UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

ORIGINAL RESEARCH REPORT

Rationality is Gendered

Olivia Pavco-Giaccia, Martha Fitch Little, Jason Stanley and Yarrow Dunham

Shared rationality is the common ground of scientific progress. However, some theorists have argued that this common ground may not be level, in that subtle assumptions embedded within lay views of rationality marginalize some would-be participants. Specifically, feminist philosophers have argued that rationality is associated with male rather than female discourse. This claim has frequently been dismissed as incoherent, but a straightforward interpretation is readily available: The concept *reason* is semantically associated with the concept *male*. We support this hypothesis in four studies (total $N > 900$), finding that at both the explicit and implicit level, *reason* is preferentially associated with *male*, *feeling* is preferentially associated with *female*, male faces prime unrelated judgments of reason/rationality, and gendered associations are related to interest in academic disciplines as well as estimates of the (mis)representation of women within those disciplines. Implications for gender stereotyping and the representation of women in different fields are discussed.

From the first moments of life children are bombarded with rich cues that pervasively convey gender roles and stereotypes. From the color of congratulations cards and nursery walls to the toys, names, and clothing they are exposed to, gender is presented as an important cultural distinction that must be mastered (Bridges, 1993). Further, the world also sees children through a prism of gender: from early infancy male and female babies, though difficult to tell apart in the absence of imposed gender cues such as clothing or names, are interpreted quite differently by observers, with infants thought to be male interpreted as more active and agentic, and infants thought to be female as more delicate and sweet (Rubin, Provenzano, & Luria, 1974). It is little surprise then, that gender becomes a defining feature of life, affecting friendship patterns (Howes & Phillipsen, 1992), clothing (Cox & Dittmar, 1995), and play preferences (Francis, 2010). Beyond these relatively benign and presumably voluntary gender-typed preferences, gender also becomes a source of inequality. Most strikingly, in the United States gender correlates with a significant pay gap (Blau & Kahn, 2006), unequal professional advancement opportunities (Bohnet, Van Geen, & Bazerman, 2016), unequal division of household labor (Bird, 1999), and occupational segregation (Charles & Grusky, 2004).

Many factors underlie these disparities, but likely sources include subtle forms of gender stereotypes that associate women with less prestigious, agentic, or lucrative roles. For example, women appear to be judged as greater in warmth and lesser in competence than men (Cuddy et al., 2009), a state of affairs sometimes referred to as "benevolent sexism" (Glick & Fiske, 2011). Further, women are associated with the home over the workplace (Rudman & Phelan, 2010) and with weakness over strength as well as diminished authority (Leach, Carraro, Garcia, & Kang, 2015; Rudman, Greenwald, & Mcghee, 2000; Rudman & Glick, 2001; Rudman & Kilianski, 2000). One overarching framework for considering these differences is that of agency versus communality, and indeed gender stereotypes associating men with agency and women with communality appear both strong and temporally stable over the last several decades (Haines, Deaux, & Lefaro, 2016). To the extent that these attributes are perceived as differentiating professional spaces, women may even seek to avoid occupations that downplay or elide aspects of personhood that are important to them (such as communal goals; Diekman, Brown, Johnston, & Clark, 2010; Eccles, 2011a), as further suggested by evidence that women are preferentially associated with the humanities and men with the sciences (Kiefer & Sekaquaptewa, 2007; Nosek, Banaji, & Greenwald, 2002a). These considerations may also extend to the gender wage gap. For example, women are thought to seek jobs requiring less effort and to be less concerned with professional progress and intellectual challenge, including the assumption that intellectual work deserves higher pay than other, more "practical" forms of work, such as care-giving (Blau & Ferber, 1991). 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 United States today. This is a

Department of Psychology, Yale University, New Haven, CT, US
Corresponding author: Yarrow Dunham (yarrow.dunham@yale.edu)

possibility bolstered by recent findings that gender gaps in STEM fields can be predicted by the extent to which success in that field is thought to depend on brilliance (Leslie, Cimpian, Meyer, & Freeland, 2015).

Here we seek to contribute to these discussions by exploring another pattern of conceptual association that could support the maintenance and acceptance of unequal treatment in professional domains, especially professional domains that place rational discourse at their center. More specifically, we explore the hypothesis that women are semantically associated with emotion and feeling and men are semantically associated with reason and thinking. Interestingly, while it has rarely made contact with psychological research or theorizing, an argument of just this sort has been central to longstanding Feminist philosophical critiques of gender roles. As early as 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..." (Hutchinson, 1982: 78). Feminist philosophers have also argued that the everyday concept of reason itself is gendered male (Lloyd, 1979), creating 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." This view also dovetails with recent research seeking to explain girls' disengagement from math and science as stemming from perceived incompatibility with prevailing gender stereotypes (Cvencek, Meltzoff, & Greenwald, 2011), as well as findings that some scientific disciplines are themselves associated with masculinity (Young, Rudman, Buettner, & McLean, 2013).

The feminist critique offered by Lloyd and others elicited a backlash, with some critics (Nussbaum, 1994) suggesting that the very notion of gendered rationality is incoherent because science is rigorously grounded in evidence and is self-correcting (Padovani, Richardson, & Tsou, 2010). Here we argue that this backlash misses the mark, and that in fact the cognitive sciences offer a simple and empirically tractable interpretation of Lloyd's argument, namely that the concept of *reason* or *rationality* is semantically associated with the concept *male*. By framing the question in this way, it becomes straightforward to test using methods of social cognition, following in the tradition of decades of work that has demonstrated semantic associations between other concepts, from *doctor* and *nurse* to *cat* and *furry*. Thus, our aim here is provide an empirical test of the longstanding feminist contention that the concept of reason is gendered male. Of course, interpreting the feminist contention in terms of conceptual associations is by no means a given, and other interpretations are no doubt possible. Our contention here is not that all philosophers in this tradition would endorse this interpretation. However, we would argue that our approach reflects at least one plausible translation of the philosophical claim into the language of the cognitive sciences, and thus that evidence in favor of these semantic associations can be taken, all

else equal, as support for the philosophical contention that rationality is gendered.

Prentice and Carranza (2002) provide perhaps the closest analog to the present work. They demonstrate that prescriptive stereotypes of men frequently relate to intelligence and rationality, while prescriptive stereotypes of women frequently emphasize emotive and empathic traits. However, our focus is not prescriptive stereotypes but rather simple conceptual associations. If present, such associations are arguably more general in their potential influence because they do not depend on explicit or ideological commitment to any particular view of gender, and indeed could be held even by those highly committed to gender equality. Ultimately, we suggest that these widely held conceptual associations plausibly support structural gender inequalities, for example by supporting more downstream associations between gendered traits and interest in various realms of professional pursuit (Eccles, 2011a, 2011b; Leslie et al., 2015).

Of course, conceptual associations between gender and rationality would not occur in a semantic vacuum, and are likely related to other previously investigated associations, such as those between gender and agency/communality, briefly reviewed above. The present effort does not attempt to disentangle gendered rationality from those quite plausibly related linkages; rather, it attempts to amass positive evidence that concepts related to rationality and more precisely to thinking versus feeling are themselves gendered, i.e., are semantically associated to male versus female, respectively. As a first investigation into these hypothesized associations we also acknowledge a certain coarseness, in that we assume rough sematic equivalence between concepts that are unlikely to be fully overlapping. For example, in what follows we assume that *reason* and *thinking* refer to roughly the same network of associations, as do *emotion* and *feeling*. Of course, this is likely an oversimplification, if not one that we expect to greatly limit our conclusions.

More generally, we seek to test the possibility of these associations at both the implicit level and the explicit level. This allows us to examine both more conscious and controlled associations as well as associations that our respondents might be unaware of or even actively disavow. Further, it allows us to test the relationship between those two levels of analysis, i.e. to examine whether implicit and explicit associations are correlated. Whether these constructs and the relationships between them further differ by respondent gender is also of interest, as we might imagine that these gendered associations could be stronger in men, who most directly benefit from the inequalities that could result from associating *male* with *rational.*

To summarize our empirical program of research, Study 1a uses the Implicit Association Test (IAT: Greenwald, McGhee, & Schwartz, 1998) as well as explicit measures to test the hypothesis that participants preferentially associate *male* with *thinking* and *female* with *feeling*. Study 1b uses a modified IAT procedure to decompose that relative inquiry into independent associations between *male* and *thinking* more than *feeling* between

female and *feeling* more than *thinking*. Study 2 provides a conceptual replication again using explicit measures as well as an alternative implicit measure, a priming procedure, to investigate whether the rapid presentation of photographs of men (women) make concepts related to thinking (feeling) more accessible, influencing ostensibly unrelated semantic judgments. Finally, Study 3 explores the consequences of these gendered associations through a pre-registered investigation of whether these explicit or implicit associations relate to men and women's interest in and estimates of the relative prevalence of men and women in several academic disciplines. Taken together, these results provide evidence that rationality is semantically gendered at the level of basic conceptual associations, and that these associations could be consequential for individual interest in and perceptions of various professional disciplines.

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 all studies reported here sample size was based on several recent studies that revealed evidence of implicit gender stereotypes (Jost & Kay, 2005; Kiefer & Sekaquaptewa, 2007; Leach et al., 2015; Rudman et al., 2000; Rudman & Glick, 2001), for which the median sample size was 98 (mean = 88). With this in mind we sought a target N of 100 per condition. We also enrolled a somewhat larger sample in Study 1a because we were concerned with the possibility of problems with a browser plugin we were using for the first time (see *procedure*, below), though no issues related to this arose.

Measures
Participants in both studies completed an Implicit Association Test (Greenwald et al., 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, stimuli were 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*). 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 this and subsequent studies are in the online supplement.

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 sentence 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
Data from all studies in this paper, as well as code to generate the primary results and figures, is available at https://osf.io/2fr8a/?view_only=8e039f7da9ce4f29b010 16dad14d94a1. This manuscript reports all the studies that

were conducted as part of our investigation of gendered rationality, with the exception of a pilot study investigating the use of a lexical decision task to explore gendered associations. Because we were unable to replicate a basic lexical decision effect (outside the realm of gendered associations) in an online environment, perhaps due to the more precise reaction time recording necessary for that task, we did not continue developing this paradigm to test our main questions of interest, instead shifting to the approach described in Study 2.

Study 1a: Relative associations between gender and rationality

Standard exclusion criteria (A. G. 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, $D = .33$ ($SD = .31$), $t(115) = 11.50$, $p < .0001$, 95% CIs [.28; .39] (**Figure 1**). There was no evidence of a difference between male participants ($M = .33$) and female participants ($M = .34$), Welch's $t(102.87) = .23$, $p = .82$. To better quantify potential support for the null hypothesis of no difference based on participant gender, here and elsewhere in this paper we employ Bayes Factors, in particular the BF01 statistic, indicating in this case the relative strength of evidence in support of the null versus in support of a difference at the $p < .05$ level. Here BF01 = 4.9, suggesting moderate evidence in support of the null that both participant genders showed similar patterns of association.

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 were not significantly different by participant gender (all BF01 > 2.9, $ts < 1.9$, $p > .28$), except for the explicit association between female and thinking, which was weaker in men ($M = 57.43$) than women ($M = 67.06$), Welch's $t(92.91) = 2.39$, $p = .019$.

To examine whether self-reported gendered associations related to implicit gendered associations we computed a comparative 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, though we note that the creating of a difference score to index explicit gendered associations raises some issues concerning the compounding of unreliability (Furr, 2017).

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
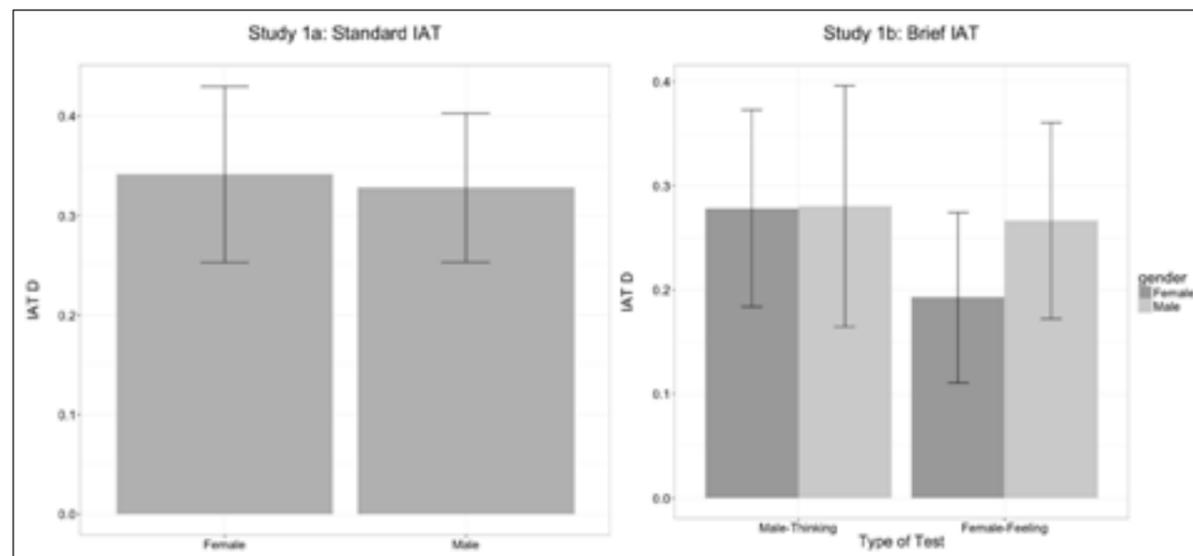

Figure 1: Associations between Gender and Rationality and Emotionality, by participant gender. Error bars are 95% CIs.
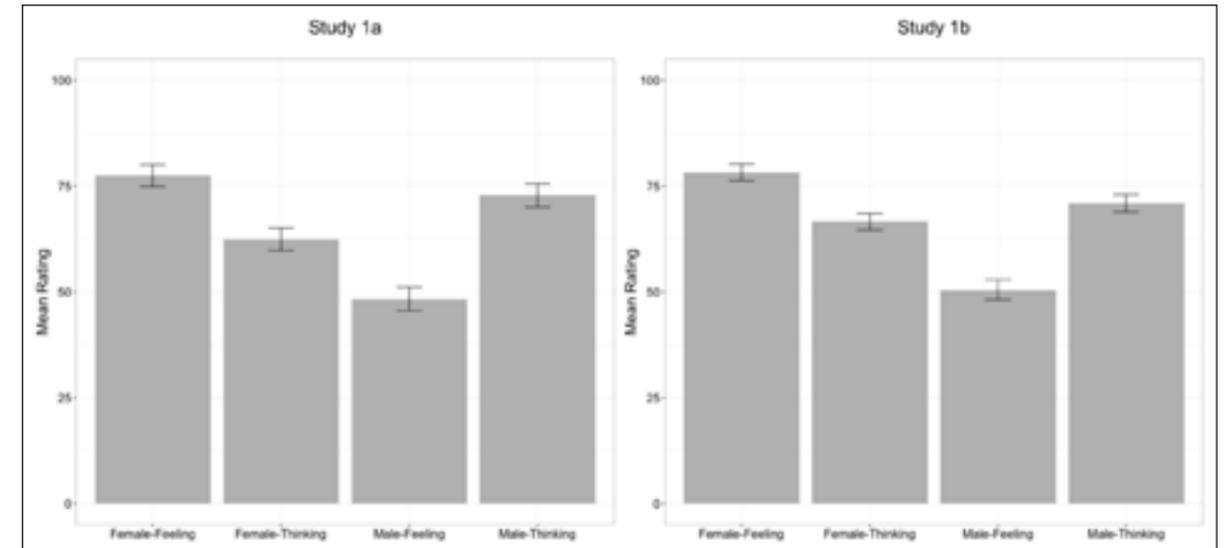

Figure 2: Average self-reported ratings for each of four gender ratings. Error bars are 95% CIs.

for this type of model, we nonetheless elected to conduct Study 1b to confirm this result (as a source of converging evidence we provide details of the processing tree approach in the online supplement).

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 $ts < 1.2$, $ps > .24$, BF01 = 3.5.

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*, $p = .63$, BF01 = 5.6.

Unlike in Study 1a, implicit and explicit gender associations did not relate to one another for either the thinking or feeling test, both $|r| < .06$, $ps > .63$, BF01 > 3.6. Given somewhat conflicting findings on these correlations across studies, we return to patterns of implicit-explicit correlation at some length in the General Discussion. 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 linked to the primary phenomenon of interest, namely semantic associations between concepts as assessed by the Affect Misattribution Procedure (AMP; Payne et al., 2014), which measures whether one concept (in this case *male* or *female*) increases the accessibility of other concepts (in this case *thinking* or *feeling*). 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 preceded by a male prime and about emotionality if preceded 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 219 adults (male = 111, female = 106, unknown = 2; M_{age} = 34 years (S = 12 years), White = 74%, Black = 9%, other = 17%) 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; 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. 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 75 ms followed by a blank screen for 125 ms, followed by a Chinese character for 100 ms, followed by a noise 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 White men or six White 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 of adults 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 are provided in the online supplement. We acknowledge that the use of solely White faces is a limitation on the generalizability of the present study.

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 and an additional item asking whether they had familiarity with a language that made use of Chinese characters (9 participants answering 'yes' to this question were excluded, leaving a total of 210 participants in the AMP portion of the study).

Results and Discussion

Demographic data for one participant was lost due to a data recording error and so that participant does 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 (53%), 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 = .14, CI [.055; .22], p = .0003. While not significant, they trended towards being less likely to judge characters as relating to rationality following female primes, b = −.062, CI [−.14; .020], p = .11. Neutral primes did not appear to be related to judgments, b = .01, CI [−.075; .088], p = .86. Odds ratios can be used to quantify these effects more intuitively: participants were 1.22 times as likely to categorize a character as relating to rationality if it was preceded by a male as compared to a female prime, a difference which was statistically significant, b = .20, CI [.11; .29], p < .0001. In summary, Study 2 provides converging evidence concerning a semantic association between male and rational, and is suggestive of a weaker link between female and emotional.

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 = 66.05, SD = 22.97) more than *feeling* (M = 46.59, SD = 23.51) and *female* with *feeling* (M = 74.74, SD = 20.68) more than *thinking* (M = 61.06, SD = 23.87); all pairwise comparisons between these ratings were significant, paired t > 4.97, p < .0001. These ratings also differed by gender in the case of the two ratings associated with *thinking*; females associated *female* with *thinking* (M = 69.59) more than did males (M = 52.72), p < .0001, and also associated *male* with *thinking* (M = 61.29) less so than did males (M = 70.91), p = .002. Females also associated *female* with *feeling* (78.19) somewhat more than did males (71.54), p = .017. The self-reported association between *male* and *feeling* did not differ by gender, p = .81, BF01 = 6.6.

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(208) = .19, p = .0048.

Study 3

While the prior studies demonstrate the presence of gendered associations with *thinking* and *feeling*, whether or not these associations are related to more face-valid
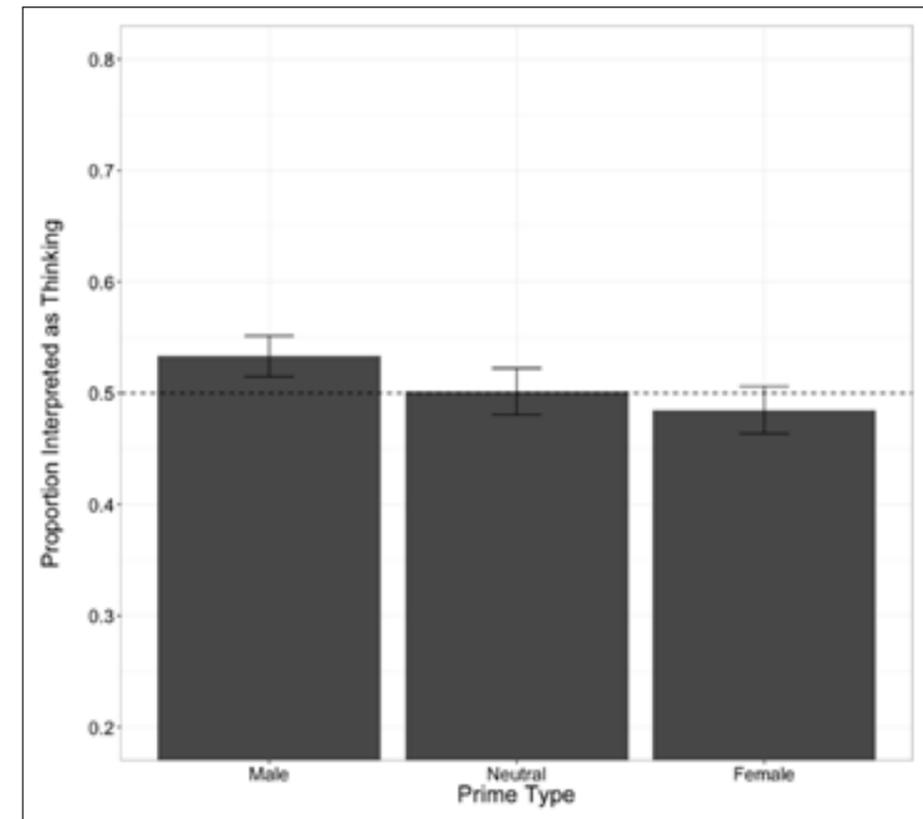
Figure 3: Proportion interpreting characters as referring to rationality, as a function of prime type. Error bars are 95% CIs.
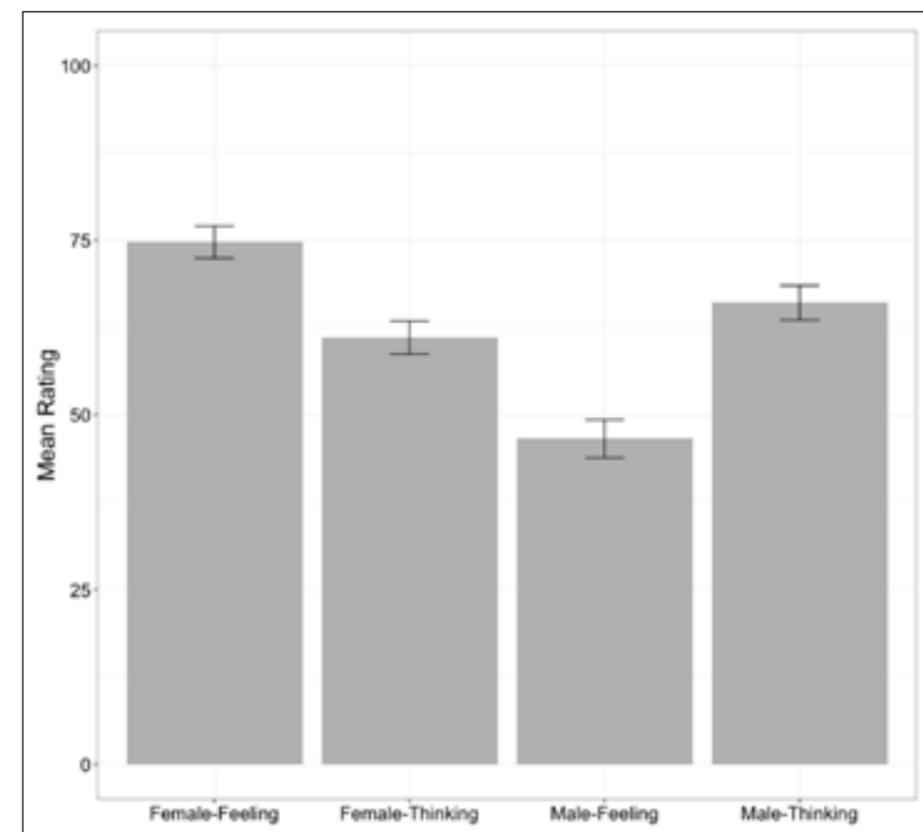
Figure 4: Average self-reported ratings for each of four gender ratings in Study 2. Error bars are 95% CIs.

outcomes remains unknown. Given their close potential connection to interest in various occupations, in this study we investigate whether individuals with stronger implicit or explicit gendered associations regarding rationality also have consonant views of various academic career paths. We elected to focus on academic disciplines because this allowed us to build on prior work looking at representation of women as well as perceptions of the attributes predicting success in the academy (e.g. Leslie et al., 2015). Thus, in addition to measuring participants' implicit and explicit views concerning gendered rationality, we also assessed participants' interest in a range of fields in the social sciences, natural sciences and humanities that were first rated by an independent group of participants concerning the extent to which they were perceived as requiring attributes associated with thinking versus feeling. Our pre-registered prediction was that individuals with stronger gendered associations would tend to show less interest in fields that provide a potential mismatch with their own gender, and would also tend to underestimate the prevalence of women in fields associated with thinking as compared to feeling. Prevalence of women was measured via participant estimates of the percent of PhDs earned by women as well as the percentage of women at the top of the profession, as indicated by tenure-track faculty lines at top departments in that discipline.

Participants

50 adults were recruited via Amazon's Mechanical Turk online labor market to provide initial ratings of 10 professions, as described below. Two of these participants did not complete the survey and so were dropped, leaving 48 adults (male = 34, female = 14, M_{age} = 34.2 (SD = 11.0), White = 69%, Asian = 17%, Black = 8%, other = 6%) in the initial ratings study. 409 adults (male = 213, female = 188, unspecified = 8, M_{age} = 37.8 years (SD = 11.6 years), White = 71%, Black = 10%, other = 19%) recruited via Amazon's Mechanical Turk participated in the main study. As above, recruitment was restricted to IP addresses within the United States and to workers with greater than 95% approval on prior tasks.

Measures

Discipline ratings pre-study

We selected 10 academic disciplines (chemistry, engineering, English, history, linguistics, molecular biology, neuroscience, philosophy, physics, and psychology) by choosing fields that varied in terms of participation rate by women as reported in the National Science Foundation's Survey of Earned Doctorates (available at https://www.nsf.gov/statistics/srvydoctorates/), selecting one natural science and one non-natural science discipline from each quintile of female participation. Participants in the ratings study then rated each of these 10 disciplines on 9 traits intended to capture the thinking dimension (*being agentic, brilliance, competence, logic,* and *reason*) and the feeling dimension (*communal qualities, emotional sensitivity, intuition,* and *warmth*). Ratings were provided via a slider that converted positions to a 100-point scale from "strongly disagree" to "strongly agree" based on

the prompt "In your opinion, how much does each of the following subjects require [trait] as an important trait?" Averaged ratings for each discipline were then submitted to Principal Components Analysis, which strongly suggested a two-factor solution (fully detailed in the online supplement). The first factor related to the *thinking* dimension, with high loadings on being agentic, brilliance, competence, logic, and reason; the second factor related to the *feeling* dimension with high loadings on communal, emotional sensitivity, intuitional and warmth. We thus extracted rotated component scores for these two components to serve as indices of each discipline's position in thinking versus feeling space. However, all reported results are robust to instead using an average of all the thinking ratings (alpha = .97) and feeling ratings (alpha = .93). For our pre-registered analysis we created a difference score to index a discipline's position in thinking-feeling space by subtracting the feeling score from the thinking score (a plot of each discipline in this two-dimensions component space is included in supplemental materials). In addition to these participant ratings we secured the percentage of women earning PhDs in each discipline via the National Science Foundation's Survey of Earned Doctorates (available at https://www.nsf.gov/statistics/srvydoctorates/) to compare to the estimates provided by participants as described below. To estimate the percentage of women who were at the top of each discipline we calculated the percentage of tenured or tenure-track faculty who were women in each of the top 20 schools in each discipline, as ranked by professional ranking organizations and as described in more detail in the online supplement. While this method is certainly imperfect, since securing a tenure-stream position is one clear indicator of success it at least provides a rough proxy for success at the highest levels.

Main study

Participants completed the same gendered rationality IAT described in Study 1, a set of survey items relating to the 10 academic disciplines, the same set of explicit ratings, and the same set of demographics from prior studies. The academic discipline items presented each of the 10 disciplines, one at a time in a random order, each on its own survey page, and asked participants to indicate their interest in the field on a 9-point scale ("Imagine you were choosing a new career. How interested would you be in a career in this field?"), their estimate of the percent of PhD students in the field that are female, and their estimate of what percent of the best 5% of people working in the field are female.

Procedure

The procedure was identical to that described in prior studies with the following exceptions relating to task order: the explicit gendered rationality ratings always immediately followed the IAT, and those two tasks were in turn presented either immediately before or immediately after the discipline ratings, with that order counterbalanced across participants. Demographics were always presented last.

Results and Discussion

We provide summary statistics to parallel the results of prior studies and then turn to confirmatory analyses (http://aspredicted.org/blind.php?x=rk4ku4), followed by some additional exploratory analyses.

Descriptive Statistics

Results of the implicit and explicit gendered association items were similar to those provided above. Beginning with the IAT, 403 participants completed the IAT portion of the study, of whom data for 57 (14%) had to be excluded via standard exclusion criteria relating to an excessive number of fast trials, leaving usable data for 346 participants. Results indicated a robust relative association between male and thinking and female and feeling relative to the opposite pairings, D = .35 (SD = .33), $t(345)$ = 19.79, p < .0001, 95% CIs [.32; .38]. There was a hint that this association was somewhat stronger in male participants (M = .38) than female partcipants (M = .32), Welch's $t(334.54)$ = 1.70, p = .09, but this was a relatively small effect, d = .18 that in fact provided slight evidence in favor of the null of no difference BF01 = 2.1. Further, the association itself was robust in both male and female participants, both t > 12, p < .001.

Participants self-reported associations between gender and thinking paralleled results described above, with *male* associated with *thinking* (M = 71.25, SD = 21.90) more than *feeling* (M = 49.77, SD = 24.10) and *female* associated with *feeling* (M = 79.11, SD = 19.05) more than *thinking* (M = 66.57, SD = 23.14); all pairwise comparisons between these ratings were significant, paired $t(401)$ > 2.91, p < .004. Participant gender differences were evident in all these ratings except for the explicit association between male and feeling (Welch $t(386.5)$ = 1.07, p = .29, BF01 = 5.2), always in the direction of stronger stereotypical gender associations in men than women (all Welch t > 3.35, p < .0009, d > .34). As in Study 1, implicit and explicit gendered associations were modestly corelated, $r(343)$ = .23, p < .001.

Because the explicit measures were identical across all three data collections, and because patterns of participant gender differences in explicit responses were not wholly consistent across studies, we provide some additional analyses combining all explicit data, including figures focusing on gender differences, in the online supplement. This analysis strongly suggests that explicit gendered associations are stronger in male than female participants.

Predicting interest in disciplines

Because interest ratings can be considered clustered within participants as well as within disciplines, we fit a linear mixed model (using the lme4 package in R), predicting interest in a particular field from participant gender, the discipline's relative thinking versus feeling score calculated as described above, and the participant's relative explicit gendered association, centered at the sample mean, with random intercepts for both participant and discipline. As predicted, this analysis revealed a 3-way interaction between these terms, b = .0032, CI [.0012; .0052], p = .0009. This relationship is depicted in **Figure 5**, and suggested that explicit gendered rationality negatively predicted women's interest in fields highly associated with thinking and men's interest in fields highly associated with feeling, but did not predict women's interest in fields highly associated with feeling or men's interest in fields highly associated with thinking. Thus, explicit beliefs about gendered rationality predicted less interest in fields that "mismatched" the participant's gender in terms of these characteristic associations. Visual inspection of **Figure 5** suggests that over the range of available data this effect constituted a difference of about 2 scale points in interest, which amounts to a shift from the median up to the first quartile or down to the third quartile of interest.

Turning to a parallel analysis with implicit gendered rationality (IAT), contrary to prediction the 3-way interaction between implicit gendered rationality, participant gender, and the discipline's relative thinking versus feeling
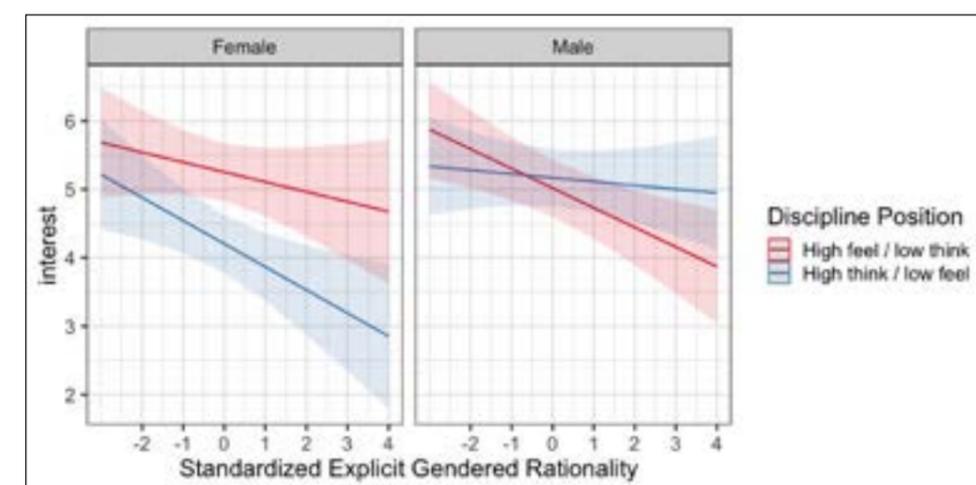
Figure 5: Model = -predicted interest in academic disciplines as a function of (standardized) explicit relative gendered associations linking male to thinking and female to feeling for fields rated as relatively high in thinking versus feeling. Faceted by gender and plotted with 95% confidence intervals.

score did not reach significance, b = −.17, CI [−.47; .12], p = .24. The final model retaining only significant terms included gender, the discipline's relative thinking versus feeling score, and their interaction, b = .43, CI [.34; .50], p < .0001, indicating that as fields increased in being seen as high thinking over feeling women but not men were generally less interested in them. However, there was no evidence that the IAT increased our ability to predict this relationship over and above the independent ratings of the field's relationship with thinking and feeling.

Predicting estimates of representation in disciplines

Descriptively, participant estimates of gender representation at the PhD level were at least somewhat calibrated to actual representation taken from the NSF Survey of Earned Doctorates, as evidenced by a mixed linear model predicting estimates (nested within participants) from gender and actual representation. Only the effect of actual representation was significant, b = .49, CI [.46; .52], p < .0001(note that as these are unstandardized betas a value of 1 would indicate that estimates were perfectly calibrated with actual representation). Estimates of gender representation at the top of the field were better calibrated to actual representation, as evidenced by a mixed linear model predicting estimates (nested within participants) from gender and actual representation, in which the effect of actual representation was strong, b = .74, CI [.70; .78], p < .0001. Here there was also a main effect of gender: men estimated there to be about 4% more women at the top of field, b = 4.26, CI [4.18; 4.34], p = .008. We also note that in general participants *underestimated* the representation of women in the disciplines at the PhD level by 8.7% and at the top of the discipline by 6.8%.

For our main analysis of representation at the PhD and top of the field we focus on predicting *miscalibration*, i.e. the extent to which a participant thought there were more or less women in the field than were actually present. We note that in our preregistered analysis plan miscalibration was described as a secondary analysis, with the primary analysis focusing on directly predicting estimates of the representation of women. However, because analyzing miscalibration merely involves subtracting a constant (i.e., actual representation) from participant ratings for each discipline, these analyses are fundamentally identical, and because miscalibration is more germane to the question of whether gendered associations bias thinking than are estimates, we focus on them here.

First, we examine estimate of PhD representation, and whether they are predicted by explicit gendered associations, participant gender, and the discipline's thinking versus feeling score. This model lead to the observation of two interactions, one between the discipline's score and participant gender, b = 1.50, CI [1.00; 2.00], p < .0001, and one between the discipline's thinking versus feeling score and explicit gendered associations, b = −.02, CI [−.029; −.017], p < .0001. The first interaction with gender shows that the degree of miscalibration in men was highest for fields thought (by independent raters) to emphasize feeling over thinking. In such fields men underestimated the

number of women in those disciplines to a greater degree than women. More critical for our question, however, is the second interaction, which demonstrates that for fields considered (by independent raters) to emphasize thinking, those with stronger gendered associations were more miscalibrated, i.e., they underestimated the representation of women to a greater degree.

Turning to the potential relationships between implicit gendered associations and calibration, this model unsurprisingly revealed the same interaction between discipline and participant gender described above, b = 1.17, CI [.56; 1.70], p < .0001. More interestingly was a second interaction between implicit gendered associations and the field's position in thinking versus feeling space, b = −1.83, CI [−2.79; −.89], p < .0001. This interaction shows that those with stronger implicit gendered associations were more miscalibrated, i.e. the tended to underestimate the representation of women, in fields that previous raters had judged as requiring more thinking, and less miscalibrated for fields previously judged as requiring more feeling. Given the effects of both implicit and explicit gendered associations we fit an additional model incorporating both these interactions; both remained highly significant (p < .008), suggesting independent relationships between implicit and explicit gendered associations and views of the representation of women at the PhD level.

Turning to representation of women at the top of the field (operationalized as tenured professors faculty at top 20 institutions), we again start with the impact of explicit gendered associations. In similar fashion to what was described for PhD representation, we observed an interaction between explicit gendered associations and the field's position in thinking versus feeling space, b = −.026, CI [−.032; −.019], p < .0001. As before, this indicated that those higher in explicit gendered associations were more miscalibrated for fields thought of as relatively high in thinking (by independent raters) and less miscalibrated for fields thought of as relatively high in feeling. Turning to a parallel model with implicit gendered associations, here we observed a 3-way interaction between implicit gendered associations, participant gender, and the field's relative thinking score, b = −3.31, CI [−5.13; −1.35], p = .0008. Decomposing by participant gender, men showed the same pattern described above, tending to be more miscalibrated for fields rated high in relative thinking and less miscalibrated for fields rated low in relative thinking (as indicated by the interaction between IAT scores and the field's position in thinking-feeling space, b = −2.76, CI [−4.05; −1.36], p = .0002). By contrast, women's estimates were not significancy related to both their IAT scores and the field's position in thinking-feeling space, both p > .53. As above, when the effects of both implicit and explicit gendered associations were included in the same model, both retained their predictive effects (p < .002), suggesting both relate to views of the gender composition at the top of these 10 academic disciplines.

In summary, this study examined whether implicit and explicit gendered associations related to individual interest in ten academic disciplines as well as estimates

of the degree of female representation in those same disciplines. Explicit gendered associations related to all of these outcomes, while implicit gendered associations related to estimates of female representation but not of individual interest in the field.

General Discussion

We tested the hypothesis that sematic associations, whether assessed explicitly or implicitly, link the concepts of *male* to *reason* and *female* to *emotion*. Our results provide evidence of robust semantic associations between both these pairs of concepts, thereby putting empirical teeth to a straightforward interpretation of a longstanding claim in feminist philosophy. Study 1a provided evidence that both men and women implicitly associated male with rationality and female with emotionality more than vice versa. Study 1b decomposed this result into a tendency to associate male with rationality more than female and a tendency to associate female with emotionality more than male; these two tendencies were of roughly equal magnitude and similarly present in both men and women. Study 2 provided convergent evidence by replicating the primary effect of interest, i.e. a link between *male* and *rationality*, with a measure derived directly from semantic priming, though its results were more equivocal concerning the link between *female* and *emotionality*. Self-reported associations between gender and reason/emotion also reflected these same stereotypical links, though more powerfully in men than women (see online supplement for an analysis of self-report data aggregated across all studies). Finally, Study 3 provided evidence that explicit gender associations relate to interest in, as well as estimates of the prevalence of, women versus men in these academic disciplines. However, contrary to hypothesis, implicit gender associations did not relate to interest in the disciplines, though they did relate to estimates of the prevalence of women. One reason for this discrepancy could be that judgments of interest could be previously held views that are merely reported as an explicit judgment; if so, this might explain why they relate to explicit rather than implicit views, and why judgments of representation, which participants are unlikely to have explicitly considered, might be more susceptible to implicit influence.

While both men and women hold implicit gendered conceptions of both reason and emotion, men's implicit and explicit cognitions tend to be directionally consistent, while women occupy a more ambivalent position, rejecting strong notions of gendered rationality in their self-reports but nonetheless showing robust evidence of *male-reason* and *female-emotion* associations at the implicit level. 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, 2002b). This pattern emerges in large part because men hold stronger explicit gendered stereotypes in this domain (see supplement for analysis aggregated across all studies), an important finding in its own right given

that, as an explicitly endorsed stereotype, those who hold it may not be motivated to curtail its potential influence on teaching, advising, hiring, mentoring, and so on.

Our partial focus on implicit social cognition opens us to some of the critiques concerning that literature. For example, Arkes, Tetlock, and colleagues have argued that results of the IAT and other implicit measures do not necessarily reflect individual beliefs (even unconscious ones) but rather reflect past exposure to cultural stereotypes (Arkes & Tetlock, 2004); they have also argued via meta-analytic synthesis that the relationship between these measures and behavior are relatively weak (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013; but see also a more recent meta-analysis revising this claim somewhat: Kurdi et al, 2018). However, our findings are useful even were we to grant the full force of these critiques, for two reasons. First, our primary goal was simply to document the widespread presence of a subtle but widely held form of cultural stereotype linking rationality to gender, independent of its role in decision making or its locus within the individual versus the culture. Second, in our final study we do directly demonstrate, in a pre-registered study, that these beliefs—whether conceptualized as explicit judgments or implicit semantic associations— relate to face-valid judgments, including interest in the disciplines themselves as well as beliefs about the degree to which women participate in them (though it must be acknowledged that this study was correlational and thus it would be unwise to interpret it in a causal manner). Third and finally, we note that many of our findings cut across cognitive level, i.e. they appear robustly with respect to both implicit and explicit beliefs.

In general we find modest correlations between implicit and explicit gendered associations, generally in the r = .2 to .3 range. These correlations are somewhat below the average implicit-explicit correlation reported in a large sample across multiple attitude domains, which was .35 (Nosek, 2005). A smaller than average correlation has been interpreted as showing larger than average social desirability concerns; in the present context this implies that explicitly reporting gendered associations might invoke some experimental demand characteristics, even in the anonymous context of these studies. Alternatively, differences in measurement technique (using a 100-point thermometer measure for explicit reporting and a relative IAT measure for most implicit measures) can also depress correlations between implicit and explicit measures (Hofmann, Gawronski, Gschwendner, & Schmitt, M., 2005). It is also worth noting that we did not find statistically significant correlations between implicit and explicit measures in Study 1b that used the single-category IAT procedure. There is in general less work exploring implicit-explicit correlations with this measure so we do not over-interpret here, but it may be that in a case like the present where both patterns of association are on average of equal strength the single-target measure provides a somewhat weaker index of the broader pattern of associations and so is less reliably linked to explicit associations.

Our results complement recent work by past researchers interested in how gender stereotypes relate to occupational

stereotypes as well as occupational choices (Eccles, 2011a, 2011b; Leslie et al., 2015; Storage, Horne, Cimpian, & Leslie, 2016). For example, some of this work finds that fields that are believed to most centrally require brilliance are more likely to be male dominated, whereas fields that are believed to require more empathy or hard work are more likely to have greater gender balance. Our work contributes to the understanding of this phenomenon by demonstrating that reason—the underlying factor driving brilliance or genius—is preferentially associated with maleness, potentially buttressing or even underlying this gender divide. It also raises interesting future questions, such as whether exposure to female role-models in STEM disciplines (which have previously been linked to more identification with and positive attitudes towards those disciplines; Young et al., 2013) might prevent gendered associations from forming as powerfully or otherwise mitigate their effects. Other important points of contact are conceptual links between *male* and *science* and *female* and *humanities* (Kiefer & Sekaquaptewa, 2007; Nosek et al., 2002a). We might expect that the driving force behind these links are in fact the associations explored here, i.e. the broader conception of gendered rationality. Of course, other possibilities should be acknowledged. As a correlational study, our Study 3 is open to third variable critiques in which an independent factor predicts both greater gendered associations and views of the different disciplines included here. This issue might be particularly thorny in the present case given that, as we discuss further below, we did not exhaustively distinguish the thinking versus feeling dichotomy from other potentially related constructs that have previously been used to characterize gender stereotypes, such as agency versus communality. This raised the possibility that one of those beliefs might drive both gendered associations as well as beliefs about the relevant fields.

Another related possibility concerns stereotype accuracy (for discussion, see Jussim et al., 2009), which could itself serve as a critical third variable. The issue here is that if stereotypes concerning the gendered nature of rationality are accurate, and if perceptions of the fields in question are largely accurate, those two factors will relate but not in a manner that suggests a causal link between them. While our study of gendered representation does assess accuracy, or lack thereof, we did not assess the accuracy of stereotypes linking gender to rationality. However, we would argue that assessing the accuracy of such stereotypes is difficult absent much more precise operationalizations of what it means to be "more rational" or "more emotional". Further, efforts to assess these links are made more difficult given the very semantic associations we document here, which will tend to exacerbate perceived links even when not actually differentially present. What's more, accurate gender differences in, e.g. the American adults tested here, would not really tell us whether such differences are inherent to men and women versus are themselves the product of socialization efforts driven by the very same semantic associations. Finally, we follow Hammond and Cimpian (2017) in believing that stereotypes are not merely statistical claims but rather reflect generic beliefs about

groups such that merely assessing accuracy elides some of their more powerful and pernicious effects (because in many cases generic beliefs such as "men are more rational" are better predictors of social judgments than are that same individual's' beliefs concerning the underlying statistics).

The present study has additional limitations. We have already alluded to debates concerning the predictive power of the IAT and other measures of implicit social cognition. While we hope our final study does some work to stave off the strongest form of such critiques, we acknowledge that self-reported judgments of interest (as well as of the relative prevalence of women) provide only a limited window into the potential downstream consequences of gendered associations, if any, or more generally the psychology underlying gender disparities in academic disciplines. For one thing, ratings of interest are somewhat ambiguous, in that an individual might find the content of a domain more or less interesting than they find the actual pursuit of professional life in that domain. Indeed, a dissociation between these two things might be predicted if observers think that some disciplines are more or less accepting of people like them. Further, we used only a small set of academic disciplines that are unlikely to span the full range of fields in thinking versus feeling space. How results would look with a larger range of fields, or with professions outside the academy, remains an open question. Richer and more varied sets of measures, or even more powerfully, actual behavior observed subsequent to attitude measurement, would provide a much more powerful window into how these associations impact real-world professional choices. Nonetheless, we do believe that our findings strongly suggest that gendered associations are not "merely" cultural stereotypes in that they are "in the head" of individuals and relate to stated interest in fields that occupy gendered positions in the social order.

One major concern is the extent to which the associations documented here are distinct from other previously reported links concerning gender, most notably the link between gender and agency versus communality (e.g. Haines, Deaux, & Lefaro, 2016). We would expect that these different families of stereotypes would be related; indeed, we would expect them to be closely linked in the semantic network of individuals and cultures. However, determining which is strongest or most central or most predictive would be a daunting challenge given the reliability of implicit measures and the extent to which semantically associated concepts co-activate one another. Relatedly, in an ideal world the scales we used to measure thinking versus feeling in Study 3, and the terms we used in the implicit measures in Study 2, would have been validated more thoroughly prior to use. Such efforts would not only increase the strength of the inferences we can make on the basis of our findings but might also go some way towards showing whether and if so how our constructs differ from the past work alluded to above.

Nonetheless, the gendered associations we document here are compatible with a longstanding feminist critique that has been influential in philosophy and other areas of the humanities, namely that gender infuses concepts of reason and rationality. Such views have been hotly

contested and even tarred as incoherent (Nussbaum, 1994; Padovani et al., 2010). While our work does not address all interpretations of the gendered rationality perspective, it does provide a cognitively grounded interpretation of one such claim, which is fully supported by the available empirical evidence. Philosophy, and perhaps in particular areas such as feminist theory and critical race theory, have long been concerned with inequality, marginalization, and the power of ideology and stereotypes. These topics are of direct relevance to social and political philosophy, and we hope scholars in the social sciences will begin to mine their insights with more regularity. Further, they can frequently be translated into straightforward psychological and empirically testable claims, the investigation of which can move our field forward (e.g. Jost, 2006). We hope our work will help to build bridges across fields that share topical interest in areas such as prejudice and discrimination but rarely communicate across the disciplinary divides that separate them.

In closing, we note that the general strategy employed here, that of identifying a philosophical hypothesis, reinterpreting it in light of the contemporary science of the mind, and then subjecting it to testing with psychological tools, is one that could fruitfully be employed more broadly.

Data Accessibility Statement
Data from all studies in this paper, as well as code to generate the primary results and figures, is available at https://osf.io/2fr8a/?view_only=8e039f7da9ce4f29b010 16dad14d94a1.

Additional File
The additional file for this article can be found as follows:

· **Supplemental Online Materials.** Additional details on stimuli selection and the supplementary analyses referred to in the text are available here. DOI: https://doi.org/10.1525/collabra.274.s1

Competing Interests
The authors have no competing interests to declare.

Author Contributions
· Contributed to conception and design of Studies 1–2: OP, JS, YD
· Contributed to conception and design of Study 3: YD, ML
· Performed analysis: YD
· Contributed to interpretation of data: OP, JS, YD, ML
· Drafted and revised article: OP, JS, YD, ML
· Approved the submitted version for publication: OP, JS, YD, ML

References
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. http://doi.org/10.2307/20447239?ref=search-gateway:1f57ebef29501935023ddbc864512614. DOI: https://doi.org/10.1207/s15327965pli1504_01

Bird, C. E. (1999). Gender, Household Labor, and Psychological Distress: The Impact of the Amount and Division of Housework. *Journal of Health and Social Behavior, 40*(1), 32–45. http://doi.org/10.2307/2676377?ref=search-gateway:61ecd615d0421a d5f71f2ca29d2753f8. DOI: https://doi.org/10.2307/2676377

Blau, F. D., & Ferber, M. A. (1991). Career plans and expectations of young women and men: The earnings gap and labor force participation. *Journal of Human Resources.* DOI: https://doi.org/10.2307/145976

Blau, F. D., & Kahn, L. M. (2006). The U.S. Gender Pay Gap in the 1990s: Slowing Convergence. *Industrial and Labor Relations Review, 60*(1), 45–66. http://doi.org/10.2307/25067574?ref=search-gateway:de99a3c 1b49fcd2df31603a08dbf3b1c. DOI: https://doi.org/10.1177/001979390606000103

Bohnet, I., Van Geen, A., & Bazerman, M. (2016). When Performance Trumps Gender Bias: Joint vs. Separate Evaluation. *Management Science, 62*(5), 1225–1234. DOI: https://doi.org/10.1287/mnsc.2015.2186

Bridges, J. S. (1993). Pink Or Blue: Gender-Stereotypic Perceptions of Infants as Conveyed by Birth Congratulations Cards. *Psychology of Women Quarterly.* DOI: https://doi.org/10.1111/j.1471-6402.1993.tb00444.x

Charles, M., & Grusky, D. B. (2004). Occupational ghettos: The worldwide segregation of women and men (Vol. 200). Stanford, CA: Stanford University Press.

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. DOI: https://doi.org/10.1037/0022-3514.89.4.469

Cox, J., & Dittmar, H. (1995). The functions of clothes and clothing (dis)satisfaction: A gender analysis among British students. *Journal of Consumer Policy, 18*(2–3), 237–265. DOI: https://doi.org/10.1007/BF01016513

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., …, Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*(1), 1–33. DOI: https://doi.org/10.1348/014466608X314935

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-Gender Stereotypes in Elementary School Children. *Child Development, 82*(3), 766–779. DOI: https://doi.org/10.1111/j.1467-8624.2010.01529.x

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking Congruity Between Goals and Roles: A New Look at Why Women Opt Out of Science, Technology, Engineering, and Mathematics Careers. *Psychological Science, 21*(8), 1051–1057. DOI: https://doi.org/10.1177/0956797610377342

Eccles, J. (2011a). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development, 35*(3), 195–201. DOI: https://doi.org/10.1177/0165025411398185

Eccles, J. S. (2011b). Understanding Educational and Occupational Choices. *Journal of Social Issues*, *67*(3), 644–648. DOI: https://doi.org/10.1111/j.1540-4560.2011.01718.x

Francis, B. (2010). Gender, toys and learning. *Oxford Review of Education*, *36*(3), 325–344. DOI: https://doi.org/10.1080/03054981003732278

Furr, R. M. (2017). *Psychometrics: An introduction.* Sage Publications.

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

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. DOI: https://doi.org/10.1037/0022-3514.74.6.1464

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. DOI: https://doi.org/10.1037/0022-3514.85.2.197

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The Times They Are a-Changing … or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*(3), 353–363. DOI: https://doi.org/10.1177/0361684316634081

Hammond, M. D., & Cimpian, A. (2017). Investigating the cognitive structure of stereotypes: Generic beliefs about groups predict social judgments better than statistical beliefs. *Journal of Experimental Psychology: General*, *146*(5), 607. DOI: https://doi.org/10.1037/xge0000297

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. DOI: https://doi.org/10.1177/0146167205275613

Howes, C., & Phillipsen, L. (1992). Gender and friendship: Relationships within peer groups of young children. *Social Development*. DOI: https://doi.org/10.1111/j.1467-9507.1992.tb00126.x

Hutchinson, L. D. (1982). Anna J. Cooper, a voice from the South. Smithsonian Institution Press.

Jost, J. T. (2006). The end of the end of ideology. *American Psychologist*, *61*(7), 651–670. DOI: https://doi.org/10.1037/0003-066X.61.7.651

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. DOI: https://doi.org/10.1037/0022-3514.88.3.498

Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. *Handbook of prejudice, stereotyping, and discrimination*, *199*, 227.

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. DOI: https://doi.org/10.1016/j.jesp.2006.08.004

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., …, Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*. DOI: https://doi.org/10.1037/amp0000364

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. DOI: https://doi.org/10.1177/1368430215603462

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. DOI: https://doi.org/10.1126/science.1261375

Lloyd, G. (1979). THE MAN OF REASON. *Metaphilosophy*, *10*(1), 18–37. DOI: https://doi.org/10.1111/j.1467-9973.1979.tb00062.x

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. DOI: https://doi.org/10.3758/s13428-014-0532-5

Nosek, B. A. (2005). Moderators of the Relationship Between Implicit and Explicit Evaluation. *Journal of Experimental Psychology: General*, *134*(4), 565–584. DOI: https://doi.org/10.1037/0096-3445.134.4.565

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*(1), 44–59. DOI: https://doi.org/10.1037//0022-3514.83.1.44

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

Nussbaum, M. C. (1994, October 20). Feminists and philosophy. *The New York Review of Books*.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. DOI: https://doi.org/10.1037/a0032734

Padovani, F., Richardson, A., & Tsou, J. Y. (Eds.) (2010). Objectivity in Science: New Perspectives from Science and Technology. Springer.

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. DOI: https://doi.org/10.1037/0022-3514.89.3.277

Prentice, D. A., & Carranza, E. (2002). What Women and Men Should Be, Shouldn't be, are Allowed to be, and don't Have to Be: The Contents of Prescriptive Gender Stereotypes. *Psychology of Women Quarterly*, *26*(4), 269–281. DOI: https://doi.org/10.1111/1471-6402.t01-1-00066

Rubin, J. Z., Provenzano, F. J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry*, *44*(4), 512–519. DOI: https://doi.org/10.1111/j.1939-0025.1974.tb00905.x

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*. DOI: https://doi.org/10.1111/0022-4537.00239

Rudman, L. A., Greenwald, M. G., & Mcghee, D. E. (2000). Implicit Self-Concept and Evaluative Implicit Gender Stereotypes: Self and Ingroup Share Desirable Traits. *Personality and Social Psychology Bulletin*, *27*(9), 1164–1178. DOI: https://doi.org/10.1177/0146167201279009

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and Explicit Attitudes Toward Female Authority. *Personality and Social Psychology Bulletin*, *26*(11), 1315–1328. DOI: https://doi.org/10.1177/0146167200263001

Rudman, L. A., & Phelan, J. E. (2010). The Effect of Priming Gender Roles on Women's Implicit Gender Beliefs and Career Aspirations. *Social Psychology*, *41*(3), 192–202. DOI: https://doi.org/10.1027/1864-9335/a000027

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

Storage, D., Horne, Z., Cimpian, A., & Leslie, S.-J. (2016). The Frequency of "Brilliant" and 'Genius' in Teaching Evaluations Predicts the Representation of Women and African Americans across Fields. *PloS One*, *11*(3), e0150194. DOI: https://doi.org/10.1371/journal.pone.0150194

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. DOI: https://doi.org/10.3758/s13428-012-0314-x

Young, D. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The Influence of Female Role Models on Women's Implicit Science Cognitions. *Psychology of Women Quarterly*, *37*(3), 283–292. DOI: https://doi.org/10.1177/0361684313482109