Author: Olivia Pavco-Giaccia
Thesis Advisors: Jason Stanley, Yarrow Dunham
April 22, 2016

Rationality is Gendered:
Using Social Cognition to Explore the Thinking/Feeling Gender Bias

Abstract

This paper used the tools of the social sciences to investigate the feminist philosophical premise that 'reason' and 'femaleness' are culturally at odds with one another, a hypothesis also known as the thinking/feeling gender bias. Results confirmed our expectations, indicating that there are robust semantic associations between male and rationality, as well as between female and emotionality. More specifically, the results of Study 1a suggested that both men and women implicitly associated male with rationality and female with emotionality (more than vice versa). Study 1b broke down these relative tendencies into their constituent parts, finding that there is a tendency to associate male with rationality more than female, and a tendency to associate female with emotionality more than male. Study 2 provided convergent evidence by means of an AMP test to confirm the semantic links between thinking/feeling and male/female concepts. Self-reported associations between gender and reason/emotion in Study 2 reflected an explicit acknowledgement of the stereotype in men, but much less so in women. Thus, in addition to confirming the presence of the thinking/feeling gender bias, the results of Study 2 indicated that women exhibit an intriguing implicit-explicit divergence within the domain of gendered rationality.

Introduction

From the first moments of life, children are bombarded with a rich array of cues that innocuously convey stereotypical gender roles (Bridges, 1992). From the color of congratulations cards and nursery walls, to the toys, names, and clothing they are exposed to, gender is presented to children as an important cultural distinction that must be mastered. Furthermore, children not only perceive themselves, but are also perceived, through a prism of gender. In fact, male and female babies are interpreted quite differently by adult observers based on imposed gender cues such as clothing or names; infants thought to be male are interpreted as more active and agentic, and infants thought to be female as more delicate and sweet (Smith & Lloyd, 1978). Little surprise, then, that gender becomes a defining feature of life, affecting friendship patterns (Howes & Philipsen, 1992), clothing (Cox & Dittmar, 1995), and play preferences (Francis, 2010).

But beyond these relatively benign and presumably voluntary gender-typed preferences, gender also becomes a source of radical inequality. Most strikingly, in the United States, gender correlates with a significant pay gap (Blau & Kahn 2006), unequal professional advancement opportunities (Richmond, 2014), unequal division of household labor (Bird 1999), and occupational segregation (Charles & Grusky 2004). Many factors underlie these disparities, and extensive academic literature has been expended in an attempt to make sense of them. For example, in his famous debate with Elizabeth Spelke, Stephen Pinker invoked several potential explanations for the gender gap. He included differences in male/female priorities, female interest in people versus abstract rule systems, male comfort with risk, male abilities to affect three dimensional mental transformations, differences in mathematical reasoning, and sex difference in variability. He then went on to imply that many of these factors have some innate, biological basis, thus we should not seek to change them (Pinker & Spelke, 2005). Spelke responded by claiming that people have equal ability to develop an intuitive

understanding of the physical world; any gender divergence in something like formal mathematical abilities could not possibly be due to evolutionary differences, because it is a relatively recent accomplishment. However, Spelke does note that many sex differences emerge at older ages, making it difficult to definitively tease apart their biological and social sources (Pinker & Spelke, 2005).

According to the research cited by both Pinker and Spelke, it is clear that some gender differences are motivated by biology. However, due to the overwhelming presence of modern gender biases, it seems likely that social factors, rather than intrinsic aptitude, account for the biggest effect. For example, despite a widespread push for gender equality, there are pervasive forms of subtle gender stereotypes that associate women with less prestigious, agentic, or lucrative roles. Both men and women associate women with the home over the workplace (Nadler, 2010), with language and the arts over math or science (White & White, 2006), and with more general attributes such as weakness over strength (Rudman & Greenwald, 2001). These findings provide insight into the content of subtle gender stereotypes, but do not yet provide a general framework for understanding the conceptual associations that support the maintenance and acceptance of unequal treatment. As it stands, a common explanation of the wage gap is the type of jobs that women prioritize. For example, women are thought to seek jobs requiring less effort and to be less concerned with professional progress and intellectual challenge (Blau & Ferber, 1991). Inherent in this argument is the assumption that intellectual work deserves higher pay than other, more "practical" forms of work, such as care-giving. Thus the intellectual – practical divide, where women are associated with the care-giving professions, and men with the "intellectual" ones, could play a role in the structural gender injustice that we see in the US.

Indeed, just this sort of argument has been central to a longstanding feminist philosophical critique of gender roles. In 1892, the civil rights activist Anna Julia Cooper gave voice to what she saw as the common view: "…as the man is more noble in reason, so the woman is more quick in sympathy.

That as he is indefatigable in pursuit of abstract truth, so is she in…striving tenderly and lovingly that not one of the least of these 'little ones' should perish," (Cooper, 1982). The question then presents itself: how do we address this gendered thinking/feeling divide? In her book*, Gender Trouble*, Judith Butler proposes that we should do our best to eliminate gender distinctions entirely. She claims that "the dogged effort to 'denaturalize' gender in this text emerges, I think, from a strong desire…to counter the normative violence implied by ideal morphologies of sex," (Butler, 1990). However, her utopian vision for a subversion of the "regulatory practice of identity," is not shared by all (Butler, 1990). Though many acknowledge the powerful implications of her ideal, gender-neutral society, few think it is possible. For example, psychologist and famous feminist scholar, Carol Gilligan, believes that boys and girls differ in terms of conceptions of self and morality. This is evidenced by her claim that boys and girls have a "different interpersonal orientation," (Gilligan, 1982). In other words, Gilligan views gender distinctions as a fact of life – one that is unlikely to change in the coming decades. Thus, Butler's solution to gender-based discrimination is, for now, untenable.

Thankfully, Butler was not the first person to write on this issue; by the late 1920s, W.E.B. Du Bois had already offered his own solution to discrimination in his formidable treatise, *Criteria of Negro Art*. Unlike Butler, who advocates for the destruction of social norms, here Du Bois suggests that it is possible to employ certain social stereotypes against themselves. Du Bois acknowledges that the African-American artist must challenge a problematic ideal (read: white supremacy,) however he recognizes that, at the time he is writing, a direct challenge to racist social norms would be ineffective. His solution is to "wrap challenges to the ideal in the tempting vocabulary of the ideal itself," (Stanley, 2015). For example, Du bois states that "If a colored man wants to publish a book, he has got to get a white publisher and a white newspaper to say it is great," (Du Bois, 1926). In order to do this, a black author must tiptoe around "all sorts of customs that have come down as second-hand soul clothes of

white patrons," (Du Bois, 1926). For Du bois, public acceptance is the first step towards redefining black artists as simply American artists. In order to correct racial divisions, the black artist must first achieve acceptance by embracing them.

This theory does not only apply to race. A modern, feminist application of Du Bois' solution to discrimination is substantiated in Evelyn Fox Keller's book, *Reflections on Gender in Science*. Here, Keller presents evidence for gender-bias in STEM; akin to Du Bois' thoughts on embracing racial divisions in black art, she concludes that science can never be gender-free until we acknowledge the current gender divisions which pervade it (Keller, 1985). Keller's argument for the acknowledgement of gender divisions in science is powerful, but she stops short at suggesting that scientific theory itself is gendered male. This is where Keller and her contemporary, Genevieve Lloyd, diverge. In her book, *The Man of Reason*, Lloyd argues that reason itself is gendered male, and if women want to be perceived as participants in 'reason,' they must conform to the masculine norms of scientific thought. Essential to Lloyd's thesis is the idea that objectivity has been gendered male, whereas subjectivity has been gendered female. Lloyd suggests that this gendering creates a dilemma for women who seek advancement in intellectual domains: If women want to participate in 'reason,' they must give up/distance themselves from some vital aspect of their "femaleness," and vice versa (Lloyd, 1984).

This feminist critique offered by Lloyd and others elicited a large backlash, with some critics (e.g. Nussbaum, 1994) suggesting that the very notion of gendered rationality is incoherent because "science is able to insulate itself from social values that would bias it and render it subjective" (Padovani, Richardson, & Tsou, 2015). Here, we argue that this backlash misses the mark; in fact, the cognitive sciences offer a simple and straightforward interpretation of Lloyd's argument that can be directly tested. Specifically, Lloyd's argument can be interpreted as the claim that the concept *reason* or *rational* is semantically associated with the concept *male*. By framing the question in this way, it

becomes straightforward to test through the tools of implicit social cognition. This methodology follows in the tradition of decades of work that has demonstrated conceptual associations between other ideas, from *doctor* and *nurse* to *cat* and *furry*. Thus, our aim here is to help bring the discussion of gendered associations of reason into the scientific mainstream by providing direct evidence in favor of the long standing feminist contention that the concept of reason is gendered male. Ultimately, we suggest that these conceptual associations plausibly support structural gender inequalities.

Study 1a uses the Implicit Association Test (IAT: Greenwald, McGhee, & Schwarz, 1988), as well as explicit measures, to demonstrate that participants preferentially associate *male* with *thinking* and *female* with *feeling*. Study 1b uses a modified IAT procedure to decompose that relative finding into two effects of roughly equal magnitude; namely, an association between *male* and *thinking* more than *feeling*, and an association between *female* and *feeling* more than *thinking.* Study 2 is a conceptual replication using a priming procedure to show that the rapid presentation of photographs of men (women) make concepts related to thinking (feeling) more accessible, influencing ostensibly unrelated semantic judgments. Taken together, these results provide direct evidence that rationality is gendered at the level of basic conceptual associations residing in semantic memory, giving empirical teeth to a long-standing assertion of feminist philosophy.

Study 1

Participants

Study 1a involved 124 adults (male = 69, female = 54, unknown =1; M_{age} = 37 years (*SD* = 12 years), White = 78%, Black = 10%, other = 11%) recruited via Amazon's Mechanical Turk online labor market. Recruitment was restricted to IP addresses within the United States and to workers with greater than 95% approval on prior tasks.

Study 1b involved a between-participants design with 204 adults (male = 98, female = 103, unknown =1; M_{age} = 36 years (*S* = 12 years), White = 73%, Black = 9%, other = 18%) recruited via the

sample procedure but excluding any participant who completed Study 1a. For both studies sample size was determined by attempting to meet or exceed the sample sizes of several recent studies that revealed evidence of implicit gender stereotypes (Jost & Kay, 2005; Kiefer & Sekaquaptewa, 2007; Leach, Carraro, Garcia, & Kang, 2015; Rudman & Glick, 2001; Rudman, Greenwald, & Mcghee, 2000), for which the median sample size was 98 (mean = 88).

Measures

Participants in both studies completed an Implicit Association Test (A. G. Greenwald, McGhee, & Schwartz, 1998) designed to measure associations between gendered category terms and terms related to thinking and feeling. The IAT is a widely used measure of association based on the logic that related terms will be easier to categorize using the same key than will unrelated or opposing terms. In the standard IAT employed in Study 1a participants rapidly categorized words referring to males (e.g. *uncle*, *brother, boy*), females (e.g. *aunt, sister, girl*), thinking (e.g. *rational, logical, deliberate)* and feeling (e.g. *emotional, intuition, empathic*) using a left and right response key. The thinking and feeling words were chosen by generating synonyms of the focal categories *thinking* and *feeling,* and then matching on word valence, frequency, and word length. Frequency data were drawn from the Corpus of Contemporary American English (http://corpus.byu.edu/coca/) and valence data were from Warriner and colleagues extensive set of word ratings (Warriner, Kuperman, & Brysbaert, 2013). Further details concerning word selection and the complete list of words used in all studies are in SOM-R.

Following the standard IAT procedure, participants first completed short practice blocks in which only male and female or only thinking and feeling words appeared, designed to familiarize participants with the words they would see in the critical blocks. In the two critical blocks that contribute to IAT scores all four word categories appear, such that participants have to either respond to both *male*

and *thinking* words using one response key and *female* and *feeling* words with the other response key, or the converse pairings (*male* and *feeling* with one key and *female* and *thinking* with the other key). Block order and the left-right position of the target categories were counterbalanced between participants. The dependent measure was the IAT *D-asis* statistic (Greenwald, Nosek, & Banaji, 2003), an effect size representing the mean difference in response latencies across critical blocks divided by the pooled within-participant standard deviation, coded such that positive values indicate a relative association between *male* and *rationality.*

Study 1b employed the Brief IAT (B-IAT) (Sriram & Greenwald, 2009), which involves only a single target attribute (i.e., only thinking or feeling rather than both), allowing an index of whether the relative effects from Study 1a (which are based on both the *male-rationality* and *female-emotionality* associations) are driven by a tendency to associate *male* with *rationality*, *female* with *emotionality*, or both. A standard B-IAT using the same stimulus items was employed; participants were randomly assigned to either the thinking or feeling condition as a between-participants factor.

In addition to the IAT measure, in both studies we elicited basic demographic information (race, gender, age, and political orientation), as well as single-item explicit measures of semantic association between male and rationality, male and emotionality, female and rationality, and female and emotionality (i.e., each of the four questions generated from the stem "how much do you associate MALE/FEMALE with THINKING/FEELING?"). These ratings were elicited via a slider ranging from "not at all" to "very strongly", and slider values were converted to a 1000-point scale for purposes of analysis.

Procedure

After accepting the task on Mechanical Turk participants downloaded a small browser plug-in allowing for the standard implementation of all tasks on their personal computer running Inquisit

(Millisecond Software, 2014). They then completed the IAT measure, followed by explicit items and then demographics.

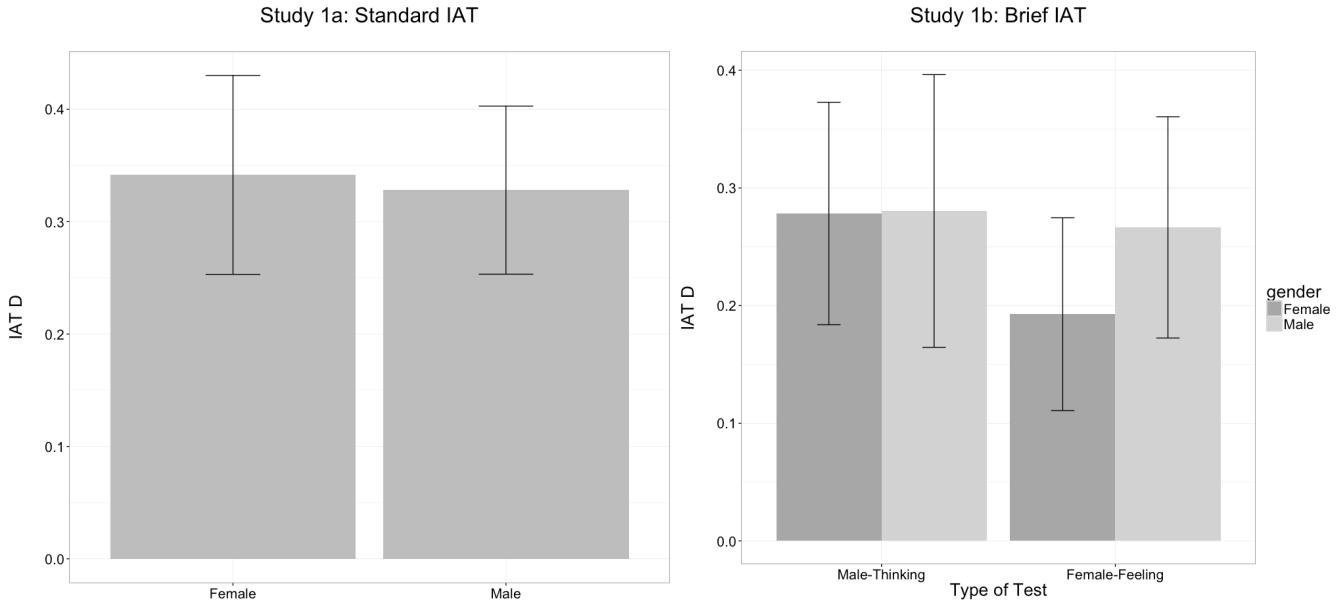
Results

Additional information on the exact IAT methodology (including words used and other relevant figures) is available in the supplementary materials section of this paper.

Study 1a: Relative associations between gender and rationality.

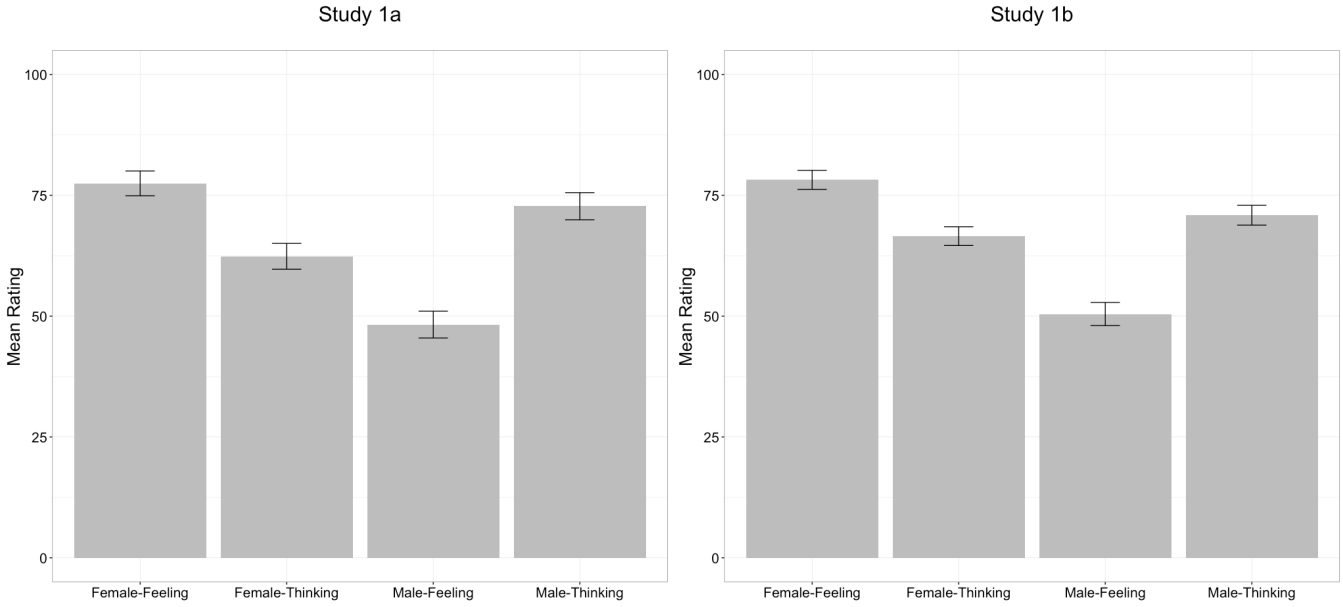
Standard exclusion criteria (Greenwald et al., 2003) led to the exclusion of 8 participants who had an excessive number of fast trials (> 10% at < 300 ms), generally indicative of rapid key pressing in order to move rapidly through the task. Thus, we base all analyses on the 116 participants with complete data. Results indicated a robust relative association between male and thinking and female and feeling relative to the opposite pairings, $D = .33$ (.31), $t(115) = 11.50$, $p < .0001$, 95% CIs [.28; .39] (Figure 1). There was no evidence of a difference between males ($M = .33$) and females ($M = .34$), Welch's $t(102.87) = .23$, $p = .82$, suggesting that both participant genders showed the same pattern of association.

Figure 1: Associations between Gender and Rationality and Emotionality, by participant gender. Error bars are 95% CIs.



Participants self-reported associations between gender and thinking versus feeling demonstrated evidence of a similar gendered pattern (Figure 2), with *male* associated with *thinking* ($M = 73.51$, $SD = 19.84$) more than *feeling* ($M = 47.25$, $SD = 19.11$) and *female* associated with *feeling* ($M = 78.31$, $SD = 19.08$) more than *thinking* ($M = 61.50$, $SD = 21.36$); all pairwise comparisons between these ratings were significant, $t(115) > 2.69$, $p < .009$. These ratings did not differ by participant gender, except for he explicit association between female and thinking, which was weaker in men ($M = 57.43$) than women ($M = 67.06$), Welch's $t(92.91) = 2.39$, $p = .019$.

Figure 2: Average self-reported ratings for each of four gender ratings. Error bars are 95% CIs.



To examine whether self-reported gendered associations related to implicit gendered associations we computed a relative measure to parallel the relative structure of the IAT, representing the difference of differences between the relative rationality advantage for female subtracted from the relative rationality advantage for males. This relative gender index correlated modestly with the IAT, $r(114) = .23$, $p = .011$, indicating that individuals who more strongly endorsed gendered rationality at the explicit level also tended to have stronger implicit associations between male and rational relative to female and emotional.

A limitation of the standard IAT in the present context is that it provides a relative rather than an absolute index of gendered associations. That is, the IAT *D*-score does not tell us whether the association is driven by associations between male and rationality, female and emotionality, or both. To address this limitation, we first attempted a data analytic approach by fitting a multinomial processing tree model of IAT errors that allows the independent estimate of the male-rational and female-emotional associations. This approach, known as the Quadruple Process Model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005), suggested that both the *male-thinking* and *female-feeling* associations contributed to IAT performance. However, model fit was not satisfactory, and while it has been argued that these fit statistics are overly conservative for this type of model, we nonetheless elected to conduct a follow-up study to confirm this result (as a source of converging evidence we provide details of the processing tree approach in SOM-R).

Study 1b: Independent associations between each gender and rationality versus emotionality.

The B-IAT drops one of the comparison categories and so provides an estimate of the association between one target category (either *male* or *female*) and the two properties (*rational* versus *emotional*). Participants completed one of those two B-IATs; standard exclusion criteria (Greenwald et al., 2003) led to the exclusion of 16 participants who had an excessive number of fast trials (>10% at < 300 ms), and one participant who terminated the study before providing complete data. Thus, we base analyses on the 185 participants with complete data ($N_{thinking}$ = 98, $N_{feeling}$ = 87). Results indicated that male was preferentially associated with reason over emotion, D = .23 (SD = .31) and also that female was preferentially associated with emotion over reason, D = .28 (SD = .36), both ts > 7.1, ps < .0001; see Figure 1, right panel. The strength of these two associations did not differ, Welch's $t(172.28) = 1.10$, p = .27, nor was there any evidence that the strength of each association differed by participant gender both ts < 1.2, ps > .24.

Participants self-reported associations between gender and *thinking* versus *feeling* demonstrated evidence of a similar gendered pattern, depicted in Figure 2, right panel, with *male* associated with *thinking* (M = 70.78, SD = 20.22) more than *feeling* (M = 49.39, SD = 22.25) and *female* associated with *feeling* (M = 78.65, SD = 17.60) more than *thinking* (M = 66.23, SD = 21.21); all pairwise comparisons between these ratings were significant at $p < .0001$, except the comparison between *male-thinking* and *female-thinking*, which, while still statistically significant, was notably smaller, paired $t(184) = 2.49$, $p = .014$. In this sample there was stronger evidence that these ratings might differ by participant gender, especially for the two scales focusing on females; females indicated a stronger association between both *female* and *feeling* (M = 83.31) and *female* and *thinking* (M = 72.42) than did males (Ms = 74.63 and 59.41, respectively), $ps < .003$. Male participants also somewhat more strongly associated *male* with *thinking*, Welch's $t(180.36) = 1.98$, $p = .049$, but there was no gender difference for explicit associations between *male* and *feeling*, $p = .63$.

Unlike in Study 1a, implicit and explicit gender associations did not relate to one another for either the thinking or feeling test, both $|r| < .06$, ps $> .63$. Overall, then, Study 1b indicates that US adults, whether measured with implicit or explicit measures, associated both male with rationality and female with emotionality.

Study 2

Study 2 sought a conceptual replication and extension of the main finding of Study 1 using an alternative measure that is more closely to linked to the primary phenomenon of interest, namely semantic associations between concepts. We chose the Affect Misattribution Procedure, or AMP (Payne, Cheng, Govorun, & Stewart, 2005), a procedural variant of evaluative or semantic priming in which participants make forced choice judgments concerning the valence (or in this case, the meaning) of unfamiliar Chinese characters after being primed with images expected to affect those judgments. More

precisely, participants were primed with photographs of men and women and then decided whether Chinese characters had meanings related to rationality versus emotionality. If photographs of men (women) activate semantic associations relating to rationality (emotionality), those associations might be *misattributed* to the Chinese character, a complex, ambiguous stimulus about which the participant has no direct knowledge. Thus, we predicted that more characters would be judged as being about rationality if followed by a male prime and about emotionality if followed by a female prime. By including neutral primes our design also includes a baseline comparison allowing us to examine both effects (*male—rationality*, *female—emotionality*) independently.

Participants

Study 2 involved 177 adults (male = 86, female = 83, unknown = 8; M_{age} = 36 years (S = 11 years), White = 75%, Black = 10%, other = 15%) recruited via Amazon's Mechanical Turk online labor market. Recruitment was restricted to IP addresses within the United States and to workers with greater than 95% approval on prior tasks.

Measures

The primary dependent measure was the Affect Misattribution Procedure (AMP). The AMP consisted of a short block of 10 practice trials followed by a test block of 60 trials. Each trial involved the presentation of a prime for 75ms followed by 125ms blank screen followed by a Chinese character for 100ms followed by a mask which remained on the screen until the participant responded by pressing a left or right response key indicating their decision concerning whether they thought the character had a meaning related to thinking or feeling. Primes were full color frontal photographs of six men or six women or a grey square which served as a neutral prime; Chinese characters were randomly selected without replacement from a set of 100. Participants completed 20 trials with each prime type. Photographs were neutral faces drawn from the Chicago Face Database (Ma, Correll, & Wittenbrink,

2015) and approximately matched on unusualness, age, attractiveness, happiness, and sadness based on the ratings provided by those authors. Details of stimulus items is provided in SOM-R.

Procedure

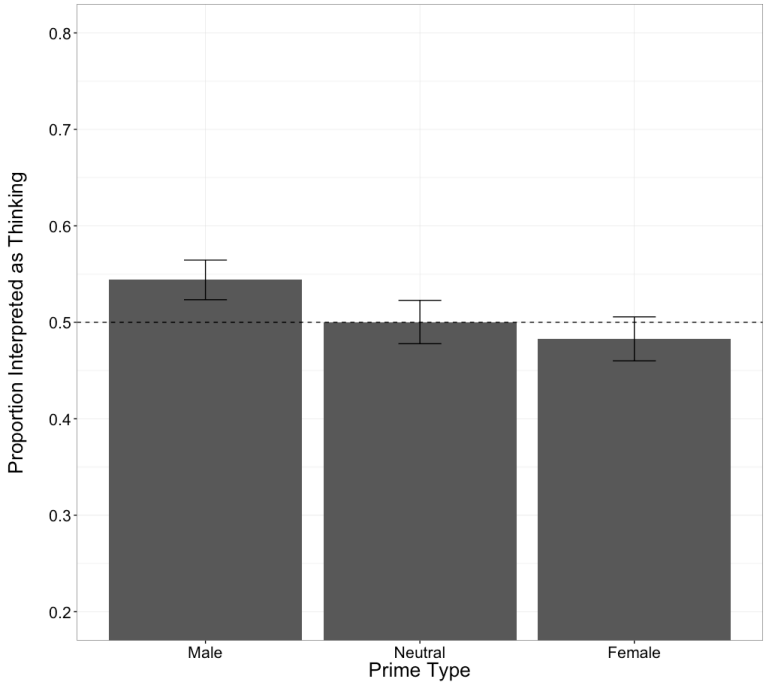
The procedure was identical to that described in Study 1, above, except that the task always began with the AMP instead of an IAT, and a brief cover story was provided. The cover story indicated that the research focused on whether people can intuit the meaning of Chinese characters through their historical link with pictograms; the primes were described as signals that indicated a target character would appear, and participants were warned to try to avoid having the primes influence their responses (Payne et al., 2005). Following the AMP participants completed the same set of demographic and explicit items described in Study 1, above, and an additional item asking whether they had familiarity with a language that made use of Chinese characters (participants answering 'yes' to this question were excluded).

Results

5 participants indicated knowledge of a language that employs Chinese characters and so were dropped from the AMP portion of the analysis. Demographic data for an additional 6 participants were lost due to a data recording error and so those participants do not figure in any analyses involving demographic factors. Overall, and as predicted, participants were most likely to judge a character to be associated with rationality when it was preceded by a male prime (54%), followed by a neutral prime (50%), followed by a female prime (48%; Figure 3). To respect the dichotomous nature of the AMP (forced choice judgments of rationality versus emotionality) data were analyzed in a mixed logistic regression with trials nested within participants, with a suppressed intercept so that each parameter reflects a comparison to chance responding (50%). This analysis confirmed the trends visible in Figure 3; participants were more likely to judge characters as relating to rationality following male primes, b =

.18, CI [.099; .26], $p < .0001$ and also (marginally) less likely to judge characters as relating to rationality following female primes, b = -.070, CI [-.15; .012], $p = .093$. Neutral primes had no effect on character judgments, b = .00, CI [-.080; .083], $p = .97$. Odds ratios can be used to quantify these effects more intuitively: participants were 1.3 times as likely to categorize a character as relating to rationality if it was preceded by a male as compared to a female prime. In summary, Study 2 provides converging evidence concerning a semantic association between male and rational, and somewhat less prominently, between female and emotional.

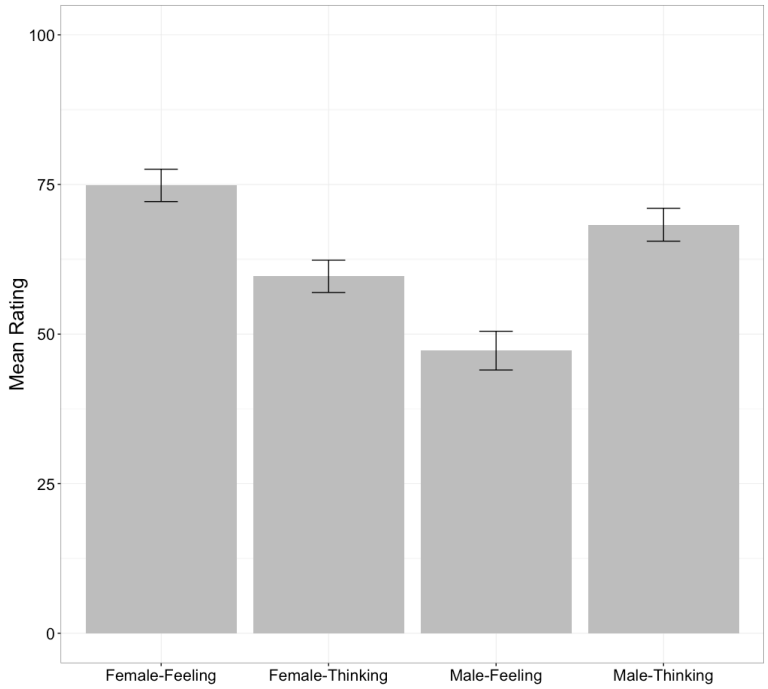
Figure 3: Proportion interpreting characters as referring to rationality, as a function of prime type. Error bars are 95% CIs.



A secondary model in which gender was included as a covariate revealed that the effect of female primes differed by gender, b = .16, $p = .049$. To understand this interaction, we fit an independent model for male and female participants; it revealed that while both genders were more likely to interpret characters as reflecting rationality when they were preceded by male primes, both b > .14, $p < .013$, the effect of female primes on judgments was statistically significant in women, b = -.16, p

= .007, but not men, b = .01, p = .93. This finding implies greater cross-gender implicit consensus concerning the *male-rationality* link than the *female-emotionality* link, but given the just-significant nature of this effect we do not interpret it with confidence.

Figure 4: Average self-reported ratings for each of four gender ratings in Study 2. Error bars are 95% CIs.



Participants self-reported associations between gender and thinking were similar to those in Study 1, and are depicted in Figure 4. Participants associated *male* with *thinking* (M = 68.27, SD = 22.67) more than *feeling* (M = 47.23, SD = 24.44) and *female* with *feeling* (M = 74.85, SD = 21.35) more than *thinking* (M = 59.65, SD = 24.79); all pairwise comparisons between these ratings were significant, paired t > 3.39, p < .0009. These ratings also differed by gender in the case of the two ratings associated with *thinking*; females associated *female* with *thinking* (M = 70.27) more than did males (M = 49.13), p < .0001, and also associated *male* with *thinking* (M = 63.86) less so than did males (M = 72.99), p = .009. The other two explicit associations did not differ by gender, p > .08. Because the explicit measures were identical across all three data collections, and because explicit responses did not reliably differ

across samples, we provide some additional analyses and figures focusing on gender differences collapsed across all three samples in SOM-R.

To compare implicit and explicit responses we computed each participant's implicit gendered association by subtracting the percent of characters judged as rational following female primes from the percent judged rational following male primes and correlated that value with the same participant's explicit scores, specifically the relative male-rationality advantage described in Study 1a, above. As in that study, a modest correlation was revealed, $r(167) = .19$, $p = .015$.

Supplementary Materials

This SOM-R includes a list of all word stimuli used in the IAT in Study 1a and 1b, as well as additional details on stimulus selection for those items as well as the photographs used in Study 2. In addition, it provides details concerning the Quadruple Process Model analysis of the IAT data in Study 1a, and some additional analyses looking at gender differences on explicit measures conducted on the pooled data from Studies 1a, 1b, and 2.

Stimuli Selection, Study 1a and 1b

Studies 1a and 1b used the same verbal stimuli. The category stimuli designed to invoke the categories *male* and *female*, respectively, were taken from past research with the IAT, and were:

Male: Man, Boy, Father, Male, Grandpa, Husband, Son, Uncle

Female: Woman, Girl, Mother, Female, Grandma, Wife, Daughter, Aunt

The attribute stimuli designed to invoke the categories *thinking* and *feeling*, respectively, were produced by the authors based on synonyms of those focal terms and were then roughly matched on valence, frequency, and length. The items to establish non-difference on each of these dimensions are:

Rational: Rational, Deliberate, Calculate, Think, Logical, Reason

Emotional: Emotional, React, Intuition, Feel, Empathic, Sense

Table S1: Characteristics of Word Stimuli in IATs (Studies 1a and 1b)

Word	Category	Valence	Log Frequency	Length
rational	Thinking	5.85	3.82	8
deliberate	Thinking	4.8	3.63	10
think	Thinking	6.68	5.81	5
reason	Thinking	6.27	4.92	6
calculate	Thinking	4.89	3.50	9
logical	Thinking	6.38	3.80	7
Thinking Average		**5.81**	**4.25**	**7.5**
emotional	Feeling	6.62	4.41	9
react	Feeling	5.15	3.86	5
feel	Feeling	6.27	5.19	4
intuition	Feeling	6.42	3.30	9
sense	Feeling	6.09	5.04	5
empathic	Feeling	7.29	2.66	8
Feeling Average		**6.31**	**4.08**	**6.67**

Stimuli selection, Study 2

Faces drawn from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) were matched on unusualness, sadness, attractiveness, age, and happiness; they were, however, well-differentiated on femininity and masculinity. Table S2 provides a list of items and their ratings on each of these dimensions.

Table S2: Mean Stimulus Ratings for Photographs used in Study 2 (AMP)

CFD Stimulus Code	Gender	Unusualness	Attractiveness	Age	Sadness	Happiness	Femininity	Masculinity
WF-039	F	2.03	3.35	23.12	3.12	2.48	4.55	2.14
WF-003	F	2.63	5.02	26.09	2.58	3.55	5.56	2.00
WF-011	F	1.78	4.06	23.75	2.52	2.46	5.15	1.75
WF-006	F	2.82	3.46	24.03	3.75	2.24	4.67	2.06
WF-038	F	2.18	3.27	28.27	2.00	2.88	4.14	2.53
WF-027	F	2.05	4.63	21.58	2.08	3.05	5.23	1.58
Female	**Average**	**2.25**	**3.97**	**24.47**	**2.67**	**2.78**	**4.88**	**2.01**
WM-003	M	2.79	3.57	22.98	3.05	2.76	2.02	4.91
WM-004	M	2.23	4.68	25.40	2.74	2.80	2.06	4.76
WM-006	M	2.86	3.65	25.53	2.54	2.88	2.63	4.74
WM-009	M	2.42	4.18	23.76	2.65	2.26	2.21	4.64
WM-029	M	1.98	4.59	28.63	2.56	2.63	1.73	5.24
WM-024	M	2.04	3.75	20.67	2.81	2.86	2.26	4.16
Male	**Average**	**2.39**	**4.07**	**24.49**	**2.73**	**2.70**	**2.15**	**4.74**

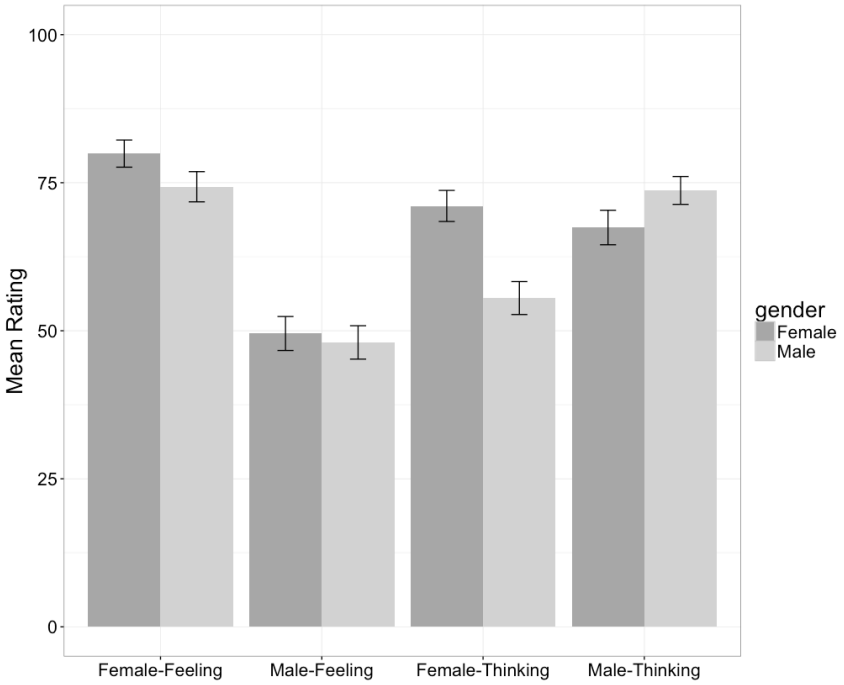
Identifying unique contributions from male—reason and female—emotion.

Due to the relative nature of the IAT standard scoring criteria based on differences in reaction time are not able to dissociate the two contributing associations (in this case the proposed link between *male* and *rationality* and *emotionality* and *female.* However, a recent data-analytic procedure seeks to remedy this gap: the *Quadruple Process Model,* or Quad Model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005). The Quad Model analyzes error rates rather than reaction times, building a multinomial processing tree model that estimates an independent parameter estimate for the influence of each association. A Quad Model analysis of our data suggested that our results reflected the joint influence of both associations, as the estimates of each parameter (and their 95% confidence intervals) both exceeded 0: $A_{\text{male-rational}} = .10$, 95% CI [.083; .13]; $A_{\text{female-emotional}} = .071$, 95% CI [.050; .093]. Unfortunately, the goodness-of-fit test assessing whether the model fit the data suggested poor fit, $G^2 = 207.14$, $df = 3$, $p < .001$. It has been argued that this test is overly conservative with larger data sets such that it rejects the null of satisfactory fit too frequently (Gonsalkorale, Sherman, & Klauer, 2009), and on that reading these results could stand. However, given the uncertainty surrounding this claim we elected to conduct a second study (Study 1b) to examine this issue more directly. The fact that Study 1b and the Quad Model analysis provide convergent evidence is therefore reassuring.

Combined analyses of explicit ratings focusing on gender differences.

Participants in all three data collections completed an identical set of explicit rating items. To gain insight into gender differences in these items we pooled the sample across the three data collections, producing a final sample of 497 participants (240 male, 254 female, 4 unreported). Mean ratings on each of the four items are presented in Figure S1. This aggregated analysis revealed consistent gender differences on three of the four items.

Figure S1: Ratings on Explicit Items
(by Participant Gender; + 95% CIs)



More specifically, while there was no difference between men and women in terms of the tendency to associate *male* with *feeling*, $t(489.2) = .74$, $p = .46$, there were moderate tendencies for women to more strongly endorse a link between *female* and *feeling* ($t(486.8) = 3.22$, $p = .001$, $d = .29$) and for men to more strongly endorse a link between *male* and *thinking* ($t(464.0) = -3.29$, $p = .001$, $d = -.30$). Further, there was a strong tendency for women to more strongly endorse a link between *female* and *thinking* than did men, $t(493.4) = 8.00$, $p < .0001$, $d = .72$.

Another way to consider these differences is to note that women saw a large difference in terms of whether *male* versus *female* as associated with *feeling* but saw very little difference in the extent to which *male* versus *female* are associated with *thinking*; by contrast, men showed very strong tendencies to see both *thinking* and *feeling* as heavily gendered.

General Discussion

In addition to extending past work on implicit gender stereotypes and confirming hypotheses generated by feminist philosophers, the main findings of this paper indicated that there are robust semantic associations between male and rationality, as well as between female and emotionality. The results of Study 1a suggested that both men and women implicitly associated male with rationality and female with emotionality more than vice versa. Study 1b broke down these relative tendencies into their constituent parts, finding that there is a tendency to associate male with rationality more than female, and a tendency to associate female with emotionality more than male. These tendencies were of roughly equal magnitude and were similarly present in both men and women, though only men admitted to holding these stereotypes explicitly. This hints at an intriguing implicit-explicit divergence for women. Study 1 used an IAT as well as the methods of self-report to capture these results. However, like most IATs, the data from each category was relative in comparison to the other. In an attempt to mitigate this issue, Study 2 used an AMP test to confirm the semantic links between thinking/feeling and male/female concepts. In this way, Study 2 accomplished its goal of providing convergent evidence. Self-reported associations between gender and reason/emotion in Study 2 once again reflected these same stereotypical links in men, but much less so in women.

The results of Study 1 and 2 indicate that, while men and women hold implicit gendered conceptions of both reason and emotion, men's implicit and explicit cognitions tend to be more consistent, while women occupy a more ambivalent position. In other words, women seem to reject strong notions of gendered rationality in their self-reports, but nonetheless show robust evidence of male-reason and female-emotion associations at the level of more basic semantic associations. This form of implicit-explicit ambivalence has been observed in several other domains, for example in race bias, in which most White Americans hold relatively egalitarian explicit views but more negative views when

measured with the tools of implicit social cognition (Nosek, Banaji & Greenwald, 2002). It is unclear why women display this divergence in terms of the thinking/feeling bias in particular, but an exploration of the issue would be an intriguing topic for future research.

In terms of potential implications for our current research: our results compliment recent work by some of our colleagues surrounding gender stereotypes, such as the overwhelming gender gap in mathematics and the hard sciences. For example, in Sarah-Jane Leslie and Andre Cimpian's paper on *Expectations of Brilliance*, the authors claim that general attitudes about a discipline reflect the representation of women in those fields. More specifically, they found that scientific fields that are believed to require attributes such as brilliance and genius are often male dominated, whereas other scientific fields that are believed to require more empathy or hard work often have a more equal gender balance (Leslie & Cimpian, 2015). This dovetails nicely with the data in our paper, because brilliance/genius are obviously associated with thinking, and therefore with maleness. Empathy is associated with feeling, and therefore femaleness. Since both genders implicitly perceive women to be more closely associated with feeling, it seems possible that they also implicitly believe women to be less competent in "innately brilliant" fields. Thus, the thinking/feeling bias is potentially correlated with the gender gap in STEM fields.

Building on this hypothesis, it seems likely that the thinking/feeling bias is not only correlated, but also causality related to the difficulty women scientists experience while attempting to advance in their respective fields. Women are linked to feeling, which is linked to subjectivity. Therefore, it is often more difficult for women scientists to establish the objective legitimacy of their research. This is the most obvious potential implication for the thinking/feeling bias, and, as such, it is a frequent topic of feminist discourse. However, there are other potential implications for the thinking/feeling bias, many of which have not yet been seriously discussed. For example, many young women are subconsciously

aware of the thinking/feeling bias on some level. They recognize that, in order to be respected as scientists, they need to sacrifice certain stereotypically feminine characteristics – characteristics that are associated with feeling, such as warmth and empathy. And it is conceivable that some women are simply unwilling to make that sacrifice. It is possible that the prospect of being forced to give up an essential element of their gender identity dissuades them from science before they are ever explicitly discriminated against. Thus, the thinking/feeling bias is worth exploring further as a potentially causal mechanism behind the STEM gender gap.

A possible critique of this conclusion is that of intersectionality. As Kim Crenshaw said in her famous paper on *Mapping the Margins*, "the problem with identity politics is not that it fails to transcend difference…but rather the opposite – that it frequently conflates or ignores intragroup differences," (Crenshaw, 1991). Not all women are affected by the thinking/feeling bias in the same way, and their experience with the bias is often shaped by other dimensions of their identities, such as race and class. Any extrapolation of our results in terms of potential implications should take this into account. Also, though Study 1 avoided any explicit racial indicators, the faces shown in Study 2 as primes were white. An interesting follow-up study might try to replicate our results with faces of different ethnicities, as it is likely that the results may change.

Another possible critique of our research centers around concerns with the IAT. In recent years, several influential papers have been published questioning the effectiveness of IAT methodology. For example, in Hal Arkes and Philip Tetlock's paper on *Attributions of Implicit Prejudice*, the authors offer three potential issues with IAT/reaction time data. One, "the data may reflect shared cultural stereotypes rather than personal animus" (Arkes & Tetlock, 2004). Two, "the affective negativity attributed to participants may be due to cognitions and emotions that are not necessarily prejudiced," (Arkes & Tetlock, 2004). And three, "the patterns of judgment deemed to be indicative of prejudice pass tests

deemed to be diagnostic of rational behavior," (Arkes & Tetlock, 2004). The authors then use these issues to come to the conclusion that reaction time data is not actually indicative of how people behave, nor is it an effective measure of prejudice.

These critiques are not applicable to the work in this paper. After all, the goal of our data is not to predict behavior or even conscious prejudice. Rather, this paper is an attempt to reveal basic cultural associations, something that even Arkes and Tetlock agree that the IAT is effective for: "the participants tested using these methodologies are not providing responses indicative of their attitudes but instead are responding to cultural stereotypes to which they have been exposed but with which they may or may not agree," (Arkes & Tetlock, 2004). In our case, explicit agreement with the cultural bias is less important than the fact that it implicitly exists – our interest here was charting conceptual associations to the oft-made claim that women's views are treated as more subjective and men's views as more objective. We assert that rationality is inherently deemed objective while emotionality/feelings are considered more personal and therefore inherently subjective. In the future, it is possible to envision linking these conceptual associations to behavior, (e.g. to evaluations of male versus female intellectual work products or judgments of the relative objectivity versus subjectivity of claims made by men versus women,) but more work is necessary before that claim can be made with confidence.

Nevertheless, though more data is necessary, it should be noted that there is no shortage of written work on the thinking/feeling gender bias. As was already mentioned in the introduction, women such as Carole Gilligan, Evelyn Fox Keller, Genevieve Lloyd, and Judith Butler have all written extensively on the above hypothesis. The fact that their work has not gained widespread recognition or inspired earlier research indicates a general neglect of feminist philosophy that seems to be a trend among researchers of the social sciences. Unfortunately, feminist philosophy is not the only genre of thought confined to the narrow sphere of the humanities; critical theory and race theory are also areas of

work that are often neglected as sources for novel psychological hypotheses. Ironically, the lack of data in this area confirms our own – scientists who draw inspiration from philosophical principles often seek "pure" philosophy (written by white men,) because it is more closely associated with analytical thinking and therefore more worthy of research. Other genres of 'activist' philosophy are associated with emotion, and therefore are often overlooked as sources for legitimate scientific hypotheses.

Conclusion

In sum, this paper used the tools of the social sciences to investigate the feminist philosophical premise that 'reason' and 'femaleness' are culturally at odds with one another. The results of this investigation provide direct evidence that rationality is gendered at the level of basic conceptual associations residing in semantic memory. This data hints at deep, systemic issues in the way that we approach rationality and objectivity in the modern scientific method; when extrapolated to its fullest conclusions, it is possible that the thinking/feeling bias forces some women to reject basic elements of their gender identity in order to advance in certain STEM fields. Change is necessary, and it starts with the acceptance of gendered subjectivity as an inescapable part of scientific research. The acknowledgement of this, along with an acknowledgement intragroup differences more generally, is the next step forward towards a more self-aware (and therefore inclusive,) scientific community. Thus, if we wish to teach the next generation of scientists that gender equality is possible, we must include the thinking/feeling bias in that education.

Author Contributions

O.P.G. and J.S. drafted the hypothesis, O.P.G. wrote the abstract, O.P.G., Y.D., and J.S. co-wrote the introduction, O.P.G. and Y.D. co-designed the experiments, Y.D. wrote the results and supplementary materials sections, O.P.G. wrote the general discussion, O.P.G wrote the conclusion.

References

1. Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or" would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry*, *15*(4), 257-278.

2. B., D. B., & Sundquist, E. J. (1996). *The Oxford W.E.B. Du Bois reader*. New York: Oxford University Press.

3. Bird, Chloe E. "Gender, household labor, and psychological distress: The impact of the amount and division of housework." *Journal of Health and Social Behavior* (1999): 32-45.

4. Blau, Francine D., and Lawrence M. Kahn. "The US gender pay gap in the 1990s: Slowing convergence." *Industrial & Labor Relations Review* 60.1 (2006): 45-66.

5. Blau, Francine D., and Lawrence M. Kahn. *Understanding international differences in the gender pay gap*. No. w8200. National Bureau of Economic Research, 2001.

6. Blau, Francine D., and Marianne A. Ferber. "Career plans and expectations of young women and men: The earnings gap and labor force participation. "*Journal of Human Resources* (1991): 581-607.

7. Bridges, Judith S. "Pink Or Blue Gender-Stereotypic Perceptions of Infants as Conveyed by Birth Congratulations Cards." *Psychology of Women Quarterly* 17.2 (1993): 193-205.

8. Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York: Routledge.

9. Charles, Maria, and David B. Grusky. *Occupational ghettos: The worldwide segregation of women and men*. Vol. 200. Stanford, CA: Stanford University Press, 2004.

10. Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating Multiple Processes in Implicit Social Cognition: The Quad Model of Implicit Task Performance. *Journal of Personality and Social Psychology*, *89*(4), 469–487. http://doi.org/10.1037/0022-3514.89.4.469

11. Cooper, A. J. (1982). *A voice from the South*. New York: Oxford University Press.

12. Cox, Jason, and Helga Dittmar. "The functions of clothes and clothing (dis) satisfaction: a gender analysis among British students." *Journal of Consumer Policy* 18.2-3 (1995): 237-265.

13. Crenshaw, Kimberley. "Mapping the margins: Intersectionality, identity politics, and violence against women of color." *Stanford law review* (1991): 1241-1299.

14. Francis, Becky. "Gender, toys and learning." *Oxford Review of Education*36.3 (2010): 325-344.

15. Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

16. Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, *45*(2), 1–5. http://doi.org/10.1016/j.jesp.2008.11.004

17. Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. http://doi.org/10.1037/0022-3514.74.6.1464

18. Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. http://doi.org/10.1037/0022-3514.85.2.197

19. Howes, Carollee, and Leslie Phillipsen. "Gender and friendship: Relationships within peer groups of young children." *Social Development* 1.3 (1992): 230-242.

20. Jost, J. T., & Kay, A. C. (2005). Exposure to Benevolent Sexism and Complementary Gender Stereotypes: Consequences for Specific and Diffuse Forms of System Justification. *Journal of Personality and Social Psychology*, *88*(3), 498–509. http://doi.org/10.1037/0022-3514.88.3.498

21. Keller, E. F. (1985). *Reflections on gender and science*. New Haven: Yale University Press.

22. Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, *43*(5), 825–832. http://doi.org/10.1016/j.jesp.2006.08.004

23. Leach, C. W., Carraro, L., Garcia, R. L., & Kang, J. J. (2015). Morality stereotyping as a basis of women's in-group favoritism: An implicit approach. *Group Processes & Intergroup Relations*, 1–20. http://doi.org/10.1177/1368430215603462

24. Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*,*347*(6219), 262-265.

25. Lloyd, G. (1984). *The man of reason: "male" and "female" in Western philosophy*. Minneapolis: University of Minnesota Press.

26. 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(4)*, 1122-1135. http://doi.org/10.3758/s/13428-014-0532-5

27. Nadler, J. T. (2010). *Explicit and implicit gender bias in workplace appraisals: How automatic prejudice affects decision making* (Doctoral dissertation, Southern Illinois University Carbondale).

28. Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101–115. http://doi.org/10.1037//1089-2699.6.1.101

29. Padovani, Flavia, Alan Richardson, and Jonathan Y. Tsou. "Objectivity in Science." (2015).

30. Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. http://doi.org/10.1037/0022-3514.89.3.277

31. Pinker, S., & Spelke, E. (2005, April 22). THE SCIENCE OF GENDER AND SCIENCE. Retrieved April 20, 2016, from https://www.edge.org/3rd_culture/debate05/debate05_index.html

32. Porter, Theodore M. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press, 1996.

33. Richmond, A. (2014, July 21). Women Still Unequal After All These Years - Career Intelligence. Retrieved April 07, 2016, from http://career-intelligence.com/women-still-unequal/

34. Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*

35. Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin*, *27*(9), 1164-1178.

36. Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology*, *56*(4), 283–294. http://doi.org/10.1027/1618-3169.56.4.283

37. Smith, C., & Lloyd, B. (1978). Maternal behavior and perceived sex of infant: Revisited. *Child development*, 1263-1265.

38. Stanley, J. (2015). *How propaganda works*. Princeton, NJ: Princeton University Press.

39. Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. http://doi.org/10.3758/s13428-012-0314-x

40. White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex roles*, *55*(3-4), 259-266.