BS – the exact opposite pattern than was predicted. Nevertheless, these findings are in line with those reported by Hayes and Swim (2013) as well as Davis et al. (2022). The statistical difference in HS was small and unlikely to hold any practical significance: Despite scoring higher on HS than White women, Black women still scored below the midpoint on the response scale, indicating that they are not hostile sexists. Black women's mean for BS, however, *is* above the response scale midpoint, but at 3.55 it is just barely above the midpoint and so we cannot confidently conclude that Black women are benevolent sexists whereas White women are not. Instead, Black women may be better classified as neutral regarding BS endorsement, whereas White women more clearly disavow it.

Past research illuminates some potential explanations for the BS difference in particular, which was in a different direction than hypothesized as well as larger than expected, with Black women's BS mean crossing the scale midpoint and just barely stepping onto the endorsement side. First, as discussed in Chapter 2, Black women typically report less traditional gender attitudes than White women; however, this difference may be domain specific, with Black women being less traditional in their

attitudes about maternal employment and a woman's family role, but more traditional in their attitudes about romantic relationships. Davis et al. (2022) argued that it is the latter that should be more influential in BS endorsement, but I suggested that age may influence how salient each attitude domain is to an individual, with older women being influenced by their attitudes about employment and family roles more than their attitudes about interpersonal relationships. Study 2's results suggest that Davis et al. may be correct in arguing that the latter are more important than the former in BS endorsement. Recent research that has explored HS and BS endorsement among gay/lesbian, bisexual, and heterosexual individuals revealed that gay men have the lowest levels of BS, providing some evidence for the claim that heterosexual intimacy is a particularly powerful driver of BS endorsement (Cowie et al., 2019).[34]

Second, it is possible that rather than making them more resistant to BS, Black women's experiences with racism may actually make them less sensitive to BS than White women. Although there is evidence that Black women may be more bothered by BS behaviors than White women (Settles et al., 2008), other researchers have argued that their experiences with racism makes Black women hypervigilant to overt threats, such as those posed by HS, and may lead them to embrace or minimize BS, which is a less threatening ideology that promotes protection of their group (Davis et al., 2022). In other words, their experiences with racism may lead Black women to be less sensitive to the negatives and more susceptible to the so-called positives of BS than White women. Third, additional variables not accounted for in the Study 2 analyses may be influencing

[34] It must be noted, however, that the MI of the ASI across the groups of interest (i.e., across sexual orientations) was not directly tested in this study.

the BS scores of one or groups. For example, past research has linked religiosity with BS endorsement (Glick et al., 2002; Burn & Busso, 2005; Mikołajczak & Pietrzak, 2014); more recently, Davis et al. (2022) concluded that for Black women, religiosity mediated the relationship between their race-by-gender identity and their BS endorsement. Religiosity is likely related to Black women's more traditional attitudes regarding gender roles in romantic relationships, explaining why it mediates BS endorsement (Davis et al., 2022). Finally, it is possible that Black women's BS simply functions differently than White women's BS and so they are able to endorse it to a greater extent despite their overall egalitarian attitudes about gender.

None of these explanations can be empirically confirmed or disconfirmed at present, thus, future research is needed to better understand why Black women seem to endorse BS to a greater degree than White women and what the consequences are for this difference in terms of behavior. As suggested in the previous section discussing the results of Study 1, qualitative research may provide the best pathway to investigating the possible explanations outlined above, as participants can provide more detailed responses about their understanding and interpretation of both AS theory and the ASI. First and foremost, however, it must be determined whether the BS subscale of the ASI is measuring the same construct for both groups and the possibility that BS functions differently for each group must also be investigated, as this could have important implications for behaviors related to BS.

### *Study 2 Hypotheses*

Of the eight Study 2 hypotheses, just two could be confirmed. Analyses revealed that for Black women, White women, and White men, there were no differences in HS or

BS endorsement across the five experimental conditions. Because of the lack of differences between the control and White prime conditions for each group, Hypotheses 2.1a and 2.1b were confirmed. However, the remaining six hypotheses all predicted that certain differences would emerged across conditions and thus were not confirmed. These conclusions were reinforced by the exploratory analyses examining the interaction between participant race-by-gender group and experimental prime, which showed consistent main effects for the participant variable but neither main effects for the target variable nor interactions between the two variables. There are several possible explanations for the lack of differences found in HS and BS endorsement.

**An Unprejudiced Sample.** Perhaps the most likely explanation for the present results is that the participants in Study 2 were simply not particularly prejudiced, against either women or different racial groups (see also Footnote 33). As previously discussed, the HS and BS means at both time points indicate that none of the three groups can be clearly classified as hostile sexists and/or benevolent sexists. Instead, the three groups clearly rejected HS and BS, with the exception of Black women who could be considered to be neutral toward BS rather than outright rejecting it. Similarly, the group means on the four racial group feeling thermometers were all positive, suggesting that while participants may have varied in just how positively they viewed a group, they did not differ in the overall valence of their racial attitudes. Thus, although the racial prime in the ASI instructions may have worked as intended, it is possible that participants themselves did not hold the expected negative stereotypes about the different racial groups that would trigger differential HS and BS responding. In other words, although participants may have been aware of negative stereotypes about women of different racial groups –

especially sexual stereotypes – they may not have endorsed them themselves, and thus would not apply them to the target women in the experiment. Future research should more directly consider this distinction between awareness of cultural stereotypes and endorsement of those same stereotypes, as activation of the former may not be enough to trigger application, i.e., the expected subtyping based on target race.

**Changes in Stereotypes.** A second possibility is that stereotypes about women of different racial groups are changing in ways that eliminate race as a potential trigger for differential HS and BS responding. More specifically, perhaps stereotypes for these groups of women are converging, leading to a new stereotype of women in general that might apply to women of all racial backgrounds. For example, perhaps the "women are wonderful" perception is no longer as relevant today as it was even just ten years ago. Future research can continue to replicate the work by Ghavami and Peplau (2013) to determine if race, gender, and race-by-gender stereotypes have grown more similar or even more distinct over time.

**Stereotypes of Men.** Stereotypes of men are also part of AS theory, but it is stereotypes of women that are given more attention in AS research due to the fact that women are the targets of HS and BS, not men. However, the original ASI does instruct respondents to think about both women *and* men, and so stereotypes of men may also play an important role in HS and BS endorsement. For example, Asian American men are stereotyped as sexually unattractive, undesirable as romantic partners, effeminate, and emasculated (Liu & Wong, 2018). These stereotypes are in stark contrast to those underlying AS, namely that men are agentic, dominant, and assertive (Ellemers, 2018; Sczesny et al., 2018), and thus respondents thinking about Asian American men and

women while responding to the ASI may be thinking about nontraditional men more than traditional women which could affect their endorsement. In contrast, Black men are stereotyped as hypermasculine and sexually aggressive (Rosenthal et al., 2020), which could affect how respondents subtype Black women, who may typically seem nontraditional but may be viewed as more traditional alongside traditional-seeming Black men. If stereotypes of men do play a role in HS and BS endorsement, it is possible that McMahon and Kahn (2016) found a main effect for target race on BS because they specifically did not ask participants to simultaneously think about any men, whereas in Study 2 participants were asked to simultaneously consider both men and women, as the original ASI instructs them to do. Additional research is needed to better understand the role that stereotypes of men, particularly intersectional stereotypes of men, may play in HS and BS endorsement.

**The Experimental Primes.** Another potential explanation for the null results is that the method intended to prime racial group stereotypes, that would in turn affect HS and BS responses, did not work as intended. It is possible that the prime was too direct, cuing participants to the researcher's interest in race and thus triggering a response that intentionally ignored the racial aspect of the ASI instructions. It is also possible that such racial priming is not as effective at the group level as it is at the individual level, as demonstrated in McMahon and Kahn's (2016) research.

**Target Race is Irrelevant.** Finally, it is possible that there is no need to explain the lack of differences found in HS and BS endorsement in this experiment. Instead, perhaps target race truly is irrelevant to HS and BS endorsement, or is at least much less relevant than participant race. This conclusion would conflict with the findings from

McMahon and Kahn (2016) – at least those regarding BS – and thus seems like it may be the least likely possibility of the five discussed here. Nevertheless, it is inappropriate to draw such a conclusion based on a single study and therefore additional research, using different kinds of racial priming, is needed to determine the boundary conditions of such an effect.

Furthermore, although it was expected that participants would stereotype (and therefore subtype) based on target race in conjunction with target gender, participants may have instead focused solely on the latter. If so, race would not be expected to influence responding. One model of intersectional perception, the lens-based model, suggests that targets are not always perceived in terms of their intersectional identities, but may instead be perceived as members of a single group (Hudson et al., 2024). Perhaps when it comes to the application of HS and BS, targets are most likely to be viewed in terms of a single group identity, i.e., their gender, with other identities having little to no influence. Future research can address this possibility by further investigating who respondents tend to think of when responding to the ASI.

### *Limitations*

Although designed to directly address some of the limitations present in McMahon and Kahn's (2016) study on target race and AS, the present study is not without its own limitations. First, as discussed in the previous section, the racial primes may not have worked as intended; this seems a particularly likely possibility given that nearly 17% of the final sample failed the manipulation check. However, it is possible that the prime worked as intended in activating race-by-gender stereotypes but participants did not act on those stereotypes.

Second, although the present study extended McMahon and Kahn's work by including Asian American and Latinx American targets, there are still other racial groups that need to be included in this research, e.g., Middle Eastern Americans and Native Americans. However, research on stereotypes about other racial groups is even more limited than the research for the groups included in Study 2, making it difficult to immediately move on to studying additional racial minority groups of women. Additionally, this limitation extends to the participant sample, which in the present study only included individuals who were Black or White. Individuals who are members of other racial minority groups, e.g., Asian or Latinx, may respond differently than Black and/or White participants.

Third, the time between the two study phases may have been too short. If participants recognized that they were being asked the same questions at Time 2 as at Time 1 (i.e., the ASI), they may not have been as careful in their reading of the Time 2 survey as expected – especially with regard to reading the instructions – and instead may have relied on memory to answer the questions rather than fully engaging with them anew and with the primed racial group in mind. A replication with a larger gap between the two time points may produce different results.

Finally, as noted in the previous section, the present study did not investigate participants' awareness of versus endorsement of cultural stereotypes of White, Black, Asian, and Latina stereotypes. Participants in the study may have been aware of the stereotypes discussed in Chapter 3's review of the literature, but did not apply them to the target women in the experiment. A different group of participants – especially one that is more prejudiced than the present group – may not only be aware of race-by-gender

stereotypes but endorse them as well, resulting in different results than those found in the present study.

### *Moving Forward*

Future research can build on and extend the present research, perhaps most effectively by directly addressing Study 2's limitations. First, if this study were to be replicated, additional care should be taken to ensure that the prime is strengthened and works as intended. Future research could also extend the current study's approach, as well as McMahon and Kahn's, by testing more subtle or even subliminal primes that are unlikely to be detected by participants. BS in particular often operates in a subtle matter, thus a subtler prime may be more effective in activating the subtype that triggers a BS response.

Second, future research must move beyond the Black/White dichotomy in its intersectional approach to AS and the ASI. In general, intersectional research on race has been criticized for focusing primarily on this dichotomy (Li, 2014) and any future research in this area in particular (including additional MI work on the ASI) must rectify this limitation, even if it might involve additional work in other domains. For example, before continuing research on target race and AS endorsement, researchers might consider taking a step back to further explore intersectional race-by-gender stereotypes that can provide insight into differential applications of HS and BS (e.g., such as those of Middle Eastern and indigenous women). Furthermore, participants with other racial identities aside from Black and White (e.g., Asian or Latinx) must be included in future research, as results with one group of racial minority participants cannot be extended to

all racial minority groups. There may be important differences regarding how these groups respond to the target's race.

Finally, future research examining the influence of target race on HS and BS endorsement should better account for the distinction between being aware of cultural stereotypes of different race-by-gender groups and personally endorsing those stereotypes and applying them to targets of HS and/or BS. As previously noted, personal endorsement of these stereotypes may be the key factor in determining whether or not individuals will differentially apply HS and BS to targets based on their (perceived) race. Studies with multiple timepoints may be able to most effectively measure personal endorsement before experimentally testing the effect of racial primes on HS and BS endorsement.

**Conclusion**

There is no question as to the impact that AS theory has had on research examining gender prejudice, both in the United States and across the globe. Furthermore, HS and BS have been clearly linked to a number of important outcomes related to gender (in)inequality. As reported by Bareket and Fiske (2023) in their systematic review of the AS literature, HS and BS have been linked to domestic and sexual violence, sexual harassment, and workplace performance, just to name a few. Proper understanding of the attitudes underlying these behaviors (and others), as facilitated by AS theory and research, can better position researchers to develop interventions to alleviate gender inequalities.

Nevertheless, my research supports the claim that, as suggested by an intersectional perspective, caution is called for regarding the generalizability of results

such as those reported in Bareket and Fiske's review (2023). Most importantly, both of the current studies reinforce the importance of testing the ASI's MI across groups of empirical interest, as the measure may not work as intended for all groups. In particular, Study 1 highlights the need to further investigate the ASI's reverse-scored items, which may be hindering the measure's performance across groups. This, or any other issues that disrupt the ASI's MI, may in turn bias results and lead to invalid conclusions about the nature and influence of AS. Relationships between AS and certain behaviors, for example, may be more nuanced than thought or perhaps entirely different depending on the group being considered. Similarly, these studies underscore the need for AS theory and research to utilize an intersectional perspective, as important group differences in endorsement as well as outcomes linked to AS may be overlooked due to reliance on samples that better represent majority versus minority groups. For example, although research has generally found a positive link between BS and right-wing authoritarianism (Bareket & Fiske, 2023), Black women's greater BS endorsement (as found in Study 2) may not be associated with the latter due to their lower levels of authoritarianism compared to White Americans (Perez & Hetherington, 2014). Instead, BS may have other correlates for Black women that have not yet been uncovered.

AS research should continue to move forward but, at the same time, steps must be taken to look back at the theory and measure, which may need to be adjusted in order to better apply to different groups, especially those underrepresented in most AS research. Retroactively applying an intersectional perspective can allow researchers to better understand contemporary sexism and the role that other social identity variables, especially race, play in this domain.

## References

Abrams, J. A., Javier, S. J., Maxwell, M. L., Belgrave, F. Z., & Nguyen, B. A. (2016). Distant but relative: Similarities and differences in gender role beliefs among African American and Vietnamese American women. *Cultural Diversity and Ethnic Minority Psychology*, *22*(2), 256–267. https://doi.org/10.1037/cdp0000038

Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between Prolific and MTurk samples. *Scientific Reports, 13*(1), 20574. https://doi.org/10.1038/s41598-023-46048-5

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

American Psychological Association. (2018, September). *A glossary: Defining transgender terms*. https://www.apa.org/monitor/2018/09/ce-corner-glossary

American Psychological Association. (2022, October). *Race and ethnicity*. https://www.apa.org/topics/race-ethnicity

America's racial reckoning. (n.d.) NBC News. https://www.nbcnews.com/americas-racial-reckoning

Andreoletti, C., Leszczynski, J. P., & Disch, W. B. (2015). Gender, race, and age: The content of compound stereotypes across the life span. *International Journal of Aging and Human Development*, *81*(1–2), 27–53. https://doi.org/10.1177/0091415015616395

Arrizón, A. (2008). Latina subjectivity, sexuality and sensuality. *Women & Performance: A Journal of Feminist Theory, 18*(3), 189-198. https://doi.org/10.1080/07407700802495928

Ashley, W. (2014). The angry black woman: The impact of pejorative stereotypes on psychotherapy with black women. *Social Work in Public Health, 29*(1), 27-34. https://doi.org/10.1080/19371918.2011.619449

Astor, M. (2020, October 9). Kamala Harris and the 'double bind' of racism and sexism. *The New York Times.* https://www.nytimes.com/2020/10/09/us/politics/kamala-harris-racism-sexism.html

Bareket, O., & Fiske, S. T. (2023). A systematic review of the ambivalent sexism literature: Hostile sexism protects men's power; benevolent sexism guards traditional gender roles. *Psychological Bulletin, 149*(11-12), 637–698. https://doi.org/10.1037/bul0000400

Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology, 35,* 633-642. https://doi.org/10.1002/ejsp.270

Bayton, J. A., McAlister, L. B., & Hamer, J. (1956). Race-class stereotypes. *The Journal of Negro Education, 25*(1), 75-78. https://doi.org/10.2307/2293141

Bayton, J. A., & Muldrow, T. W. (1968). Interacting variables in the perception of racial personality traits. *Journal of Experimental Research in Personality, 3*(1), 39–44.

Becker, J. C., & Wright, S. C. (2011). Yet another dark side of chivalry: Benevolent sexism undermines and hostile sexism motivates collective action for social change. *Journal of Personality and Social Psychology, 101,* 62-77. https://doi.org/10.1037/a0022615

Beinart, P. (2016, October). Fear of a female president: Hillary Clinton's candidacy has provoked a wave of misogyny—one that may roil American life for years to

come. *The Atlantic*. https://www.theatlantic.com/magazine/archive/2016/10/fear-of-a-female-president/497564/

Belgrave, F. Z., Abrams, J. A., Hood, K. B., Moore, M. P., & Nguyen, A. B. (2016). Development and validation of a preliminary measure of African American women's gender role beliefs. *Journal of Black Psychology*, *42*(4), 320–342. https://doi.org/10.1177/0095798415576614

Bellovary, A., Armenta, A. D., & Reyna, C. (2020). Stereotypes of immigrants and immigration in the United States. In E. C. Voyles & J. T. Nadler (Eds.), *Stereotypes: The incidence and impacts of bias* (pp. 146-164). Bloomsbury Publishing.

Bowleg, L. (2008). When Black + lesbian + woman ≠ Black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research. *Sex Roles*, *59*(5–6), 312–325. https://doi.org/10.1007/s11199-008-9400-z

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues, 28*(2), 59-78. https://doi.org/10.1111/j.1540-4560.1972.tb00018.x

Brown-Iannuzzi, J. L., Cooley, E., Cipolli, W., & Mehta, S. (2022). Race, ambivalent sexism, and perceptions of situations when police shoot Black women. *Social Psychological and Personality Science, 13*(1), 127-138. https://doi.org/10.1177/1948550620987659

Bu, W., & Borgida, E. (2021). A four-dimensional model of Asian American stereotypes. *Group Processes & Intergroup Relations, 24*(8), 1262-1283. https://doi.org/10.1177/1368430220936360

Burn, S. M., & Busso, J. (2005). Ambivalent sexism, scriptural literalism, and religiosity. *Psychology of Women Quarterly, 29*(4), 412-418. https://doi.org/10.1111/j.1471-6402.2005.00241.x

Buss, D. M., & Schmitt, D. P. (1993). Sexual Strategies Theory: An evolutionary perspective on human mating. *Psychological Review, 100*(2), 204–232. https://doi.org/10.1037/0033-295X.100.2.204

Butler-Barnes, S. T., Allen, P. C., Williams, M. A., & Jackson, A. N. (2020). Stereotypes of African Americans. In E. C. Voyles & J. T. Nadler (Eds.), Stereotypes: The incidence and impacts of bias (pp. 109-127). Bloomsbury Publishing.

Carlson, M., Wilcox, R., Chou, C. P., Chang, M., Yang, F., Blanchard, J., ... & Clark, F. (2011). Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychological Assessment, 23*(2), 558-562. https://doi.org/10.1037/a0022484.

Cauce, A. M., & Domenech-Rodriguez, M. (2002). Latino families: Myths and realities. In J. M. Contreras, K. A. Kerns, & A. M. Neal-Barnett (Eds.), *Latino children and families in the United States: Current research and future directions* (pp. 3-25). Praeger.

Christopher, A. N., & Mull, M. S. (2006). Conservative ideology and ambivalent sexism. *Psychology of Women Quarterly, 30*, 223-230. https://doi.org/10.1111/j.1471-6402.2006.00284.x

Carter, J., Corra, M., & Carter, S. K. (2009). The interaction of race and gender: Changing gender-role attitudes, 1974–2006. *Social Science Quarterly*, *90*(1), 196-211. https://doi.org/10.1111/j.1540-6237.2009.00611.x

Cheeseborough, T., Overstreet, N., & Ward, L. M. (2020). Interpersonal sexual objectification, Jezebel stereotype endorsement, and justification of intimate partner violence toward women. *Psychology of women quarterly, 44*(2), 203-216.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*(3), 471–492. https://doi.org/10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Clarke, H.M. (2020). Gender stereotypes and gender-typed work. In K. Zimmermann (Ed.), *Handbook of labor, human resources and population economics*. Springer, Cham. https://doi.org/10.1007/978-3-319-57365-6_21-1

Clifton, J. D. (2020). Managing validity versus reliability trade-offs in scale-building decisions. *Psychological Methods, 25*(3), 259-270. http://dx.doi.org/10.1037/met0000236

Cole, E. R. (2009). Intersectionality and research in psychology. *American Psychologist, 64*(3), 170–180. https://doi.org/10.1037/a0014564

Cole, E. R., & Zucker, A. N. (2007). Black and White women's perspectives on femininity. *Cultural Diversity and Ethnic Minority Psychology*, *13*(1), 1–9. https://doi.org/10.1037/1099-9809.13.1.1

Collins, P. H. (2022). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge.

Connor, R. A., Glick, P., & Fiske, S. T. (2016). Ambivalent sexism in the twenty-first century. In C. Sibley & F. Barlow (Eds.), *The Cambridge handbook of the psychology of prejudice* (pp. 295–320). Cambridge University Press. https://doi.org/10.1017/9781316161579.013

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *The University of Chicago Legal Forum*, *140*, 139–168.

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology, 93*(5), 764-779. https://doi.org/10.1037/0022-3514.93.5.764

Davis, A. W., Levant, R. F., & Pryor, S. (2018). Traditional femininity versus strong Black women ideologies and stress among Black women. *Journal of Black Studies, 49*(8), 820-841. https://doi.org/10.1177/0021934718799016

Davis, D. E., Rice, K., Van Tongeren, D. R., Hook, J. N., DeBlaere, C., Worthington Jr, E. L., & Choe, E. (2016). The moral foundations hypothesis does not replicate well in Black samples. *Journal of Personality and Social Psychology, 110*(4), e23-e30. https://doi.org/10.1037/pspp0000056.

Davis, T. M., Settles, I. H., & Jones, M. K. (2022). Standpoints and situatedness: Examining the perception of benevolent sexism in Black and white undergraduate women and men. *Psychology of Women Quarterly, 46*(1), 8-26. https://doi.org/10.1177/03616843211043108

Davis, S., & Tucker-Brown, A. (2013). Effects of Black sexual stereotypes on sexual decision making among African American women. *Journal of Pan African Studies, 5*(9), 111-128. https://doi.org/10.1177/0021934713481

Deliovsky, K. (2008). Normative White femininity: Race, gender and the politics of beauty. *Atlantis: Critical Studies in Gender, Culture & Social Justice*, *33*(1), 49–59.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121–149. https://doi.org/10.1177/0748175610373459

Donovan, R. A. (2011). Tough or tender: (dis)similarities in White college students' perceptions of Black and White women. *Psychology of Women Quarterly*, *35*(3), 458–468. https://doi.org/10.1177/0361684311406874

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and

SONA. *PLoS ONE, 18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Durik, A. M., Hyde, J. S., Marks, A. C., Roy, A. L., Anaya, D., & Schultz, G. (2006). Ethnicity and gender stereotypes of emotion. *Sex Roles*, *54*(7-8), 429-445. https://doi.org/10.1007/s11199-006-9020-4

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology, 69*, 275-298. https://doi.org/10.1146/annurev-psych-122216-011719

Else-Quest, N. M., & Hyde, J. S. (2016a). Intersectionality in quantitative psychological research: I. Theoretical and epistemological issues. *Psychology of Women Quarterly, 40*(2), 155–170. https://doi.org/10.1177/0361684316629797

Else-Quest, N. M., & Hyde, J. S. (2016b). Intersectionality in quantitative psychological research: II. Methods and techniques. *Psychology of Women Quarterly, 40*(3), 319–336. https://doi.org/10.1177/0361684316647953

Fehr, B. (1996). *Friendship processes*. Sage Publications, Inc.

Fields, A. M., Swan, S., & Kloos, B. (2010). "What it means to be a woman:" Ambivalent sexism in female college students' experiences and attitudes. *Sex Roles, 62*(7), 554-567. https://doi.org/10.1007/s11199-009-9674-9

Fowers, A.F., Fowers, B.J. (2010). Social dominance and sexual self-schema as moderators of sexist reactions to female subtypes. *Sex Roles 62*, 468–480. https://doi.org/10.1007/s11199-009-9607-7

Fox, J., & Weisburg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE Publications, Inc. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frasure-Yokley, L. (2018). Choosing the velvet glove: Women voters, ambivalent sexism, and vote choice in 2016. *Journal of Race, Ethnicity and Politics*, *3*(1), 3–25. https://doi.org/10.1017/rep.2017.35

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science*, *24*(4), 498-506. https://doi.org/10.1177/095679761245778

Gann-Bociek, M., & Harvey, R. D. (2020). Gender stereotypes. In E. C. Voyles & J. T. Nadler (Eds.), *Stereotypes: The incidence and impacts of bias* (pp. 175-193). Bloomsbury Publishing.

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes: Testing three hypotheses. *Psychology of Women Quarterly*, *37*(1), 113–127. https://doi.org/10.1177/0361684312464203

Glick, P., Diebold, J., Bailey-Werner, B., & Zhu, L. (1997). The two faces of Adam: Ambivalent sexism and polarized attitudes toward women. *Personality and Social Psychology Bulletin*, *23*(12), 1323–1334.

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491–512. https://doi.org/10.1037/0022-3514.70.3.491

Glick, P., & Fiske, S. T. (1997). Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychology of Women Quarterly*, *21*(1), 119–135. https://doi.org/10.1111/j.1471-6402.1997.tb00104.x

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*(2), 109–118. https://doi.org/10.1037/0003-066X.56.2.109

Glick, P., & Fiske, S. T. (2011). Ambivalent sexism revisited. *Psychology of Women Quarterly*, *35*(3), 530–535. https://doi.org/10.1177/0361684311414832

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., Adetoun, B., Osagie, J. E., Akande, A., Alao, A., Brunner, A., Willemsen, T. M., Chipeta, K., Dardenne, B., Dijksterhuis, A., Wigboldus, D., Eckes, T., Six-Materna, I., Expósito, F., … López, W. L. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, *79*(5), 763–775. https://doi.org/10.1037/0022-3514.79.5.763

Glick, P., Lameiras, M., & Castro, Y. R. (2002). Education and Catholic religiosity as predictors of hostile and benevolent sexism toward women and men. *Sex Roles*, *47*(9), 433-441. https://doi.org/10.1023/A:1021696209949

Glick, P., & Raberg, L. (2018). Benevolent sexism and the status of women. *APA Handbook of the Psychology of Women: History, Theory, and Battlegrounds (Vol. 1).*, *1*, 363–380. https://doi.org/10.1037/0000059-018

Goff, P. A., & Kahn, K. B. (2013). How psychological science impedes intersectional thinking. *Du Bois Review*, *10*(2), 365–384. https://doi.org/10.1017/S1742058X13000313

Goff, P. A., Thomas, M. A., & Jackson, M. C. (2008). “Ain’t I a woman?”: Towards an intersectional approach to person perception and group-based harms. *Sex Roles*, *59*(5–6), 392–403. https://doi.org/10.1007/s11199-008-9505-4

Goh, J. X., & McCue, J. (2021). Perceived prototypicality of Asian subgroups in the United States and the United Kingdom. *Journal of Experimental Social Psychology, 97*, 104201. https://doi.org/10.1016/j.jesp.2021.104201

Greil, A. L., McQuillan, J., Shreffler, K. M., Johnson, K. M., & Slauson-Blevins, K. S. (2011). Race-ethnicity and medical services for infertility: Stratified reproduction in a population-based sample of US women. *Journal of Health and Social Behavior, 52*(4), 493-509. https://doi.org/10.1177/0022146511418236

Hammond, M. D., Milojev, P., Huang, Y., & Sibley, C. G. (2018). Benevolent sexism and hostile sexism across the ages. *Social Psychological and Personality Science, 9*(7), 863–874. https://doi.org/10.1177/1948550617727588

Harnois, C. E. (2017). Intersectional masculinities and gendered political consciousness: How do race, ethnicity and sexuality shape men's awareness of gender inequality and support for gender activism? *Sex Roles*, *77*(3), 141-154. https://doi.org/10.1007/s11199-016-0702-2

Harrington, E. F., Crowther, J. H., & Shipherd, J. C. (2010). Trauma, binge eating, and the "strong Black woman". *Journal of Consulting and Clinical Psychology, 78*(4), 469–479. https://doi.org/10.1037/a0019174

Harris, K., Armenta, A. D., Reyna, C., & Zárate, M. A. (2020). Latinx stereotypes: Myths and realities in the twenty-first century. In E. C. Voyles & J. T. Nadler (Eds.), *Stereotypes: The incidence and impacts of bias* (pp. 128-143). Bloomsbury Publishing.

Heaney, M. T. (2019, July 8). Is the Women's March focused on white women — or does it promote intersectional activism? *The Washington Post*.

https://www.washingtonpost.com/politics/2019/07/08/is-womens-march-focused-white-women-or-does-it-promote-intersectional-activism/

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences, 33*(2-3), 111-135. https://doi.org/10.1017/S0140525X10000725

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review, 109*(1), 1-17. https://doi.org/10.1017/S0003055414000604

Hudson, S. K. T. J., Myer, A., & Berney, E. C. (2024). Stereotyping, prejudice, and discrimination at the intersection of race and gender: An intersectional theory primer. *Social and Personality Psychology Compass, 18*(2), e12939. https://doi.org/10.1111/spc3.12939

Hurtado, A. (1989). Relating to privilege: Seduction and rejection in the subordination of white women and women of color. *Signs: Journal of Women in Culture and Society*, *14*(4), 833-855.

Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 166–184. https://doi.org/10.1177/2515245919882903

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist, 74*(2), 171–193. https://doi.org/10.1037/amp0000307

Jackman, M. R. (1994). *The velvet glove: Paternalism and conflict in gender, class, and race relations.* University of California Press.

Johnson, K. L., Freeman, J. B., & Pauker, K. (2011). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102*(1), 116–131. https://doi.org/10.1037/a0025335

Johnson, K. R., & Loscocco, K. (2015). Black marriage through the prism of gender, race, and class. *Journal of Black Studies*, *46*(2), 142–171. https://doi.org/10.1177/0021934714562644

Kane, E. W. (2000). Racial and ethnic variations in gender-related attitudes. *Annual Review of Sociology*, *26*, 419–439.

Kang, S. K., & Bodenhausen, G. V. (2015). Multiple identities in social perception and interaction: Challenges and opportunities. *Annual Review of Psychology, 66*, 547–574. https://doi.org/10.1146/annurev-psych-010814-015025

Kassambara A (2023). rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.7.2, https://CRAN.R-project.org/package=rstatix

Katz, I. (1981). *Stigma: A social-psychological perspective*. Lawrence Erlbaum.

Katz, I. (2014). *Stigma: A social-psychological perspective*. Psychological Press.

King, E. B., Botsford, W., Hebl, M. R., Kazama, S., Dawson, J. F., & Perkins, A. (2012). Benevolent sexism at work: Gender differences in the distribution of challenging developmental experiences. *Journal of Management, 38*(6), 1835-1866. https://doi.org/10.1177/0149206310365902

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford publications.

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: a meta-analysis. *Psychological Bulletin*, *125*(4), 470 –500. https://doi.org/10.1037/0033-2909.125.4.470

Kyriazos, T. A. (2018). Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(08), 86856. https://doi.org/ 10.4236/psych.2018.98126

Landrine, H. (1985). Race × class stereotypes of women. *Sex Roles*, *13*, 65-75. https://doi.org/10.1007/BF00287461

Langone, A. (2018, March 8). #MeToo and Time's Up founders explain the difference between the 2 movements — And how they're alike. *Time*. https://time.com/5189945/whats-the-difference-between-the-metoo-and-times-up-movements/

Lee, J. (2019). Many dimensions of Asian American pan-ethnicity. *Sociology Compass, 13*(12), e12751. https://doi.org/10.1111/soc4.12751

Lee, J., & Ramakrishnan, K. (2020). Who counts as Asian. *Ethnic and Racial Studies, 43*(10), 1733-1756. https://doi.org/10.1080/01419870.2019.1671600

Lee, S. T. H. (2018, September 28). *Testing for measurement invariance: Does your measure mean the same thing for different participants?* Association for Psychological Science. https://www.psychologicalscience.org/observer/testing-for-measurement-invariance

Lee, S. J., & Hong, J. J. (2020). Model minorities and perpetual foreigners: Stereotypes of Asian Americans. In E. C. Voyles & J. T. Nadler (Eds.), Stereotypes: The incidence and impacts of bias (pp. 165-174). Bloomsbury Publishing.

Li, P. (2014). Hitting the ceiling: An examination of barriers to success for Asian American women. *Berkeley Journal of Gender, Law & Justice*, 29, 140-167.

Lin, M. H., Kwan, V. S., Cheung, A., & Fiske, S. T. (2005). Stereotype content model explains prejudice for an envied outgroup: *Scale of anti-Asian American stereotypes. Personality and Social Psychology Bulletin, 31*(1), 34-47. https://doi.org/10.1177/0146167204271320

Littlefield, M. B. (2008). The media as a system of racialization: Exploring images of African American women and the new racism. *American Behavioral Scientist, 51*(5), 675-685. https://doi.org/10.1177/0002764207307747

Liu, T., & Wong, Y. J. (2018). The intersection of race and gender: Asian American men's experience of discrimination. *Psychology of Men & Masculinity, 19*(1), 89–101. https://doi.org/10.1037/men0000084

López, V., & Chesney-Lind, M. (2014). Latina girls speak out: Stereotypes, gender and relationship dynamics. *Latino Studies, 12*, 527-549. https://doi.org/10.1057/lst.2014.54

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*, 1122-1135. https://doi.org/10.3758/s13428-014-0532-5.

Masters, N. T., Lindhorst, T. P., & Meyers, M. K. (2014). Jezebel at the welfare office: How racialized stereotypes of poor women's reproductive decisions and relationships shape policy implementation. *Journal of Poverty, 18*(2), 109-129. https://doi.org/10.1080/10875549.2013.833159

Mays, V. M., & Ghavami, N. (2018). History, aspirations, and transformations of intersectionality: Focusing on gender. In C. B. Travis, J. W. White, Rutherford, A., & Williams, W. S. (Eds.), *APA handbook of the psychology of women: Vol. 1. History, theory, and battlegrounds* (pp. 541–566). American Psychological Association. https://doi.org/10.1037/0000059-028

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society*, *30*(31), 1771–1800. https://doi.org/10.4324/9780203890882

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, *25*(4), 563-579. https://doi.org/10.1177/002200278102500401

McCormick-Huhn, K., Warner, L. R., Settles, I. H., & Shields, S. A. (2019). What if psychology took intersectionality seriously? Changing how psychologists think about participants. *Psychology of Women Quarterly*, *43*(4), 445–456. https://doi.org/10.1177/0361684319866430

McMahon, J. M., & Kahn, K. B. (2016). Benevolent racism? The impact of target race on ambivalent sexism. *Group Processes and Intergroup Relations*, *19*(2), 169–183. https://doi.org/10.1177/1368430215583153

McMahon, J. M., & Kahn, K. B. (2018). When sexism leads to racism: Threat, protecting women, and racial bias. *Sex Roles*, *78*(9–10), 591–605. https://doi.org/10.1007/s11199-017-0828-x

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling: A*

*Multidisciplinary Journal, 14*(4), 611-635. https://doi.org/10.1080/10705510701575461

Merskin, D. (2007). Three faces of Eva: Perpetuation of the hot-Latina stereotype in Desperate Housewives. *The Howard Journal of Communications, 18*(2), 133-151. https://doi.org/10.1080/10646170701309890

Mikołajczak, M., & Pietrzak, J. (2014). Ambivalent sexism and religion: Connected through values. *Sex Roles, 70,* 387-399. https://doi.org/10.1007/s11199-014-0379-3

Miller, A. L., & Borgida, E. (2016). The separate spheres model of gendered inequality. *PLoS ONE, 11*(1), e0147315. https://doi.org/10.1371/journal.pone.0147315

Mukkamala, S., & Suyemoto, K. L. (2018). Racialized sexism/sexualized racism: A multimethod study of intersectional experiences of discrimination for Asian American women. *Asian American Journal of Psychology, 9*(1), 32–46. https://doi.org/10.1037/aap0000104

Nicolas, G., la Fuente, M. de, & Fiske, S. T. (2017). Mind the overlap in multiple categorization: A review of crossed categorization, intersectionality, and multiracial perception. *Group Processes and Intergroup Relations, 20*(5), 621–631. https://doi.org/10.1177/1368430217708862

Niemann, Y. F., Jennings, L., Rozelle, R. M., Baxter, J. C., & Sullivan, E. (1994). Use of free responses and cluster analysis to determine stereotypes of eight groups. *Personality and Social Psychology Bulletin, 20*(4), 379-390. https://doi.org/10.1177/0146167294204005

Pauls, F., Daseking, M., & Petermann, F. (2020). Measurement invariance across gender on the second-order five-factor model of the German Wechsler Intelligence Scale for Children–Fifth Edition. *Assessment, 27*(8), 1836-1852. https://doi.org/10.1177/1073191119847762

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Pérez, E. O., & Hetherington, M. J. (2014). Authoritarianism in black and white: Testing the cross-racial validity of the child rearing scale. *Political Analysis*, *22*(3), 398–412. https://doi.org/10.1093/pan/mpu002

Prasso, S. (2005). *The Asian mystique: Dragon ladies, geisha girls, and our fantasies of the exotic Orient*. Public Affairs.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Pyke, K. D., & Johnson, D. L. (2003). Asian American women and racialized femininities: "Doing" gender across cultural worlds. *Gender & Society*, *17*(1), 33-53. https://doi.org/10.1177/0891243202238977

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reid, P. T., Lewis, L. J., & Wyche, K. F. (2014). An intersectional framework for a multicultural analysis of gender. In F. T. L. Leong, L. Comas-Díaz, G. C.

Nagayama Hall, V. C. McLoyd, & J. E. Trimble (Eds.), *APA handbook of multicultural psychology: Vol. 1. Theory and research* (pp. 379–394). American Psychological Association. https://doi.org/10.1037/14189-020

Rodebaugh, T. L., Heimberg, R. G., Brown, P. J., Fernandez, K. C., Blanco, C., Schneier, F. R., & Liebowitz, M. R. (2011). More reasons to be straightforward: Findings and norms for two scales relevant to social anxiety. *Journal of Anxiety Disorders, 25*(5), 623-630. https://doi.org/10.1016/j.janxdis.2011.02.002

Rollero, C., Glick, P., & Tartaglia, S. (2014). Psychometric properties of short versions of the ambivalent sexism inventory and ambivalence toward men inventory. *Testing, Psychometrics, Methodology in Applied Psychology, 21*(2), 149-159. https://dx.doi.org/10.4473/TPM21.2.3

Roman, E. (2000). Who exactly is living La Vida Loca: The legal and political consequences of Latino-Latina ethnic and racial stereotypes in film and other media. *Journal of Gender, Race & Justice, 4*, 37-68.

Rosenthal, L. (2016). Incorporating intersectionality into psychology: An opportunity to promote social justice and equity. *American Psychologist, 71*(6), 474–485. https://doi.org/10.1037/a0040323

Rosenthal, L., & Lobel, M. (2016). Stereotypes of Black American women related to sexuality and motherhood. *Psychology of Women Quarterly*, *40*(3), 414–427. https://doi.org/10.1177/0361684315627459

Rosenthal, L., Overstreet, N. M., Khukhlovich, A., Brown, B. E., Godfrey, C. J., & Albritton, T. (2020). Content of, sources of, and responses to sexual stereotypes of Black and Latinx women and men in the United States: A qualitative intersectional

exploration. *Journal of Social Issues, 76*(4), 921-948. https://doi.org/10.1111/josi.12411

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rudnev, M., Lytkina, E., Davidov, E., Schmidt, P., & Zick, A. (2018). Testing measurement invariance for a second-order factor: A cross-national test of the Alienation Scale. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology, 12*(1), 47-76. https://doi.org/10.12758/mda.2017.11

Schlegel, A. (n.d.). *Games-Howell post-hoc test*. RPubs by RStudio. https://rpubs.com/aaronsc32/games-howell-test

Sczesny, S., Nater, C., & Eagly, A. H. (2018). Agency and communion: Their implications for gender stereotypes and gender identities. In A. Abele & B Wojciszke (Eds.), *Agency and communion in social psychology* (pp. 103-116). Routledge. https://doi.org/10.4324/9780203703663

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology, 46*(2), 356–360. https://doi.org/10.1016/j.jesp.2009.10.016

Settles, I. H. (2006). Use of an intersectional framework to understand Black women's racial and gender identities. *Sex Roles*, *54*(9–10), 589–601. https://doi.org/10.1007/s11199-006-9029-8

Settles, I. H., Pratt-Hyatt, J. S., & Buchanan, N. T. (2008). Through the lens of race: Black and White women's perceptions of womanhood. *Psychology of Women Quarterly*, *32*(4), 454–468. https://doi.org/10.1111/j.1471-6402.2008.00458.x

Settles, I. H., Warner, L. R., Buchanan, N. C. T., & Jones, M. K. (2020). Understanding psychology's resistance to intersectionality theory using a framework of epistemic exclusion and invisibility. *Journal of Social Issues, 76*(4), 796–813. https://doi.org/10.1111/josi.12403

Shields, S. A. (2007). Passionate men, emotional women: Psychology constructs gender difference in the late 19th century. *History of Psychology*, *10*(2), 92–110. https://doi.org/10.1037/1093-4510.10.2.92

Shields, S. A. (2008). Gender: An intersectionality perspective. *Sex Roles*, *59*(5–6), 301–311. https://doi.org/10.1007/s11199-008-9501-8

Sibley, C. G., & Wilson, M. S. (2004). Differentiating hostile and benevolent sexist attitudes toward positive and negative sexual female subtypes. *Sex Roles, 51*, 687–696. https://doi.org/10.1007/s11199-004-0718-x

Spence, J. T., Helmreich, R., & Stapp, J. (1973). A short version of the Attitudes toward Women Scale (AWS). *Bulletin of the Psychonomic Society*, *2*(4), 219-220. https://doi.org/10.3758/BF03329252

StataCorp. (2017). *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC.

Stenner, K. (2005). *The authoritarian dynamic*. Cambridge University Press. https://doi.org/10.1017/CBO9780511614712

Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist*, *54*(12), 1070-1077. https://doi.org/10.1037/0003-066X.54.12.1070

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68*(2), 199–214. https://doi.org/10.1037/0022-3514.68.2.199.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2, 53-55*. https://doi.org/10.5116/ijme.4dfb.8dfd

Thomas, A. J., Witherspoon, K. M. C., & Speight, S. L. (2004). Toward the development of the stereotypic roles for black women scale. *Journal of Black Psychology*, *30*(3), 426–442. https://doi.org/10.1177/0095798404266061

Trivers, R. (1972). Title of chapter. In E. E. Editor & F. F. Editor (Eds.), Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man: 1871-1971* (pp. 136-179). Aldine Publishing Company.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in Psychology, 6*, Article 1064. https://doi.org/10.3389/fpsyg.2015.01064

Villarreal, A. (2020, March 16). 'Anchor babies': The 'ludicrous' immigration myth that treats people as pawns. *The Guardian*. https://www.theguardian.com/us-news/2020/mar/16/anchor-babies-the-ludicrous-immigration-myth-that-treats-people-as-pawns

Voyles, E. C., & Nadler, J. T. (2020). Intersectionality and future research directions. In E. C. Voyles & J. T. Nadler (Eds.), *Stereotypes: The incidence and impacts of bias* (pp. 299-312). Bloomsbury Publishing.

Warner, L. R., Settles, I. H., & Shields, S. A. (2018). Intersectionality theory in the psychology of women. In C. B. Travis, J. W. White, Rutherford, A., & Williams,

W. S. (Eds.), *APA handbook of the psychology of women: Vol. 1. History, theory, and battlegrounds* (pp. 541–566). American Psychological Association. https://doi.org/10.1037/0000059-027

Warner, L. R., & Shields, S. A. (2018). Intersectionality as a framework for theory and research in feminist psychology. In N. Dess, J. Maracek & L. Bell (Eds)., *Gender, sex, and sexualities: Psychological perspectives* (pp. 29-52). Oxford University Press. https://doi.org/10.1093/oso/9780190658540.003.0002

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weitz, R., & Gordon, L. (1993). Images of Black women among Anglo college students. *Sex Roles, 28*, 19-34. https://doi.org/10.1007/BF00289745

West, C. M. (1995). Mammy, Sapphire, and Jezebel: Historical images of black women and their implications for psychotherapy. *Psychotherapy*, *32*(3), 458–466.

West, C. M. (2008). Mammy, Jezebel, Sapphire, and their homegirls: Developing an "oppositional" gaze toward the images of Black women. In J. C. Chrisler, C. Golden & P. D. Rozee (Eds.), *Lectures on the psychology of women* (4th ed., pp. 287–299). McGraw-Hill.

Woodard, J. B., & Mastin, T. (2005). Black womanhood: *Essence* and its treatment of stereotypical images of Black women. *Journal of Black Studies, 36*(2), 264-281. https://doi.org/10.1177/002193470427315

Yamashita, L. (2022). “I just couldn’t relate to that Asian American narrative”: How Southeast Asian Americans reconsider panethnicity. *Sociology of Race and Ethnicity, 8*(2), 250-266. https://doi.org/10.1177/23326492221078953

Yoo, H. C., Steger, M. F., & Lee, R. M. (2010). Validation of the subtle and blatant racism scale for Asian American college students (SABR-A²). *Cultural Diversity and Ethnic Minority Psychology, 16*(3), 323–334. https://doi.org/10.1037/a0018674

Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of Personality and Social Psychology, 112*(5), 696–717. https://doi.org/10.1037/pspa0000080

**Table 1**

*Sample Size and Mean Age for Each Race-by-Gender Group (Study 1)*

| Group | *N* | Percentage of Sample | Age[†] |
|---|---|---|---|
| Black Men | 247 | 11% | 49.10 (16.07) |
| Black Women | 385 | 18% | 45.71 (15.86) |
| White Men | 731 | 34% | 51.07 (17.78) |
| White Women | 811 | 37% | 51.44 (17.51) |

[†] Mean age is followed by standard deviation in parentheses.

**Table 2**

*Demographic Profile of Final Sample: Education (Study 1)*

| Education | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| No High School | 7.3% | 7.8% | 2.6% | 3.6% | 4.4% |
| High School graduate | 42.5% | 38.7% | 29.7% | 29.7% | 32.8% |
| Some college | 21.5% | 19.2% | 23.5% | 21.0% | 21.6% |
| 2-year degree | 11.7% | 13.0% | 10.4% | 11.0% | 11.2% |
| 4-year degree | 12.2% | 12.5% | 21.5% | 22.1% | 19.0% |
| Post-grad degree | 4.9% | 8.8% | 12.3% | 12.7% | 11.0% |

**Table 3**

*Demographic Profile of Final Sample: Income (Study 1)*

| Family Income | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| Less than $10,000 | 13.0% | 18.2% | 3.7% | 5.9% | 8.1% |
| $10,000 - $19,999 | 16.2% | 12.7% | 5.6% | 8.3% | 9.1% |
| $20,000 - $29,999 | 12.6% | 12.0% | 5.5% | 10.1% | 9.2% |
| $30,000 - $39,999 | 6.5% | 13.0% | 9.7% | 8.8% | 9.6% |
| $40,000 - $49,999 | 8.5% | 6.5% | 9.2% | 8.4% | 8.3% |
| $50,000 - $59,999 | 6.9% | 7.0% | 9.4% | 9.4% | 8.7% |
| $60,000 - $69,999 | 4.5% | 3.4% | 5.3% | 5.7% | 5.0% |
| $70,000 - $79,999 | 6.5% | 3.6% | 6.4% | 6.4% | 5.9% |
| $80,000 - $99,999 | 5.7% | 3.9% | 9.2% | 6.0% | 6.7% |
| $100,000 - $119,999 | 2.0% | 3.1% | 9.0% | 7.6% | 6.7% |
| $120,000 - $149,999 | 2.8% | 1.8% | 6.4% | 3.8% | 4.2% |
| $150,000 - $199,999 | 2.8% | 1.6% | 5.3% | 4.1% | 3.9% |
| $200,000 - $249,999 | 0.4% | 0.5% | 1.5% | 1.7% | 1.3% |
| $250,000 - $349,999 | 0.4% | 0.3% | 1.4% | 0.5% | 0.7% |
| $350,000 - $499,999 | 0.0% | 0.0% | 0.7% | 0.3% | 0.3% |
| $500,000 or more | 0.0% | 0.0% | 0.4% | 0.3% | 0.2% |
| Prefer not to say | 11.3% | 12.5% | 11.2% | 12.8% | 12.1% |

**Table 4**

*Demographic Profile of Final Sample: Political Party Identity (Study 1)*

| Political Party ID | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| Democrat | 58.3% | 68.8% | 30.5% | 37.0% | 42.9% |
| Republican | 7.7% | 4.7% | 28.5% | 30.5% | 22.6% |
| Independent | 28.7% | 18.2% | 31.5% | 23.4% | 25.8% |
| Other | 1.2% | 1.3% | 7.1% | 3.1% | 3.9% |
| Not sure | 4.1% | 7.0% | 2.5% | 6.0% | 4.8% |

**Table 5**

*Demographic Profile of Final Sample: Political Ideology (Study 1)*

| Political Ideology | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| Extremely liberal | 11.7% | 10.7% | 11.9% | 10.6% | 11.2% |
| Liberal | 23.5% | 21.0% | 15.6% | 17.5% | 18.2% |
| Slightly liberal | 7.3% | 8.8% | 8.8% | 8.6% | 8.6% |
| Moderate | 39.3% | 43.6% | 26.0% | 27.4% | 31.1% |
| Slightly conservative | 3.6% | 5.2% | 8.9% | 8.3% | 7.4% |
| Conservative | 9.3% | 7.3% | 18.9% | 17.1% | 15.1% |
| Extremely conservative | 5.3% | 3.4% | 10.0% | 10.5% | 8.5% |

**Table 6**

*Cronbach's Alpha* (α) *Coefficients for Each Scale and Subscale of the ASI by Group (Study 1)*

| Group | AS | HS | BS | INT | PAT | DIFF |
|---|---|---|---|---|---|---|
| Black Men | .86 | .84 | .76 | **.63** | **.53** | .72 |
| Black Women | .83 | .83 | .74 | **.51** | **.54** | .73 |
| White Men | .90 | .92 | .83 | .74 | .74 | .8 |
| White Women | .92 | .92 | .85 | .75 | .71 | .75 |
| Full Sample | .89 | .90 | .83 | .72 | .70 | .77 |
| GF96 Avg | .87 | .88 | .79 | n/a | n/a | n/a |

*Note.* AS = ambivalent sexism scale; HS = hostile sexism scale; BS = benevolent sexism scale; INT = heterosexual intimacy subscale; PAT = protective paternalism subscale; DIFF = complementary gender differentiation subscale; GF96 Avg = the average reliability coefficients reported by Glick and Fiske (1996). The bolded coefficients are highlighted as particularly low compared to conventional levels of acceptable internal consistency.

**Table 7**

*Goodness-of-Fit Statistics for the Full Model by Group (Study 1)*

| Group | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| Black Men | 730.93 | 205 | <.001 | **.757** | **.102** | **.125** |
| Black Women | 739.29 | 205 | <.001 | **.817** | **.082** | **.101** |
| White Men | 1181.71 | 205 | <.001 | **.886** | **.081** | .078 |
| White Women | 1140.75 | 205 | <.001 | **.897** | .075 | .059 |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 8**

*Nonsignificant Factor Loadings for the Configural Model: Unstandardized and (Standardized) (Study 1)*

| Item | Black Men | Black Women | White Men | White Women |
|---|---|---|---|---|
| hs4 | **.154 (.122)** | .303 (.235)* | .720 (.592)* | .856 (.660)* |
| hs10 | -.253 (-.210)* | **-.127 (-.099)** | .329 (.332)* | .354 (.315)* |
| hs11 | **.026 (.022)** | **.000 (.000)** | .599 (.496)* | .833 (.656)* |
| int2 | **.181 (.147)** | **.017 (.014)** | .470 (.445)* | .496 (.442)* |
| pat1 | **.026 (.019)** | **.019 (.013)** | .370 (.293)* | .450 (.366)* |

*Note.* Factor loadings significant at the $\alpha$ =.05 level are marked with an asterisk (*). All nonsignificant loadings (i.e., *p*s are greater than .05) are bolded.

**Table 9**

*Full Wording of Non-Significantly Loading Items (Study 1)*

| Item | Exact Wording |
|---|---|
| hs4 | Feminists are not seeking for women to have more power than men. |
| hs10 | There are actually very few women who get a kick out of teasing men by seeming sexually available and then refusing male advances. |
| hs11 | Feminists are making entirely reasonable demands of men. |
| int2 | People are often truly happy in life without being romantically involved with a member of the other sex. |
| pat1 | In a disaster, women ought not necessarily to be rescued before men. |

**Table 10**

*Revised ASI Model (17 items): Indices of Model Fit (Study 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1368.43 | 460 | <.001 | .952 | .060 | .055 |
| M2: First-Order Metric | 1461.87 | 499 | <.001 | .949 | .060 | .063 |
| M3: Second-Order Metric | 1484.71 | 505 | <.001 | .948 | .060 | .068 |
| M4: First-Order Scalar | 1706.95 | 544 | <.001 | .938 | .063 | .070 |
| M5: Second-Order Scalar | 1824.55 | 553 | <.001 | .932 | .065 | .080 |

**Table 11**

*Revised ASI Model (17 items): Model Comparisons and Changes in Fit (Study 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 93.44 | 39 | <.001 | -.003 | .000 | .008 |
| M3 vs M2 | 22.85 | 6 | <.001 | -.001 | .000 | .005 |
| M4 vs M3 | 222.23 | 39 | <.001 | -.010† | .003 | .002 |
| M5 vs M4 | 117.60 | 9 | <.001 | -.006† | .002 | .010 |

† These changes in CFI are potentially concerning for MI, with the strict cutoff for change in CFI being .005 and the more lenient cutoff being .010 (Chen, 2007).

**Table 12**

*White Men and White Women Only: Indices of Model Fit (Study 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 2322.46 | 410 | <.001 | .891 | .078 | .068 |
| M2: First-Order Metric | 2379.21 | 428 | <.001 | .889 | .077 | .072 |
| M3: Second-Order Metric | 2389.86 | 430 | <.001 | .889 | .077 | .074 |
| M4: First-Order Scalar | 2555.33 | 448 | <.001 | .880 | .078 | .076 |
| M5: Second-Order Scalar | 2616.49 | 451 | <.001 | .877 | .079 | .081 |

**Table 13**

*White Men and White Women Only: Model Comparisons (Study 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 56.75 | 18 | <.001 | -.002 | -.001 | .004 |
| M3 vs M2 | 10.64 | 2 | <.001 | .000 | .000 | .002 |
| M4 vs M3 | 165.47 | 18 | <.001 | -.009† | .001 | .002 |
| M5 vs M4 | 61.17 | 3 | <.001 | -.003 | .001 | .005‡ |

† The change in CFI is potentially problematic for MI, with the strict cutoff for change in CFI being .005 and the more lenient cutoff being .010. ‡ The change in SRMR between the first- and second-order scalar models is equal to the recommended strict cutoff of .050; however, it is below the less strict cutoff, .010.

**Table 14**

*Pearson Correlation Coefficients (*rs*) for Identity Centrality Composites (*Importance *and* Self-View*) by Group (Study 1)*

| Group | Black Identity | White Identity | Man Identity | Woman Identity |
|---|---|---|---|---|
| Black Men | .81 | | .75 | |
| Black Women | .74 | | | .79 |
| White Men | | .57 | .66 | |
| White Women | | .55 | | .69 |
| Combined Group | .78[a] | .56[b] | .70[c] | .73[d] |

*Note.* All correlations are significant, all *p*s < .001.

Combined Groups: [a] Black women and Black men. [b] White men and White women. [c] Black men and White men. [d] Black women and White women.

**Table 15**

*Identity Centrality Means (SDs) by Group (Study 1)*

| Group | Racial Identity Centrality† | Gender Identity Centrality‡ |
|---|---|---|
| Black Men | 4.26 (1.12)[ab] | 4.38 (0.96)[ab] |
| Black Women | 4.35 (1.03)[cd] | 4.47 (0.89)[cd] |
| White Men | 2.84 (1.29)[ac] | 3.49 (1.20)[ace] |
| White Women | 2.93 (1.24)[bd] | 3.93 (1.08)[bde] |
| Full Sample | 3.30 (1.37) | 3.93 (1.14) |

*Note.* Significantly different means (all *p*s <.001) within each column are marked by shared superscript letters. For example, within the Racial Identity Centrality column, the mean for Black Men is significantly greater than the mean for White Men (superscript *a*) as well as the mean for White Women (superscript *b*), but it is not significantly different from the mean for Black Women.

† Black Identity or White Identity, as appropriate. ‡ Man Identity or Woman Identity, as appropriate.

**Table 16**

*"To what extent were the women and men you had in mind [racial group]?": Means (SDs) by Group (Study 1)*

| Group | White | Black |
|---|---|---|
| Black Men | 2.34 (1.10)[a] | 2.65 (1.13)[ab] |
| Black Women | 2.25 (1.10)[bc] | 2.65 (1.18)[cd] |
| White Men | 2.51 (1.23)[b] | 1.67 (0.92)[ac] |
| White Women | 2.62 (1.19)[ac] | 1.73 (0.90)[bd] |
| Full Sample | 2.49 (1.18) | 1.98 (1.08) |

*Note.* Significantly different means (all *p*s <.01) within each column are marked by shared superscript letters. For example, within the Black column, the mean for Black Men is significantly greater than the mean for White Men (superscript *a*) as well as the mean for White Women (superscript *b*), but it is not significantly different from the mean for Black Women.

**Table 17**

*Sample Size and Mean Age for Each Race-by-Gender Group (Study 2)*

| Group | *N* | Percentage of Sample | Age[†] |
|---|---|---|---|
| Black Men | 412 | 24% | 36.42 (11.33) |
| Black Women | 401 | 23% | 39.62 (14.24) |
| White Men | 467 | 27% | 40.47 (12.41) |
| White Women | 453 | 26% | 44.85 (14.54) |

[†] Mean age is followed by standard deviation in parentheses.

**Table 18**

*Demographic Profile of Final Sample: Education (Study 2)*

| Education | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| Less than high school | 1.2% | 0.5% | 0.9% | 1.6% | 1.0% |
| High school diploma or equivalent | 16.0% | 12.7% | 16.3% | 11.5% | 14.1% |
| Some college | 21.4% | 24.4% | 18.8% | 21.6% | 21.5% |
| Associate's degree | 6.1% | 13.2% | 9.6% | 12.4% | 10.3% |
| Bachelor's degree | 41.3% | 35.2% | 38.3% | 36.0% | 37.7% |
| Master's degree | 12.6% | 12.5% | 12.0% | 14.1% | 12.8% |
| Advanced Degree | 1.5% | 1.5% | 4.1% | 2.9% | 2.5% |

**Table 19**

*Demographic Profile of Final Sample: Income (Study 2)*

| Household Income | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| Less than $10,000 | 7.3% | 7.8% | 4.3% | 2.7% | 5.4% |
| $10,000 - $14,999 | 2.9% | 5.3% | 3.6% | 2.9% | 3.6% |
| $15,000 - $24,999 | 7.8% | 10.5% | 7.3% | 7.7% | 8.3% |
| $25,000 - $34,999 | 9.5% | 13.8% | 12.4% | 11.0% | 11.7% |
| $35,000 - $49,999 | 11.7% | 13.8% | 12.4% | 18.1% | 14.0% |
| $50,000 - $74,999 | 21.1% | 22.0% | 20.3% | 22.5% | 21.5% |
| $75,000 - $99,999 | 13.1% | 11.8% | 13.9% | 13.9% | 13.2% |
| $100,000 - $149,999 | 20.9% | 11.5% | 16.5% | 13.7% | 15.7% |
| $150,000 - $199,999 | 3.9% | 3.0% | 4.1% | 3.8% | 3.7% |
| $200,000 or more | 1.9% | 0.8% | 5.1% | 3.8% | 3.0% |

**Table 20**

*Demographic Profile of Final Sample: Political Party Identity (Study 2)*

| Political Party ID | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| Democrat | 7.5% | 3.5% | 21.4% | 20.3% | 13.7% |
| Republican | 58.3% | 64.1% | 44.5% | 52.8% | 54.5% |
| Independent | 29.9% | 25.7% | 31.3% | 22.3% | 27.3% |
| Other | 4.4% | 6.7% | 2.8% | 4.6% | 4.6% |

**Table 21**

*Demographic Profile of Final Sample: Political Ideology (Study 2)*

| Political Ideology | Black Men | Black Women | White Men | White Women | Total |
|---|---|---|---|---|---|
| Extremely liberal | 12.1% | 17.5% | 16.3% | 17.7% | 15.9% |
| Liberal | 27.4% | 32.2% | 24.0% | 28.3% | 27.8% |
| Slightly liberal | 14.3% | 13.2% | 12.4% | 14.8% | 13.7% |
| Moderate | 31.1% | 23.9% | 19.1% | 15.9% | 22.2% |
| Slightly conservative | 6.3% | 7.5% | 12.2% | 8.2% | 8.7% |
| Conservative | 5.1% | 4.7% | 12.2% | 12.4% | 8.8% |
| Extremely conservative | 3.6% | 1.0% | 3.9% | 2.9% | 2.9% |

**Table 22**

*Number and Percentage of Participants in Each Experimental Condition, by Race-by-Gender Group (Study 2)*

| Group | Experimental Condition | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | White | | Black | | Asian | | Latinx | | Control | | | |
| | *N* | % | *N* | % | *N* | % | *N* | % | *N* | % | *N* | % |
| Black Men | 83 | 4.8% | 81 | 4.7% | 82 | 4.7% | 84 | 4.8% | 82 | 4.7% | 412 | 23.8% |
| Black Women | 79 | 4.6% | 79 | 4.6% | 80 | 4.6% | 83 | 4.8% | 80 | 4.6% | 401 | 23.1% |
| White Men | 94 | 5.4% | 93 | 5.4% | 94 | 5.4% | 94 | 5.4% | 92 | 5.3% | 467 | 26.9% |
| White Women | 91 | 5.3% | 91 | 5.3% | 91 | 5.3% | 90 | 5.2% | 90 | 5.2% | 453 | 26.1% |
| Total | 347 | 20.0% | 344 | 19.8% | 347 | 20.0% | 351 | 20.3% | 344 | 19.8% | 1733 | 100.0% |

**Table 23**

*Manipulation Check Failures by Condition and Race-by-Gender Group (Study 2)*

| Group | Experimental Condition | | | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | White | | Black | | Asian | | Latinx | | Control | | | |
| | *N* | % | *N* | % | *N* | % | *N* | % | *N* | % | *N* | % |
| Black Men | 29 | 10.0% | 12 | 4.1% | 23 | 7.9% | 31 | 10.7% | 11 | 3.8% | 106 | 36.4% |
| Black Women | 20 | 6.9% | 9 | 3.1% | 13 | 4.5% | 19 | 6.5% | 4 | 1.4% | 65 | 22.3% |
| White Men | 24 | 8.2% | 11 | 3.8% | 13 | 4.5% | 11 | 3.8% | 3 | 1.0% | 62 | 21.3% |
| White Women | 14 | 4.8% | 10 | 3.4% | 19 | 6.5% | 14 | 4.8% | 1 | 0.3% | 58 | 19.9% |
| Total | 87 | 29.9% | 42 | 14.4% | 68 | 23.4% | 75 | 25.8% | 19 | 6.5% | 291 | 100.0% |

**Table 24**

*Cronbach's Alpha (α) Coefficients for Each Scale and Subscale of the ASI by Group (Study 2 – Time 1)*

| Group | AS | HS | BS | INT | PAT | DIFF |
|---|---|---|---|---|---|---|
| Black Men | .87 | .90 | .82 | .72 | **.65** | .82 |
| Black Women | .89 | .89 | .84 | .73 | **.64** | .82 |
| White Men | .92 | .95 | .89 | .82 | .82 | .88 |
| White Women | .93 | .91 | .89 | .80 | .78 | .85 |
| Full Sample | .92 | .93 | .88 | .79 | .76 | .85 |
| GF96 Avg | .87 | .88 | .79 | n/a | n/a | n/a |

*Note.* AS = ambivalent sexism scale; HS = hostile sexism scale; BS = benevolent sexism scale; INT = heterosexual intimacy subscale; PAT = protective paternalism subscale; DIFF = complementary gender differentiation subscale; GF96 Avg = the average reliability coefficients reported by Glick and Fiske (1996). The bolded coefficients are highlighted as particularly low (i.e., not meeting typical levels of acceptance).

**Table 25**

*Goodness-of-Fit Statistics for the Full Model by Group (Study 2 – Time 1)*

| Group | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| Black Men | 728.37 | 205 | <.001 | **.876** | .079 | .072 |
| Black Women | 543.38 | 205 | <.001 | .907 | .064 | .065 |
| White Men | 703.98 | 205 | <.001 | .932 | .072 | .059 |
| White Women | 713.18 | 205 | <.001 | .908 | .074 | .057 |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 26**

*Full Model (22 items): Indices of Invariance Model Fit (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 2688.91 | 820 | <.001 | .910 | .073 | .063 |
| M2: First-Order Metric | 2909.11 | 874 | <.001 | .902 | .073 | .076 |
| M3: Second-Order Metric | 2935.46 | 880 | <.001 | .901 | .073 | .078 |
| M4: First-Order Scalar | 3260.39 | 934 | <.001 | **.888** | .076 | **.081** |
| M5: Second-Order Scalar | 3562.38 | 943 | <.001 | **.873** | .08 | **.118** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 27**

*Full ASI Model (22 items): Invariance Model Comparisons and Changes in Fit (Study 2 – Time 1)*

| Comparison | **Δ**$\chi$2 | **Δ***df* | *p*-value | **Δ**CFI | **Δ**RMSEA | **Δ**SRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 220.20 | 54 | <.001 | *-.008* | .000 | .013 |
| M3 vs M2 | 26.36 | 6 | <.001 | -.001 | .000 | .002 |
| M4 vs M3 | 324.93 | 54 | <.001 | **-.013** | .003 | .003 |
| M5 vs M4 | 301.99 | 9 | <.001 | **-.015** | .004 | **.037** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 28**

*More and Less Conservative Cutoffs for Acceptable Changes in Fit*

| Test Statistic | More Conservative | Less Conservative |
|---|---|---|
| **Δ**CFI | > .005 | > .010 |
| **Δ**RMSEA | > .010 | > .015 |
| **Δ**SRMR - Metric | > .025 | > .030 |
| **Δ**SRMR - Scalar | > .005 | > .010 |

*Note.* Based on Chen (2007).

**Table 29**

*Indices of Invariance Model Fit: Black Men vs. Black Women (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1271.75 | 410 | <.001 | **.890** | .072 | .069 |
| M2: First-Order Metric | 1316.33 | 428 | <.001 | **.887** | .071 | .074 |
| M3: Second-Order Metric | 1318.10 | 430 | <.001 | **.887** | .071 | .074 |
| M4: First-Order Scalar | 1405.51 | 448 | <.001 | **.878** | .073 | .076 |
| M5: Second-Order Scalar | 1593.86 | 451 | <.001 | **.854** | .079 | **.121** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 30**

*Model Comparisons and Changes in Fit: Black Men vs. Black Women (Study 2 – Time 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 44.58 | 18 | <.001 | -.003 | -.001 | .005 |
| M3 vs M2 | 1.77 | 2 | .413 | .000 | .000 | .000 |
| M4 vs M3 | 87.41 | 18 | <.001 | *-.009* | .002 | .002 |
| M5 vs M4 | 188.36 | 3 | <.001 | **-.024** | .006 | **.045** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 31**

*Indices of Invariance Model Fit: Black Men vs. White Men (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1432.35 | 410 | <.001 | .910 | .075 | .065 |
| M2: First-Order Metric | 1504.73 | 428 | <.001 | .907 | .076 | .077 |
| M3: Second-Order Metric | 1510.43 | 430 | <.001 | .906 | .076 | .078 |
| M4: First-Order Scalar | 1578.55 | 448 | <.001 | .902 | .076 | .079 |
| M5: Second-Order Scalar | 1635.88 | 451 | <.001 | **.897** | .077 | **.092** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 32**

*Model Comparisons and Changes in Fit: Black Men vs. White Men (Study 2 – Time 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 72.37 | 18 | <.001 | -.003 | .001 | .012 |
| M3 vs M2 | 5.71 | 2 | <.001 | -.001 | .000 | .001 |
| M4 vs M3 | 68.11 | 18 | <.001 | -.004 | .000 | .001 |
| M5 vs M4 | 57.33 | 3 | <.001 | -.005 | .001 | **.013** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 33**

*Indices of Invariance Model Fit: Black Men vs. White Women (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1441.55 | 410 | <.001 | **.894** | .070 | .064 |
| M2: First-Order Metric | 1562.70 | 428 | <.001 | **.884** | .078 | .079 |
| M3: Second-Order Metric | 1570.18 | 430 | <.001 | **.883** | .078 | .080 |
| M4: First-Order Scalar | 1669.46 | 448 | <.001 | **.875** | .079 | **.082** |
| M5: Second-Order Scalar | 1900.05 | 451 | <.001 | **.851** | .080 | **.148** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 34**

*Model Comparisons and Changes in Fit: Black Men vs. White Women (Study 2 – Time 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 121.15 | 18 | <.001 | -.010 | .008 | .015 |
| M3 vs M2 | 7.48 | 2 | .024 | -.001 | .000 | .001 |
| M4 vs M3 | 99.28 | 18 | <.001 | *-.008* | .001 | .002 |
| M5 vs M4 | 230.59 | 3 | <.001 | **-.024** | .001 | **.066** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 35**

*Indices of Invariance Model Fit: Black Women vs. White Men vs. White Women (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1960.54 | 615 | <.001 | .918 | .070 | .060 |
| M2: First-Order Metric | 2094.65 | 651 | <.001 | .912 | .071 | .068 |
| M3: Second-Order Metric | 2117.71 | 655 | <.001 | .911 | .071 | .071 |
| M4: First-Order Scalar | 2386.71 | 691 | <.001 | **.897** | .075 | .074 |
| M5: Second-Order Scalar | 2480.21 | 697 | <.001 | **.892** | .076 | **.085** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 36**

*Model Comparisons and Changes in Fit: Black Women vs. White Men vs. White Women (Study 2 – Time 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 134.11 | 36 | <.001 | *-.006* | .001 | .008 |
| M3 vs M2 | 23.07 | 4 | .413 | -.001 | .000 | .003 |
| M4 vs M3 | 269.00 | 36 | <.001 | **-.014** | .004 | .003 |
| M5 vs M4 | 93.50 | 6 | <.001 | -.005 | .001 | **.011** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 37**

*Indices of Invariance Model Fit: Black Women vs. White Women (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1256.56 | 410 | <.001 | .907 | .070 | .061 |
| M2: First-Order Metric | 1327.24 | 428 | <.001 | .902 | .070 | .068 |
| M3: Second-Order Metric | 1342.93 | 430 | <.001 | .900 | .071 | .071 |
| M4: First-Order Scalar | 1416.35 | 448 | <.001 | **.894** | .071 | .073 |
| M5: Second-Order Scalar | 1442.68 | 451 | <.001 | **.892** | .072 | .075 |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 38**

*Model Comparisons and Changes in Fit: Black Women vs. White Women (Study 2 – Time 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 70.68 | 18 | <.001 | -.005 | .000 | .007 |
| M3 vs M2 | 15.69 | 2 | .413 | -.002 | .001 | .003 |
| M4 vs M3 | 73.42 | 18 | <.001 | *-.006* | .000 | .002 |
| M5 vs M4 | 26.34 | 3 | <.001 | -.002 | .001 | .002 |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 39**

*Indices of Invariance Model Fit: White Men vs. White Women (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1417.16 | 410 | <.001 | .922 | .073 | .058 |
| M2: First-Order Metric | 1490.73 | 428 | <.001 | .917 | .073 | .062 |
| M3: Second-Order Metric | 1494.54 | 430 | <.001 | .917 | .073 | .064 |
| M4: First-Order Scalar | 1616.51 | 448 | <.001 | .909 | .075 | .065 |
| M5: Second-Order Scalar | 1680.98 | 451 | <.001 | .904 | .077 | .080 |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 40**

*Model Comparisons and Changes in Fit: White Men vs. White Women (Study 2 – Time 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 73.58 | 18 | <.001 | -.005 | .000 | .004 |
| M3 vs M2 | 3.81 | 2 | .149 | .000 | .000 | .002 |
| M4 vs M3 | 121.97 | 18 | <.001 | *-.008* | .002 | .001 |
| M5 vs M4 | 64.47 | 3 | <.001 | -.005 | .002 | **.015** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 41**

*Indices of Invariance Model Fit: Black Women vs. White Men (Study 2 – Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1247.36 | 410 | <.001 | .923 | .069 | .062 |
| M2: First-Order Metric | 1304.36 | 428 | <.001 | .920 | .069 | .068 |
| M3: Second-Order Metric | 1319.54 | 430 | <.001 | .919 | .069 | .071 |
| M4: First-Order Scalar | 1524.31 | 448 | <.001 | .902 | .074 | .076 |
| M5: Second-Order Scalar | 1583.78 | 451 | <.001 | **.896** | .076 | **.084** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table 42**

*Model Comparisons and Changes in Fit: Black Women vs. White Men (Study 2 – Time 1)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 57.00 | 18 | <.001 | -.003 | .000 | .006 |
| M3 vs M2 | 15.18 | 2 | <.001 | -.001 | .000 | .003 |
| M4 vs M3 | 204.77 | 18 | <.001 | **-.017** | .005 | .005 |
| M5 vs M4 | 59.47 | 3 | <.001 | *-.006* | .002 | *.008* |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table 43**

*Time 1 HS and BS Scores: Means (SDs) by Race-by-Gender Group (Study 2)*

| Group | HS | BS |
|---|---|---|
| Black Women | 2.38 (1.01)[a] | 3.55 (0.95)[a] |
| White Women | 2.22 (1.02)[ab] | 2.81 (1.04)[ab] |
| White Men | 2.77 (1.27)[b] | 3.20 (1.07)[b] |

*Note.* Significantly different means (all *p*s < .02) within each column are marked by shared superscript letters. For example, within the HS column, the mean for White men is significantly greater than the mean for White women (superscript *a*). Black women and White men were not directly compared with one another on these measures.

**Table 44**

*Time 2 HS Means (and SDs) by Race-by-Gender Group and Experimental Condition (Study 2)*

| Group | Experimental Condition | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | White | | Black | | Asian | | Latinx | | Control | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Black Women | 2.50 | 1.12 | 2.32 | 0.96 | 2.48 | 1.02 | 2.29 | 0.99 | 2.49 | 1.11 |
| White Women | 2.20 | 1.07 | 2.14 | 0.94 | 2.32 | 0.95 | 2.05 | 1.01 | 2.41 | 1.09 |
| White Men | 2.81 | 1.30 | 2.70 | 1.23 | 2.66 | 1.28 | 2.75 | 1.24 | 2.91 | 1.40 |

**Table 45**

*Time 2 BS Means (and SDs) by Race-by-Gender Group and Experimental Condition (Study 2)*

| Group | Experimental Condition | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | White | | Black | | Asian | | Latinx | | Control | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Black Women | 3.55 | 1.13 | 3.52 | 0.96 | 3.82 | 0.96 | 3.61 | 0.89 | 3.88 | 1.00 |
| White Women | 2.77 | 1.14 | 2.77 | 0.98 | 2.97 | 1.00 | 2.99 | 1.13 | 2.84 | 0.97 |
| White Men | 3.07 | 1.07 | 3.23 | 1.14 | 3.27 | 1.06 | 3.31 | 1.13 | 3.18 | 1.06 |

**Table 46**

*Study 2 Hypotheses: Support by Group*

| Group | 2.1a | 2.1b | 2.2a | 2.2b | 2.3a | 2.3b | 2.4a | 2.4b |
|---|---|---|---|---|---|---|---|---|
| Black Women | Yes | Yes | No | No | No | No | No | No |
| White Women | Yes | Yes | No | No | No | No | No | No |
| White Men | Yes | Yes | No | No | No | No | No | No |

**Table 47**

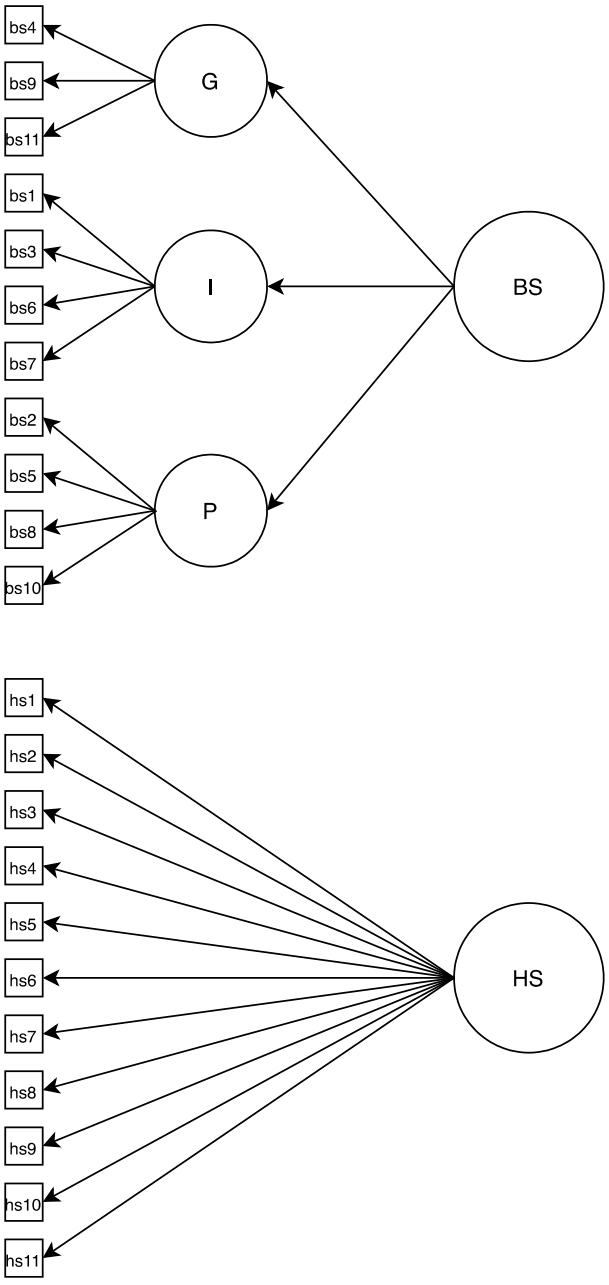
*Mean FT Ratings for Four Racial Groups by Participant Race-by-Gender Groups, as Measured at Time 1 (Study 2)*

| Group | White FT | Black FT | Asian FT | Latinx FT |
|---|---|---|---|---|
| Black Women | 55.21 (25.10)$^{ab}$ | 85.47 (16.25)$^{ab}$ | 68.40 (20.91)$^{a}$ | 70.33 (21.58)$^{a}$ |
| White Women | 71.88 (20.22)$^{a}$ | 74.08 (21.09)$^{ac}$ | 75.60 (20.40)$^{ab}$ | 74.07 (20.93)$^{ab}$ |
| White Men | 68.96 (20.82)$^{b}$ | 65.97 (22.93)$^{bc}$ | 70.61 (19.23)$^{b}$ | 68.23 (21.13)$^{b}$ |

*Note.* Significantly different means (all *p*s < .03) within each column are marked by shared superscript letters. For example, within the White FT column, the mean for Black women is significantly smaller than the mean for White women (superscript *a*) and the mean for White men (superscript *b*).

**Figure 1**

*Measurement Model of the Ambivalent Sexism Inventory*




*Note.* The model contains four first-order factors: complementary gender differentiation (G), heterosexual intimacy (I), protective paternalism (P), and hostile sexism (HS). The lone second-order factor, benevolent sexism (BS), is made up of three first-order factors: G, I, and P.

**Figure 2**

*"To what extent were the women and men you had in mind White/Caucasian?": Frequency of Each Response Option by Group (Study 1)*

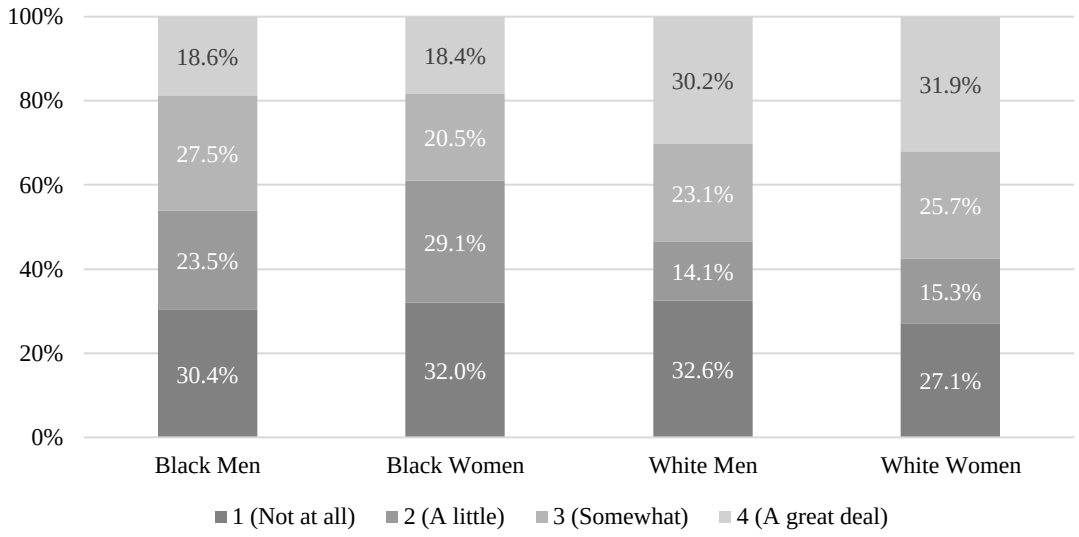


**Figure 3**

*"To what extent were the women and men you had in mind Black/African American?": Frequency of Each Response Option by Group (Study 1)*

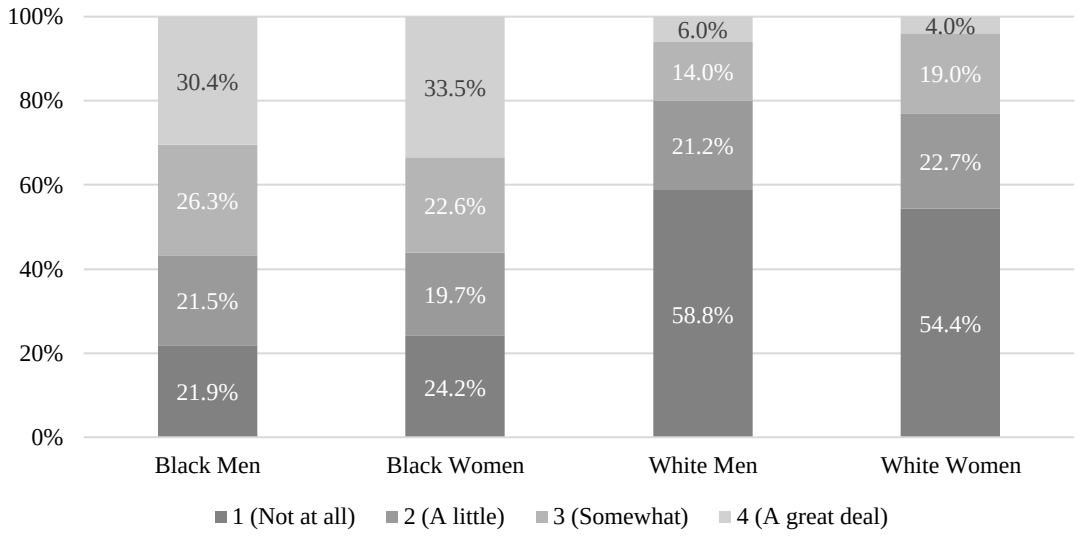


**Figure 4**

*ASI Instruction Screen in Phase 1 Survey and for Control Condition in Phase 2 Experiment (Study 2)*

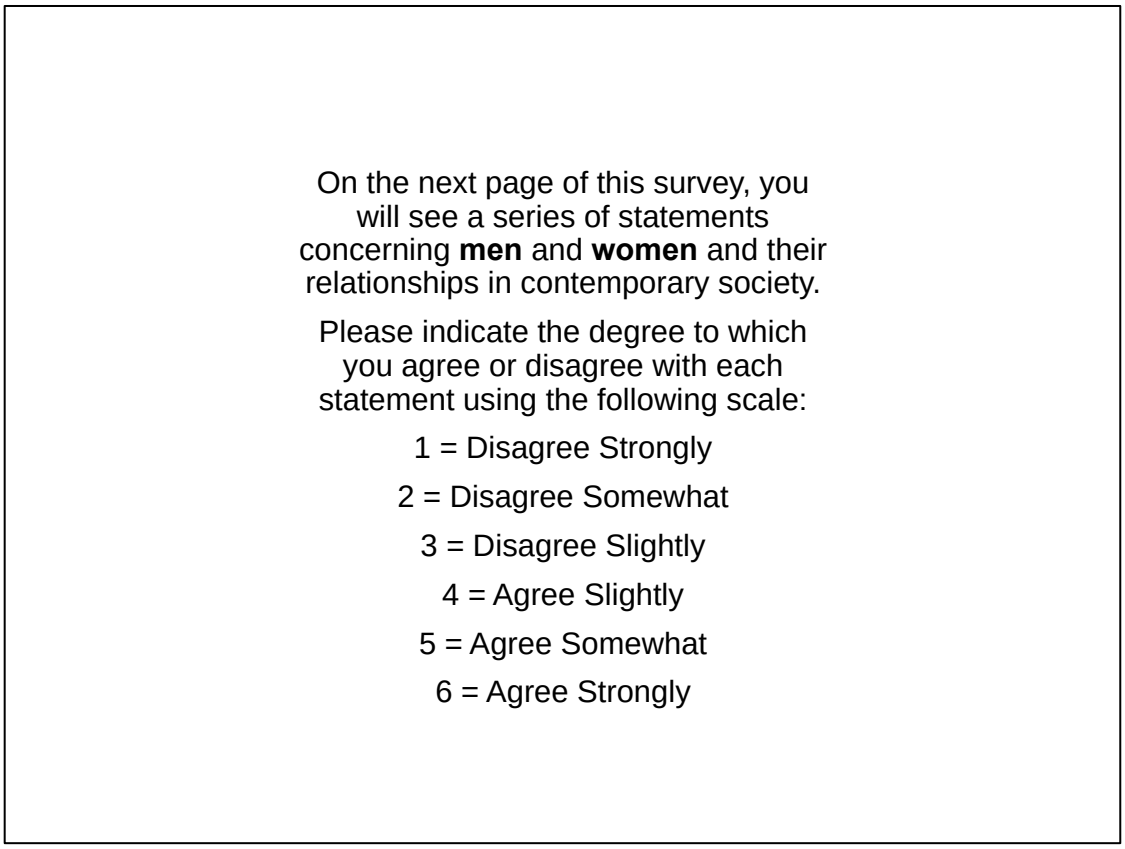
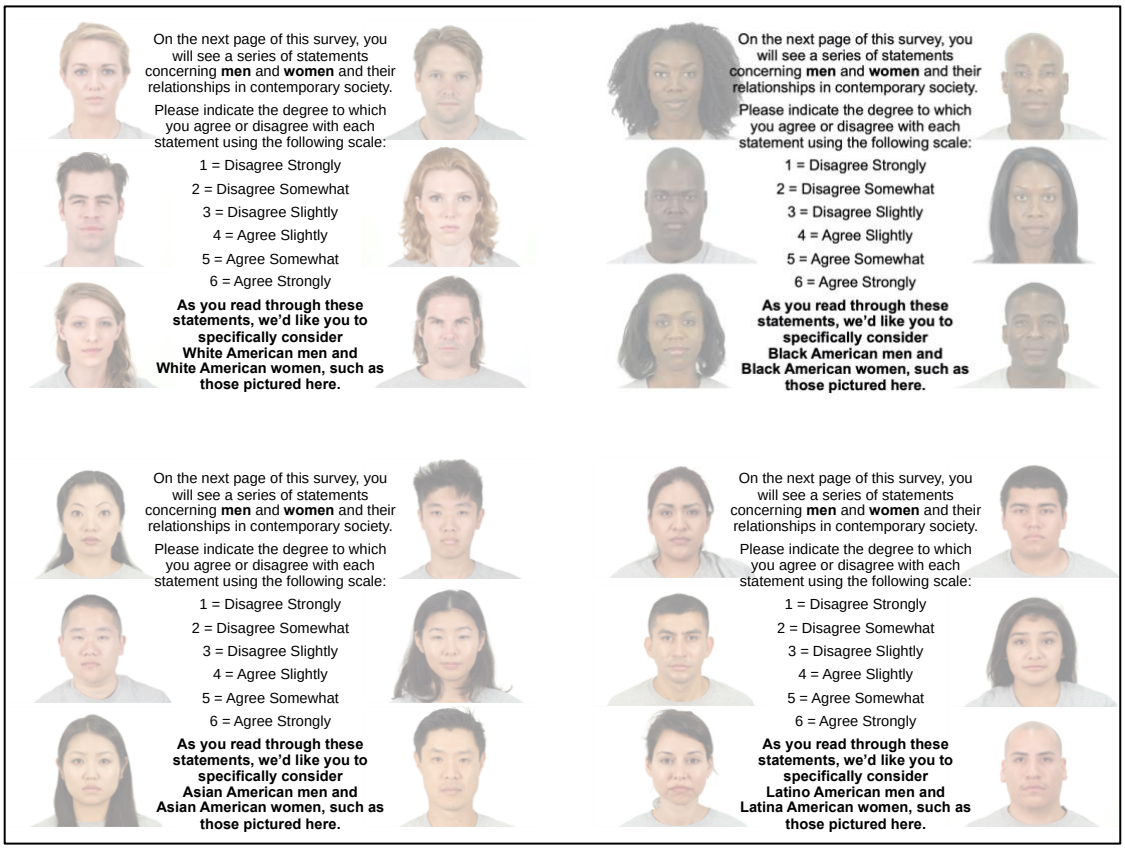


**Figure 5**

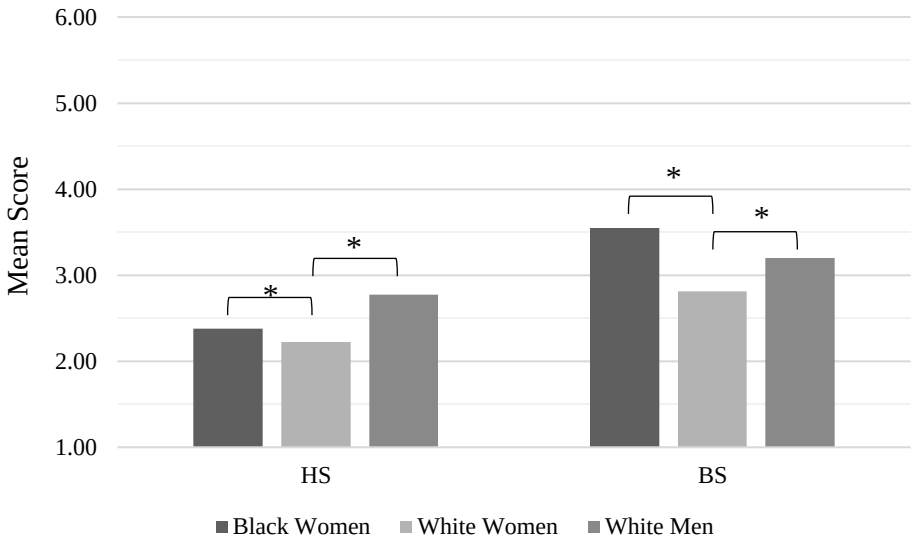
*Racial Primes in the Phase 2 Experiment (Study 2)*




*Note*. Clockwise from the upper-left corner is the White prime, Black prime, Latino/a prime, and Asian prime. All images are from the Chicago Face Database (Version 3.0; Ma et al., 2015).

**Figure 6**

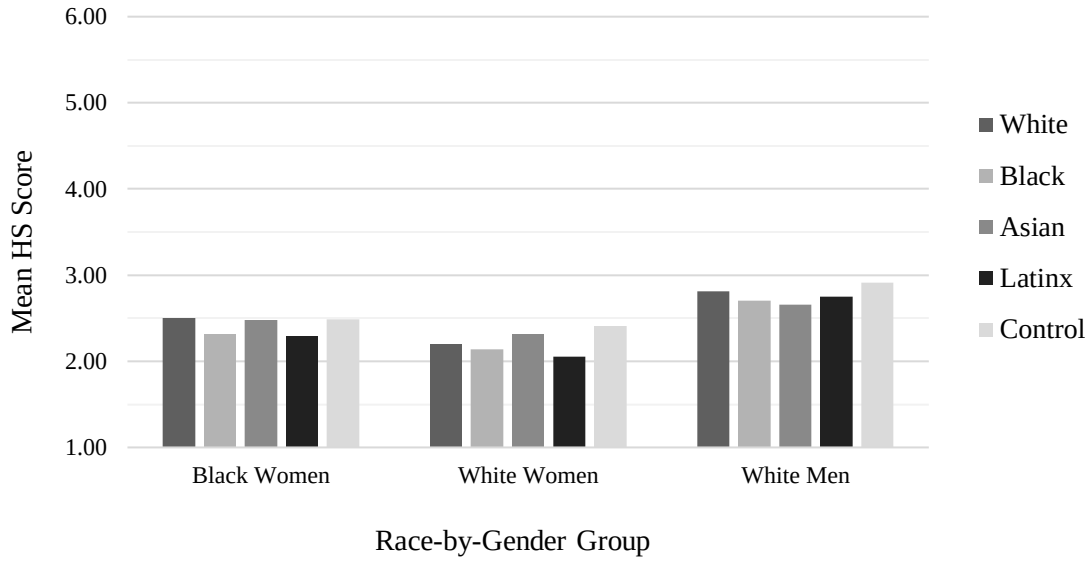
*Time 1 HS and BS Scores: Means by Race-by-Gender Group (Study 2)*




*Note.* Pairs of bars (i.e., means) that are significantly different from each other (all *p*s < .02) are marked with asterisks (*). For example, the HS mean for White men is significantly greater than the HS mean for White women. Black women and White men were not directly compared with one another on these measures.

**Figure 7**

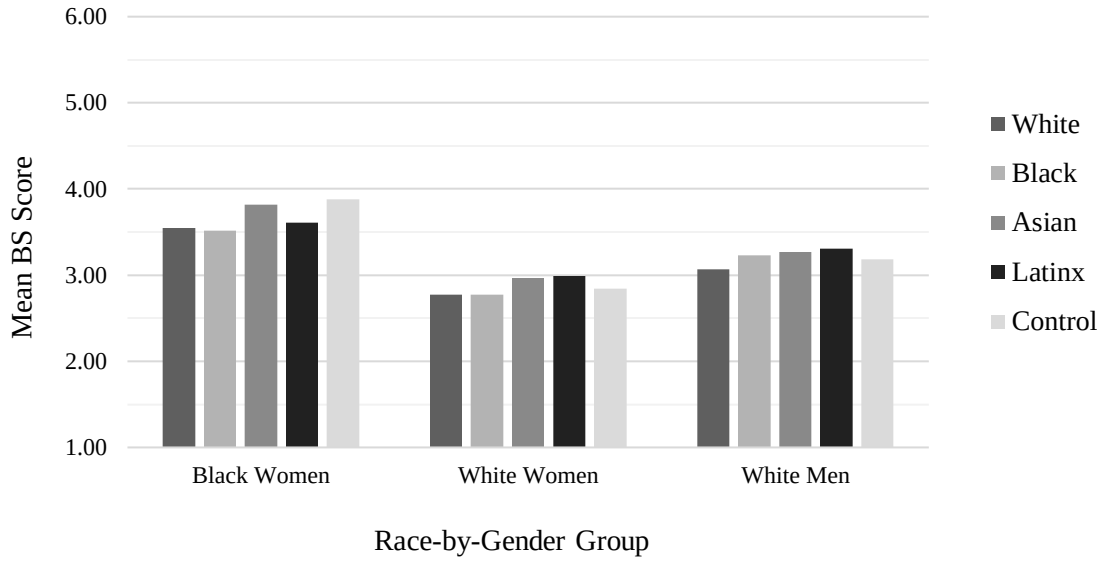
*Time 2 HS Means (and SDs) by Race-by-Gender Group and Experimental Condition (Study 2)*




*Note.* Within each race-by-gender group, there were no significant differences between any of the mean bars.

**Figure 8**

*Time 2 BS Means (and SDs) by Race-by-Gender Group and Experimental Condition (Study 2)*




*Note.* Within each race-by-gender group, there were no significant differences between any of the mean bars.

# Appendix

The Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996)

HS Items

1. Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for "equality."
2. Most women interpret innocent remarks or acts as being sexist.
3. Women are too easily offended.
4. *Feminists are not seeking for women to have more power than men.
5. Most women fail to appreciate fully all that men do for them
6. Women seek to gain power by getting control over men.
7. Women exaggerate problems they have at work.
8. Once a woman gets a man to commit to her, she usually tries to put him on a tight leash.
9. When women lose to men in a fair competition, they typically complain about being discriminated against.
10. *There are actually very few women who get a kick out of teasing men by seeming sexually available and then refusing male advances.
11. *Feminists are making entirely reasonable demands of men.

BS Items

1. No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman. (I)
2. *In a disaster, women ought not necessarily to be rescued before men. (P)

3. *People are often truly happy in life without being romantically involved with a member of the other sex. (I)
4. Many women have a quality of purity that few men possess. (G)
5. Women should be cherished and protected by men. (P)
6. Every man ought to have a woman whom he adores. (I)
7. *Men are complete without women. (I)
8. A good woman should be set on a pedestal by a man
9. Women, compared to men, tend to have a superior moral sensibility. (G)
10. Men should be willing to sacrifice their own well being in order to provide financially for the women in their lives. (P)
11. Women, as compared to men, tend to have a more refined sense of culture and taste. (G)

* = Item is reverse-scored

G = Complementary gender differentiation

P = Protective paternalism

I = Heterosexual intimacy

# Supplemental Materials

## Part 1: Study 1 One-Factor Model, Two-Factor Model, and Full Factor Model Fit by Group

As noted in the main text (Chapter 2 *Results* section, *Measurement Invariance/Initial Steps* subsection), the fit of each possible ASI model – one-factor, two-factor, and full – was checked for each of the four race-by-gender groups. For each group, as well as the full sample, the full model proved to best fit the data, followed by the two-factor model, and finally the one-factor model. However, the full model did not have acceptable fit indices for all groups. The full model results for each group was reported in the main text; the fit statistics for the additional two models are reported in the here in Tables S1-S5 (one table for the full sample and one table per race-by-gender group).

The one-factor model consists of a single ambivalent sexism (AS) factor measured by 22 items. The two-factor model consists of two factors, hostile sexism (HS) and benevolent sexism (BS), each measured by 11 items. The full model (see Figure 1 in the main text) contains four first-order factors: complementary gender differentiation (G), heterosexual intimacy (I), protective paternalism (P), and HS. The lone second-order factor, HS, is made up of three first-order factors: G, I, and P. The HS factor is measured by 11 items, the G factor by 3 items, the I factor by 4 items, and the P factor by 4 items.

**Table S1**

*Fit Indices for Three ASI Models: Full Sample (Study 1)*

| Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| One-factor | 8229.44 | 209 | <.001 | .647 | .133 | .127 |
| Two-factor | 4016.81 | 208 | <.001 | .833 | .092 | .080 |
| Full | 3090.10 | 205 | <.001 | .873 | .080 | .075 |

**Table S2**

*Fit Indices for Three ASI Models: Black Men (Study 1)*

| Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| One-factor | 1133.90 | 209 | <.001 | .573 | .134 | .142 |
| Two-factor | 777.83 | 208 | <.001 | .737 | .105 | .125 |
| Full | 703.93 | 205 | <.001 | .757 | .102 | .125 |

**Table S3**

*Fit Indices for Three ASI Models: Black Women (Study 1)*

| Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| One-factor | 1316.30 | 209 | <.001 | .621 | .117 | .124 |
| Two-factor | 806.64 | 208 | <.001 | .795 | .086 | .101 |
| Full | 739.29 | 205 | <.001 | .817 | .082 | .101 |

**Table S4**

*Fit Indices for Three ASI Models: White Men (Study 1)*

| Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| One-factor | 3085.39 | 209 | <.001 | .663 | .137 | .128 |
| Two-factor | 1733.16 | 208 | <.001 | .821 | .100 | .088 |
| Full | 1181.71 | 205 | <.001 | .886 | .081 | .078 |

**Table S5**

*Fit Indices for Three ASI Models: White Women (Study 1)*

| Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| One-Factor | 2683.29 | 209 | <.001 | .727 | .121 | .102 |
| Two-Factor | 1409.76 | 208 | <.001 | .867 | .084 | .064 |
| Full | 1140.75 | 205 | <.001 | .897 | .075 | .059 |

## Part 2: Study 2 Measurement Invariance (MI) Analysis at Time 2

As noted in the main text (Chapter 3 *Results* section, *Measurement Invariance/ Formal MI Analysis* subsection), the conclusions based on MI analyses at Time 1 and Time 2 are substantively the same; therefore, only the Time 1 results were reported in the main text. The Time 2 analyses are reported here, with Tables S5-S8 corresponding to Tables 24-27 in the main text and Tables S9-S22 corresponding to Tables 29-42.

**Table S6**

*Cronbach's Alpha (α) Coefficients for Each Scale and Subscale of the ASI by Group (Time 2)*

| Group | AS | HS | BS | INT | PAT | DIFF |
|---|---|---|---|---|---|---|
| Black Men | .86 | .90 | .83 | .74 | **.66** | .83 |
| Black Women | .90 | .89 | .86 | .76 | **.65** | .83 |
| White Men | .92 | .95 | .90 | .84 | .83 | .88 |
| White Women | .93 | .91 | .90 | .80 | .80 | .88 |
| Full Sample | .92 | .93 | .89 | .81 | .77 | .86 |
| GF96 Avg | .87 | .88 | .79 | n/a | n/a | n/a |

*Note.* AS = ambivalent sexism scale; HS = hostile sexism scale; BS = benevolent sexism scale; INT = heterosexual intimacy subscale; PAT = protective paternalism subscale; DIFF = complementary gender differentiation subscale; GF96 Avg = the average reliability coefficients reported by Glick and Fiske (1996). The bolded coefficients are highlighted as particularly low (i.e., not meeting typical levels of acceptance).

**Table S7**

*Goodness-of-Fit Statistics for the Full Model by Group (Time 2)*

| Group | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| Black Men | 679.82 | 205 | <.001 | **.891** | .075 | .077 |
| Black Women | 599.04 | 205 | <.001 | .901 | .069 | .060 |
| White Men | 670.49 | 205 | <.001 | .941 | .070 | .065 |
| White Women | 696.45 | 205 | <.001 | .917 | .073 | .055 |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S8**

*Full Model (22 items): Indices of Invariance Model Fit at Time 2*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 2645.80 | 820 | <.001 | .917 | .072 | .064 |
| M2: First-Order Metric | 2864.77 | 874 | <.001 | .910 | .073 | .078 |
| M3: Second-Order Metric | 2896.43 | 880 | <.001 | .909 | .073 | **.081** |
| M4: First-Order Scalar | 3215.65 | 934 | <.001 | **.897** | .075 | **.084** |
| M5: Second-Order Scalar | 3467.03 | 943 | <.001 | **.886** | .079 | **.113** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S9**

*Full ASI Model (22 items): Invariance Model Comparisons and Changes in Fit at Time 2*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 218.97 | 54 | <.001 | -.007 | .001 | .014 |
| M3 vs M2 | 31.65 | 6 | <.001 | -.001 | .000 | .003 |
| M4 vs M3 | 319.23 | 54 | <.001 | **-.012** | .002 | .003 |
| M5 vs M4 | 251.38 | 9 | <.001 | **-.011** | .004 | **.029** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table S10**

*Indices of Invariance Model Fit: Black Men vs. Black Women (Time 2)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1278.86 | 410 | <.001 | **.896** | .072 | .069 |
| M2: First-Order Metric | 1315.55 | 428 | <.001 | **.894** | .071 | .074 |
| M3: Second-Order Metric | 1318.18 | 430 | <.001 | **.893** | .071 | .074 |
| M4: First-Order Scalar | 1400.63 | 448 | <.001 | **.886** | .072 | .076 |
| M5: Second-Order Scalar | 1564.81 | 451 | <.001 | **.866** | .078 | **.111** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S11**

*Model Comparisons and Changes in Fit: Black Men vs. Black Women (Time 2)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 36.68 | 18 | .006 | -.002 | -.001 | .005 |
| M3 vs M2 | 2.63 | 2 | .268 | -.001 | .000 | .000 |
| M4 vs M3 | 82.45 | 18 | <.001 | *-.007* | .001 | .002 |
| M5 vs M4 | 164.18 | 3 | <.001 | **-.020** | .006 | **.035** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table S12**

*Indices of Invariance Model Fit: Black Men vs. White Men (Time 2)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1350.31 | 410 | <.001 | .923 | .072 | .071 |
| M2: First-Order Metric | 1453.36 | 428 | <.001 | .916 | .074 | **.086** |
| M3: Second-Order Metric | 1457.94 | 430 | <.001 | .916 | .074 | **.087** |
| M4: First-Order Scalar | 1546.41 | 448 | <.001 | .910 | .075 | **.088** |
| M5: Second-Order Scalar | 1594.45 | 451 | <.001 | .906 | .076 | **.097** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S13**

*Model Comparisons and Changes in Fit: Black Men vs. White Men (Time 2)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 103.04 | 18 | <.001 | *-.007* | .002 | .015 |
| M3 vs M2 | 4.58 | 2 | <.001 | .000 | .000 | .001 |
| M4 vs M3 | 88.47 | 18 | <.001 | *-.006* | .001 | .001 |
| M5 vs M4 | 48.04 | 3 | <.001 | -.004 | .001 | *.009* |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table S14**

*Indices of Invariance Model Fit: Black Men vs. White Women (Time 2)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1376.27 | 410 | <.001 | .906 | .074 | .065 |
| M2: First-Order Metric | 1484.84 | 428 | <.001 | **.897** | .076 | **.081** |
| M3: Second-Order Metric | 1488.43 | 430 | <.001 | **.897** | .075 | **.081** |
| M4: First-Order Scalar | 1590.97 | 448 | <.001 | **.889** | .077 | **.082** |
| M5: Second-Order Scalar | 1788.69 | 451 | <.001 | **.870** | **.083** | **.134** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S15**

*Model Comparisons and Changes in Fit: Black Men vs. White Women (Time 2)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | **Δ**CFI | **Δ**RMSEA | **Δ**SRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 108.58 | 18 | <.001 | *-.009* | .002 | .016 |
| M3 vs M2 | 3.58 | 2 | .167 | .000 | -.001 | .000 |
| M4 vs M3 | 102.54 | 18 | <.001 | *-.008* | .002 | .001 |
| M5 vs M4 | 197.72 | 3 | <.001 | **-.019** | .006 | **.052** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table S16**

*Indices of Invariance Model Fit: Black Women vs. White Men vs. White Women (Time 1)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1965.98 | 615 | <.001 | .924 | .071 | .060 |
| M2: First-Order Metric | 2082.31 | 651 | <.001 | .919 | .071 | .069 |
| M3: Second-Order Metric | 2112.65 | 655 | <.001 | .918 | .071 | .073 |
| M4: First-Order Scalar | 2355.89 | 691 | <.001 | .906 | .074 | .075 |
| M5: Second-Order Scalar | 2424.23 | 697 | <.001 | .903 | .075 | **.084** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S17**

*Model Comparisons and Changes in Fit: Black Women vs. White Men vs. White Women (Time 2)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 116.32 | 36 | <.001 | -.005 | .000 | .009 |
| M3 vs M2 | 30.34 | 4 | <.001 | -.001 | .000 | .004 |
| M4 vs M3 | 243.24 | 36 | <.001 | **-.012** | .003 | .002 |
| M5 vs M4 | 68.34 | 6 | <.001 | -.003 | .001 | *.009* |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table S18**

*Indices of Invariance Model Fit: Black Women vs. White Women (Time 2)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1295.49 | 410 | <.001 | .910 | .071 | .057 |
| M2: First-Order Metric | 1366.80 | 428 | <.001 | .905 | .072 | .067 |
| M3: Second-Order Metric | 1386.04 | 430 | <.001 | .903 | .072 | .071 |
| M4: First-Order Scalar | 1455.09 | 448 | <.001 | **.898** | .073 | .072 |
| M5: Second-Order Scalar | 1465.23 | 451 | <.001 | **.897** | .073 | .074 |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S19**

*Model Comparisons and Changes in Fit: Black Women vs. White Women (Time 2)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 71.31 | 18 | <.001 | -.005 | .001 | .010 |
| M3 vs M2 | 19.24 | 2 | <.001 | -.002 | .000 | .004 |
| M4 vs M3 | 69.05 | 18 | <.001 | -.005 | .001 | .001 |
| M5 vs M4 | 10.14 | 3 | <.001 | -.001 | .000 | .002 |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table S20**

*Indices of Invariance Model Fit: White Men vs. White Women (Time 2)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1366.94 | 410 | <.001 | .930 | .071 | .060 |
| M2: First-Order Metric | 1415.97 | 428 | <.001 | .928 | .071 | .063 |
| M3: Second-Order Metric | 1419.12 | 430 | <.001 | .928 | .071 | .065 |
| M4: First-Order Scalar | 1545.04 | 448 | <.001 | .920 | .073 | .067 |
| M5: Second-Order Scalar | 1598.93 | 451 | <.001 | .917 | .074 | .078 |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S21**

*Model Comparisons and Changes in Fit: White Men vs. White Women (Time 2)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | ΔCFI | ΔRMSEA | ΔSRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 49.02 | 18 | <.001 | -.002 | .000 | .003 |
| M3 vs M2 | 3.16 | 2 | .206 | .000 | .000 | .002 |
| M4 vs M3 | 125.92 | 18 | <.001 | *-.008* | .002 | .002 |
| M5 vs M4 | 53.89 | 3 | <.001 | -.003 | .001 | **.011** |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

**Table S22**

*Indices of Invariance Model Fit: Black Women vs. White Men (Time 2)*

| Invariance Model | $\chi 2$ | *df* | *p*-value | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| M1: Configural | 1269.54 | 410 | <.001 | .927 | .070 | .063 |
| M2: First-Order Metric | 1327.65 | 428 | <.001 | .924 | .070 | .071 |
| M3: Second-Order Metric | 1350.57 | 430 | <.001 | .922 | .070 | .075 |
| M4: First-Order Scalar | 1518.13 | 448 | <.001 | .910 | .074 | .079 |
| M5: Second-Order Scalar | 1566.54 | 451 | <.001 | .906 | .075 | **.087** |

*Note.* The bolded fit statistics fail to reach the typical, less conservative cutoffs for acceptable fit.

**Table S23**

*Model Comparisons and Changes in Fit: Black Women vs. White Men (Time 2)*

| Comparison | $\Delta\chi 2$ | $\Delta df$ | *p*-value | **Δ**CFI | **Δ**RMSEA | **Δ**SRMR |
|---|---|---|---|---|---|---|
| M2 vs M1 | 58.11 | 18 | <.001 | -.003 | .000 | .008 |
| M3 vs M2 | 22.92 | 2 | <.001 | -.002 | .000 | .004 |
| M4 vs M3 | 167.56 | 18 | <.001 | **-.012** | .004 | .004 |
| M5 vs M4 | 48.41 | 3 | <.001 | -.004 | .001 | *.008* |

*Note.* The italicized numbers represent changes in fit statistics that are unacceptable according to the more conservative cutoffs. The bolded numbers represent changes in fit statistics that are unacceptable according to the less conservative cutoffs. See Table 28 for cutoff information.

## Part 3: Study 2 Analysis of Feeling Thermometer (FT) Effects on HS and BS Endorsement

As noted in the main text (Chapter 3 *Results* section, *Exploratory Analysis/Racial Attitudes (Feeling Thermometers)* subsection), the effects of racial FTs on HS and BS endorsement in this experiment were minimal, possibly due to the low levels of prejudicial racial attitudes among the Black women, White women, and White men in the sample. (Black men were excluded from analyses based on the MI analyses.)

The FT effects were examined in multiple ways via OLS regression analyses. Reported here are the results in their most simple, straightforward form. For each eligible group (i.e., Black women, White women, and White men), the effects of the four racial FTs on HS and BS were examined within each experimental condition. Therefore, for each eligible participant group, ten models are reported in the following tables. Although there are some significant effects, they are quite small and unlikely to have any practical meaning (e.g., distinguish between a hostile sexist and a non-sexist).

**Table S24**

*Regression Coefficients of Four Racial FTs on HS: Black Women in the White Prime Condition (*N = *79)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 4.59 | 0.58 | 7.93 | <.001 | [3.44, 5.74] |
| White FT | 0.01 | 0.00 | 2.89 | .005 | [0.00, 0.02] |
| Black FT | -0.02 | 0.01 | -2.47 | .016 | [-0.03, -0.00] |
| Asian FT | -0.02 | 0.01 | -2.25 | .028 | [-0.04, -0.00] |
| Latinx FT | 0.00 | 0.01 | -0.02 | .982 | [-0.02, 0.02] |
| $R^2$ = .256 | | | | | |

**Table S25**

*Regression Coefficients of Four Racial FTs on HS: Black Women in the Black Prime Condition (*N = *78)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.29 | 0.70 | 4.72 | <.001 | [1.90, 4.67] |
| White FT | 0.01 | 0.01 | 2.83 | .006 | [0.00, 0.03] |
| Black FT | -0.01 | 0.01 | -0.60 | .548 | [-0.02, 0.01] |
| Asian FT | -0.01 | 0.01 | -0.96 | .342 | [-0.03, 0.01] |
| Latinx FT | -0.01 | 0.01 | -1.02 | .309 | [-0.03, 0.01] |
| $R^2$ = .136 | | | | | |

**Table S26**

*Regression Coefficients of Four Racial FTs on HS: Black Women in the Asian Prime Condition (*N = *80)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.44 | 0.59 | 5.82 | <.001 | [2.26, 4.61] |
| White FT | 0.01 | 0.01 | 1.47 | .146 | [-0.00, 0.02] |
| Black FT | 0.00 | 0.01 | -0.45 | .651 | [-0.02, 0.01] |
| Asian FT | -0.01 | 0.01 | -1.40 | .166 | [-0.03, 0.00] |
| Latinx FT | -0.01 | 0.01 | -0.67 | .502 | [-0.02, 0.01] |
| $R^2$ = .074 | | | | | |

**Table S27**

*Regression Coefficients of Four Racial FTs on HS: Black Women in the Latinx Prime Condition (*N = *83)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.44 | 0.56 | 6.19 | <.001 | [2.33, 4.54] |
| White FT | 0.02 | 0.01 | 2.90 | .005 | [0.01, 0.03] |
| Black FT | 0.00 | 0.01 | -0.69 | .492 | [-0.02, 0.01] |
| Asian FT | 0.00 | 0.01 | -0.30 | .764 | [-0.02, 0.02] |
| Latinx FT | -0.02 | 0.01 | -2.32 | .023 | [-0.04, -0.00] |
| $R^2$ = .181 | | | | | |

**Table S28**

*Regression Coefficients of Four Racial FTs on HS: Black Women in the Control Condition (*N = *80)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 4.97 | 0.66 | 7.50 | <.001 | [3.65, 6.29] |
| White FT | 0.01 | 0.00 | 2.12 | .037 | [0.00, 0.02] |
| Black FT | -0.02 | 0.01 | -2.46 | .016 | [-0.04, -0.00] |
| Asian FT | -0.02 | 0.01 | -2.41 | .018 | [-0.04, -0.00] |
| Latinx FT | 0.00 | 0.01 | 0.19 | .848 | [-0.02, 0.02] |
| $R^2$ = .280 | | | | | |

**Table S29**

*Regression Coefficients of Four Racial FTs on BS: Black Women in the White Prime Condition (*N = *79)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 4.74 | 0.62 | 7.65 | <.001 | [3.51, 5.98] |
| White FT | 0.01 | 0.01 | 0.99 | .323 | [-0.01, 0.02] |
| Black FT | 0.00 | 0.01 | 0.63 | .533 | [-0.01, 0.02] |
| Asian FT | -0.01 | 0.01 | -1.57 | .120 | [-0.03, 0.00] |
| Latinx FT | -0.01 | 0.01 | -1.34 | .184 | [-0.03, 0.01] |
| $R^2$ = .157 | | | | | |

**Table S30**

*Regression Coefficients of Four Racial FTs on BS: Black Women in the Black Prime Condition (*N = *78)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.68 | 0.73 | 5.03 | <.001 | [2.22, 5.13] |
| White FT | 0.01 | 0.01 | 1.28 | .206 | [-0.00, 0.02] |
| Black FT | 0.01 | 0.01 | 0.64 | .525 | [-0.01, 0.02] |
| Asian FT | 0.00 | 0.01 | -0.34 | .737 | [-0.02, 0.02] |
| Latinx FT | -0.01 | 0.01 | -1.05 | .298 | [-0.03, 0.01] |
| $R^2$ = .044 | | | | | |

**Table S31**

*Regression Coefficients of Four Racial FTs on BS: Black Women in the Asian Prime Condition (*N = *80)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.97 | 0.54 | 5.55 | <.001 | [1.91, 4.04] |
| White FT | 0.01 | 0.01 | 1.72 | .090 | [-0.00, 0.02] |
| Black FT | 0.02 | 0.01 | 2.77 | .007 | [0.01, 0.03] |
| Asian FT | -0.01 | 0.01 | -1.81 | .075 | [-0.03, 0.00] |
| Latinx FT | -0.01 | 0.01 | -0.80 | .424 | [-0.02, 0.01] |
| $R^2$ = .140 | | | | | |

**Table S32**

*Regression Coefficients of Four Racial FTs on BS: Black Women in the Latinx Prime Condition (*N = *83)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.15 | 0.50 | 6.28 | <.001 | [2.15, 4.15] |
| White FT | 0.02 | 0.01 | 3.94 | <.001 | [0.01, 0.03] |
| Black FT | 0.01 | 0.01 | 1.42 | .160 | [-0.00, 0.02] |
| Asian FT | -0.01 | 0.01 | -1.47 | .145 | [-0.03, 0.00] |
| Latinx FT | -0.01 | 0.01 | -0.98 | .329 | [-0.02, 0.01] |
| $R^2$ = .176 | | | | | |

**Table S33**

*Regression Coefficients of Four Racial FTs on BS: Black Women in the Control Condition (*N = 80*)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 5.03 | 0.67 | 7.49 | <.001 | [3.70, 6.37] |
| White FT | 0.00 | 0.00 | 0.21 | .836 | [-0.01, 0.01] |
| Black FT | 0.00 | 0.01 | -0.59 | .560 | [-0.02, 0.01] |
| Asian FT | -0.01 | 0.01 | -0.73 | .470 | [-0.02, 0.01] |
| Latinx FT | -0.01 | 0.01 | -0.64 | .524 | [-0.02, 0.01] |
| $R^2$ = .096 | | | | | |

**Table S34**

*Regression Coefficients of Four Racial FTs on HS: White Women in the White Prime Condition (*N = *91)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.13 | 0.53 | 4.03 | <.001 | [1.08, 3.18] |
| White FT | 0.02 | 0.01 | 2.95 | .004 | [0.01, 0.03] |
| Black FT | 0.00 | 0.01 | -0.18 | .860 | [-0.02, 0.01] |
| Asian FT | 0.01 | 0.01 | 0.72 | .475 | [-0.01, 0.03] |
| Latinx FT | -0.02 | 0.01 | -1.98 | .051 | [-0.04, 0.00] |
| $R^2$ = .141 | | | | | |

**Table S35**

*Regression Coefficients of Four Racial FTs on HS: White Women in the Black Prime Condition (*N = *91)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.24 | 0.43 | 5.26 | <.001 | [1.39, 3.09] |
| White FT | 0.02 | 0.01 | 4.08 | <.001 | [0.01, 0.03] |
| Black FT | -0.01 | 0.01 | -0.80 | .424 | [-0.03, 0.01] |
| Asian FT | 0.00 | 0.01 | 0.40 | .688 | [-0.02, 0.02] |
| Latinx FT | -0.02 | 0.01 | -1.34 | .182 | [-0.04, 0.01] |
| $R^2$ = .237 | | | | | |

**Table S36**

*Regression Coefficients of Four Racial FTs on HS: White Women in the Asian Prime Condition (*N = *91)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 1.73 | 0.45 | 3.86 | <.001 | [0.84, 2.62] |
| White FT | 0.02 | 0.01 | 3.56 | .001 | [0.01, 0.03] |
| Black FT | 0.00 | 0.01 | -0.42 | .674 | [-0.03, 0.02] |
| Asian FT | 0.00 | 0.01 | -0.10 | .924 | [-0.02, 0.02] |
| Latinx FT | -0.01 | 0.01 | -0.65 | .520 | [-0.03, 0.01] |
| $R^2$ = .130 | | | | | |

**Table S37**

*Regression Coefficients of Four Racial FTs on HS: White Women in the Latinx Prime Condition (*N = *90)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.04 | 0.44 | 4.59 | <.001 | [1.16, 2.93] |
| White FT | 0.02 | 0.01 | 2.58 | .012 | [0.00, 0.03] |
| Black FT | -0.02 | 0.01 | -1.79 | .078 | [-0.05, 0.00] |
| Asian FT | -0.01 | 0.01 | -1.01 | .316 | [-0.03, 0.01] |
| Latinx FT | 0.02 | 0.01 | 1.13 | .260 | [-0.01, 0.04] |
| $R^2$ = .104 | | | | | |

**Table S38**

*Regression Coefficients of Four Racial FTs on HS: White Women in the Control Condition (*N = *89)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.29 | 0.41 | 8.11 | <.001 | [2.48, 4.10] |
| White FT | 0.02 | 0.01 | 3.91 | <.001 | [0.01, 0.03] |
| Black FT | -0.02 | 0.01 | -1.76 | .082 | [-0.03, 0.00] |
| Asian FT | -0.01 | 0.01 | -1.00 | .319 | [-0.02, 0.01] |
| Latinx FT | -0.01 | 0.01 | -1.41 | .162 | [-0.03, 0.00] |
| $R^2$ = .288 | | | | | |

**Table S39**

*Regression Coefficients of Four Racial FTs on BS: White Women in the White Prime Condition (*N = *91)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.63 | 0.54 | 4.83 | <.001 | [1.55, 3.71] |
| White FT | 0.01 | 0.01 | 2.35 | .021 | [0.00, 0.03] |
| Black FT | -0.01 | 0.01 | -0.88 | .383 | [-0.02, 0.01] |
| Asian FT | 0.02 | 0.01 | 1.97 | .052 | [-0.00, 0.04] |
| Latinx FT | -0.03 | 0.01 | -2.32 | .022 | [-0.05, -0.00] |
| $R^2$ = .151 | | | | | |

**Table S40**

*Regression Coefficients of Four Racial FTs on BS: White Women in the Black Prime Condition (*N = *91)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.37 | 0.48 | 6.97 | <.001 | [2.41, 4.34] |
| White FT | 0.01 | 0.01 | 1.26 | .210 | [-0.00, 0.02] |
| Black FT | -0.01 | 0.01 | -0.72 | .476 | [-0.03, 0.01] |
| Asian FT | 0.00 | 0.01 | -0.06 | .950 | [-0.02, 0.02] |
| Latinx FT | -0.01 | 0.01 | -0.46 | .649 | [-0.04, 0.02] |
| $R^2$ = .075 | | | | | |

**Table S41**

*Regression Coefficients of Four Racial FTs on BS: White Women in the Asian Prime Condition (*N = *91)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.72 | 0.43 | 8.64 | <.001 | [2.87, 4.58] |
| White FT | 0.01 | 0.01 | 2.73 | .008 | [0.00, 0.03] |
| Black FT | 0.01 | 0.01 | 1.08 | .281 | [-0.01, 0.03] |
| Asian FT | -0.01 | 0.01 | -0.89 | .376 | [-0.03, 0.01] |
| Latinx FT | -0.03 | 0.01 | -2.90 | .005 | [-0.05, -0.01] |
| $R^2$ = .229 | | | | | |

**Table S42**

*Regression Coefficients of Four Racial FTs on BS: White Women in the Latinx Prime Condition (*N = *90)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.16 | 0.48 | 6.61 | <.001 | [2.21, 4.11] |
| White FT | 0.02 | 0.01 | 2.05 | .044 | [0.00, 0.03] |
| Black FT | -0.03 | 0.01 | -2.38 | .019 | [-0.06, -0.01] |
| Asian FT | 0.01 | 0.01 | 1.17 | .244 | [-0.01, 0.03] |
| Latinx FT | 0.00 | 0.01 | 0.07 | .942 | [-0.03, 0.03] |
| $R^2$ = .134 | | | | | |

**Table S43**

*Regression Coefficients of Four Racial FTs on BS: White Women in the Control Condition (*N = *89)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.72 | 0.41 | 6.71 | <.001 | [1.91, 3.53] |
| White FT | 0.01 | 0.01 | 1.92 | .059 | [-0.00, 0.02] |
| Black FT | 0.00 | 0.01 | -0.37 | .711 | [-0.02, 0.01] |
| Asian FT | -0.01 | 0.01 | -1.73 | .087 | [-0.03, 0.00] |
| Latinx FT | 0.01 | 0.01 | 1.11 | .271 | [-0.01, 0.02] |
| $R^2$ = .079 | | | | | |

**Table S44**

*Regression Coefficients of Four Racial FTs on HS: White Men in the White Prime Condition (*N *=94*)

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.62 | 0.56 | 6.51 | <.001 | [2.51, 4.72] |
| White FT | 0.02 | 0.01 | 3.29 | .001 | [0.01, 0.04] |
| Black FT | -0.01 | 0.01 | -1.12 | .265 | [-0.03, 0.01] |
| Asian FT | -0.01 | 0.01 | -0.94 | .350 | [-0.03, 0.01] |
| Latinx FT | -0.02 | 0.01 | -1.74 | .086 | [-0.03, 0.00] |
| $R^2$ = .240 | | | | | |

**Table S45**

*Regression Coefficients of Four Racial FTs on HS: White Men in the Black Prime Condition (*N *=93*)

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.92 | 0.51 | 5.71 | <.001 | [1.91, 3.94] |
| White FT | 0.02 | 0.01 | 2.98 | .004 | [0.01, 0.03] |
| Black FT | -0.03 | 0.01 | -3.74 | <.001 | [-0.05, -0.01] |
| Asian FT | 0.01 | 0.01 | 0.81 | .421 | [-0.01, 0.03] |
| Latinx FT | 0.00 | 0.01 | -0.29 | .773 | [-0.02, 0.01] |
| $R^2$ = .228 | | | | | |

**Table S46**

*Regression Coefficients of Four Racial FTs on HS: White Men in the Asian Prime Condition (*N *=94*)

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.89 | 0.49 | 7.98 | <.001 | [2.92, 4.86] |
| White FT | 0.02 | 0.01 | 2.72 | .008 | [0.01, 0.04] |
| Black FT | -0.03 | 0.01 | -4.63 | <.001 | [-0.05, -0.02] |
| Asian FT | 0.00 | 0.01 | -0.29 | .771 | [-0.03, 0.02] |
| Latinx FT | 0.00 | 0.01 | -0.06 | .953 | [-0.03, 0.03] |
| $R^2$ = .381 | | | | | |

**Table S47**

*Regression Coefficients of Four Racial FTs on HS: White Men in the Latinx Prime Condition (*N *=94*)

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.91 | 0.43 | 6.77 | <.001 | [2.06, 3.77] |
| White FT | 0.02 | 0.01 | 4.11 | <.001 | [0.01, 0.04] |
| Black FT | 0.01 | 0.01 | 1.56 | .123 | [-0.00, 0.03] |
| Asian FT | 0.00 | 0.01 | -0.31 | .757 | [-0.02, 0.02] |
| Latinx FT | -0.04 | 0.01 | -3.39 | .001 | [-0.06, -0.02] |
| $R^2$ = .280 | | | | | |

**Table S48**

*Regression Coefficients of Four Racial FTs on HS: White Men in the Control Condition (*N = *92)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.79 | 0.50 | 5.56 | <.001 | [1.80, 3.79] |
| White FT | 0.03 | 0.01 | 4.15 | <.001 | [0.01, 0.04] |
| Black FT | -0.02 | 0.01 | -2.58 | .011 | [-0.04, -0.01] |
| Asian FT | 0.01 | 0.01 | 0.81 | .422 | [-0.02, 0.04] |
| Latinx FT | -0.01 | 0.01 | -0.97 | .335 | [-0.04, 0.01] |
| $R^2$ = .349 | | | | | |

**Table S49**

*Regression Coefficients of Four Racial FTs on BS: White Men in the White Prime Condition (*N = *94)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.15 | 0.51 | 6.13 | <.001 | [2.13, 4.17] |
| White FT | 0.01 | 0.01 | 1.36 | .178 | [-0.00, 0.02] |
| Black FT | 0.00 | 0.01 | -0.10 | .921 | [-0.02, 0.02] |
| Asian FT | -0.01 | 0.01 | -0.95 | .344 | [-0.03, 0.01] |
| Latinx FT | 0.00 | 0.01 | -0.02 | .987 | [-0.02, 0.02] |
| $R^2$ = .030 | | | | | |

**Table S50**

*Regression Coefficients of Four Racial FTs on BS: White Men in the Black Prime Condition (*N = *93)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.43 | 0.52 | 4.64 | <.001 | [1.39, 3.47] |
| White FT | 0.01 | 0.01 | 1.40 | .166 | [-0.00, 0.02] |
| Black FT | -0.01 | 0.01 | -1.52 | .133 | [-0.03, 0.00] |
| Asian FT | 0.01 | 0.01 | 1.10 | .276 | [-0.01, 0.03] |
| Latinx FT | 0.00 | 0.01 | 0.15 | .879 | [-0.02, 0.02] |
| $R^2$ = .060 | | | | | |

**Table S51**

*Regression Coefficients of Four Racial FTs on BS: White Men in the Asian Prime Condition (*N = *94)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 3.41 | 0.50 | 6.78 | <.001 | [2.41, 4.40] |
| White FT | 0.01 | 0.01 | 1.03 | .304 | [-0.01, 0.02] |
| Black FT | -0.01 | 0.01 | -1.16 | .249 | [-0.02, 0.01] |
| Asian FT | -0.01 | 0.01 | -0.47 | .637 | [-0.03, 0.02] |
| Latinx FT | 0.00 | 0.01 | 0.30 | .768 | [-0.02, 0.03] |
| $R^2$ = .044 | | | | | |

**Table S52**

*Regression Coefficients of Four Racial FTs on BS: White Men in the Latinx Prime Condition (*N = *94)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 1.71 | 0.42 | 4.04 | <.001 | [0.87, 2.55] |
| White FT | 0.01 | 0.01 | 1.19 | .236 | [-0.00, 0.02] |
| Black FT | 0.01 | 0.01 | 0.71 | .481 | [-0.01, 0.03] |
| Asian FT | 0.02 | 0.01 | 2.09 | .040 | [0.00, 0.04] |
| Latinx FT | -0.01 | 0.01 | -0.93 | .355 | [-0.03, 0.01] |
| $R^2$ = .155 | | | | | |

**Table S53**

*Regression Coefficients of Four Racial FTs on BS: White Men in the Control Condition (*N = *92)*

| Variable | *B* | *SE* | *t* | *p* | 95% CI |
|---|---|---|---|---|---|
| Constant | 2.34 | 0.44 | 5.27 | <.001 | [1.46, 3.22] |
| White FT | 0.01 | 0.01 | 2.63 | .010 | [0.00, 0.03] |
| Black FT | 0.00 | 0.01 | 0.08 | .934 | [-0.02, 0.02] |
| Asian FT | 0.01 | 0.01 | 0.96 | .340 | [-0.01, 0.04] |
| Latinx FT | -0.01 | 0.01 | -1.21 | .230 | [-0.04, 0.01] |
| $R^2$ = .111 | | | | | |