COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 42 (2018) 2229–2249

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12637

"Girls Are as Good as Boys at Math" Implies That Boys Are Probably Better: A Study of Expressions of Gender Equality

Eleanor K. Chestnut, Ellen M. Markman

Department of Psychology, Stanford University

Received 22 May 2018; received in revised form 7 May 2018; accepted 25 May 2018

Abstract

Although "Girls are as good as boys at math" explicitly expresses equality, we predict it could nevertheless suggest that boys have more raw talent. In statements with this subject-complement structure, the item in the complement position serves as the reference point and is thus considered more typical and prominent. This explains why "Tents are like houses," for instance, sounds better than "Houses are like tents"—people generally think of houses as more typical. For domains about ability, the reference point should be the item that is typically more skilled. We further propose that the reference point should be *naturally* more skilled. In two experiments, we presented adults with summaries of actual scientific evidence for gender equality in math (Experiment 1) or verbal ability (Experiment 2), but we manipulated whether the reference point in the statements of equality in the summaries (e.g., "Boys' verbal ability is as good as girls'") was girls or boys. As predicted, adults attributed more natural ability to each gender when it was in the complement rather than subject position. Yet, in Experiment 3, we found that when explicitly asked, participants judged that such sentences were *not* biased in favor of either gender, indicating that subject-complement statements must be transmitting this bias in a subtle way. Thus, statements such as "Girls are as good as boys at math" can actually backfire and perpetuate gender stereotypes about natural ability.

Keywords: Language; Gender equality; Inference; Ability; Syntax

1. Introduction

To increase the representation of women in science, technology, engineering, and math (STEM), a common, natural strategy is to affirm girls' abilities by saying, for example,

Correspondence should be sent to Eleanor K. Chestnut, Stanford University, Department of Psychology, Jordan Hall, 450 Serra Mall (Bldg. 420), Stanford, CA 94305. E-mail: eleanor.chestnut@gmail.com

"Girls are as good as boys at math." In three experiments, we question whether such statements of equality can be effective in reducing gender stereotypes.

The problem is that sentences with this subject-complement structure imply *differences* between the items in the subject and complement positions, even when the statement explicitly expresses similarity. One explanation for this is that the subject position is for *variants*, while the complement position is for *reference points* (Bowdle & Medin, 2001; Bruckmüller & Abele, 2010; Bruckmuller, Hegarty, & Abele, 2012; Chestnut & Markman, 2016; Gleitman, Gleitman, Miller, & Ostrin, 1996; Rosch, 1975; Shen, 1989; cf. Tversky, 1977, whose seminal work helped bring this asymmetry to the forefront, and Bowdle & Gentner, 1997, who provide an interesting alternative account for this phenomenon). Reference points are generally the more *prominent* items—they are more typical or more important, serving as "anchors" for variants (Rosch, 1975). Consider, for example, the statements "Zebras are like horses" and "Horses are like zebras." Both express similarity between horses and zebras. However, "Zebras are like horses" sounds better because when discussing similarity, we usually think of horses, the more prototypical animal, as an appropriate reference point for zebras rather than vice versa. Importantly, listeners can thus *infer* differences between items based on their syntactic positions. Upon hearing, "Zum is like Gax," for instance, listeners infer that *Gax* is more typical and important than *Zum* (Bruckmüller & Abele, 2010; Chestnut & Markman, 2016; Gleitman et al., 1996).

The specific features that make the reference point more prominent vary according to the predicate. Because the predicate specifies the dimension along which the two items are related, what it means to be prominent must be relevant to that dimension. For symmetrical spatial relations, we prefer, for example, "The bicycle is near the building" over "The building is near the bicycle" because in this context, reference points are defined by physical prominence and stability. Similarly, we prefer "The child met the President" over "The President met the child" because the features defining prominence in encounters include greater fame and noteworthiness.

What might we infer, then, when we hear "Girls are as good as boys at math"? In domains about ability the reference point is likely the one who is typically considered to be highly skilled. Although someone can become skilled by either working hard, having raw talent, or both, we predict that adults will be biased to associate raw talent rather than effort with the reference point. We suggest this because of people's tendency—at least in Western cultures—to attribute others' behaviors to internal features rather than to external circumstances (e.g., Gawronski, 2004; Gilbert & Malone, 1995; Jones, 1979; Ross, Amabile, & Steinmetz, 1977). If the reference point is typically considered skilled in the domain, then the easiest, most cognitively accessible explanation for their status could be that they were *made* to do that activity (Cimpian & Salomon, 2014).

If adults do associate higher levels of raw talent with the reference point, then they might also reason that the variant, by comparison, has to work harder to be equally skilled. Although several researchers have shown that adults (and children) vary in their perception of raw talent and effort as inversely related (Heyman & Compton, 2006; Miele, Son, & Metcalfe, 2013; Muenks & Miele, 2017), in contexts of *social comparison*, adults will often use this inverse relation to reason about relative ability (Jagacinski & Nicholls, 1987; Karabenick & Heller, 1976; Nicholls, Patashnick, & Mettetal, 1986). If two students perform equally on an exam but one clearly exerted more effort than the other, for example, most adults—and perhaps even children—will infer that the student who worked harder has less raw talent. Given that statements such as "Girls are as good as boys at math" involve social comparison between the variant and reference point, adults might reason that the variant must compensate for its lack of raw talent with increased effort. Despite explicitly expressing equality, then, subject-complement statements could *implicitly* communicate important differences in status and ability.

The subtle implications of subject-complement statements are particularly important to investigate because of how powerfully implicit information can influence thought and behavior outside conscious awareness (e.g., Banaji & Hardin, 1996; Banaji, Hardin, & Rothman, 1993; Bargh, Chen, & Burrows, 1996; Eberhardt, Goff, Purdie, & Davies, 2004; Payne, 2001). Participants who are subliminally primed with black faces, for example, are faster to recognize crime-relevant objects (e.g., guns) than those who are primed with white faces or no faces, because of their association of blackness with crime (Eberhardt et al., 2004; Payne, 2001). Similarly, participants who are primed with words associated with gender stereotypes (e.g., *nurse*, *doctor*) are faster to recognize gendered pronouns that are consistent rather than inconsistent with the gender of the stereotype (Banaji & Hardin, 1996). Thus, the implicit messages communicated by "Girls are as good as boys at math" could be equally influential, implying that boys are naturally more talented.

To test whether subject-complement statements about gender equality subtly communicate gender *differences* in raw talent, we presented participants with paragraphs summarizing scientific evidence that there are no gender differences in either math (Experiment 1) or verbal ability (Experiment 2). Critically, we manipulated how this similarity was framed across conditions. While some participants read statements with a subject-complement structure (e.g., "Boys do as well as girls at math"), others read statements with both genders in the subject position (e.g., "Boys and girls do equally well at math"). Participants then identified either which gender is naturally more skilled in the domain, or which gender has to work harder to have the ability. Finally, in Experiment 3, we confirmed the implicit nature of this effect by directly asking participants whether they believe that subject-complement statements of equality are at all biased against the group framed as the variant.

2. Experiment 1

In Experiment 1, we focused on math ability, a domain in which boys are stereotyped as more talented (e.g., Tiedemann, 2000). We predicted that statements of equality with boys in the complement position (e.g., "Girls do as well as boys at math") would not substantially change participants' belief that boys are naturally better than girls at math. In contrast, statements with *girls* in the complement position, which frame girls as the standard, might more successfully reduce this stereotype. We also examined statements with both genders in the subject position (e.g., "Girls and boys are equally good at math"). Here, since neither gender is framed as the reference point, the order of the genders in the subject position should not influence participants' beliefs. Instead, these statements might reduce participants' biases, regardless of the word order.

2.1. Method

2.1.1. Participants

Participants were 640 English-speaking adults from the United States ages 18 to 69 (M = 33, 354 men) who participated through Amazon Mechanical Turk for a payment of $0.25. A total of 128 participants participated in each of the five conditions (*Girls EQUAL Boys*, *Boys EQUAL Girls*, *Girls AND Boys*, *Boys AND Girls*, and *Baseline*). In a previous study using multiple trials and non-gender categories (e.g., zebras and horses), we had sufficient power to detect adults' sensitivity to the framing of similarity with only 48 participants (Chestnut & Markman, 2016). We significantly increased the number of participants in this study because we expected a smaller effect size, given that participants' preexisting beliefs about gender might influence their responses, and because the current task consisted of only one trial.

2.1.2. Materials and procedure

In all conditions, participants completed only one trial.

In the *Baseline* condition, participants were simply asked to judge either which gender (girls or boys) is *naturally* more skilled at math, or which gender has to *work harder* to be good at math.

Participants in the *Girls EQUAL Boys* condition read a brief summary of an actual scientific article that reported the results of Hyde, Lindberg, Linn, Ellis, and Williams (2008), a meta-analysis showing a lack of gender differences in math ability (Girls = Boys at Math, 2008). The summary contained three statements of gender equality, each of which placed "boys" in the complement position (e.g., "Girls do as well as boys at math"). After reading the paragraph, participants were asked to judge either which gender is naturally more skilled at math ("Based on these findings, who do you think is naturally more skilled at math?") or which gender has to work harder to be good at math ("Based on these findings, who do you think has to work harder to be good at math?").

> ***Girls EQUAL Boys* condition** [Underlined here are the three statements of equality.]
>
> <u>Girls = Boys at Math</u>
>
> A recent study has shown that <u>girls do just as well as boys at math</u>. At the University of Wisconsin, a team of researchers analyzed scores from standardized tests taken in 2005, 2006, and 2007 by approximately seven million students in ten different states. Overall, they found that <u>girls performed as well as boys</u> in grades two through eleven.

A troubling finding from the study, however, is that many tough math questions seem to have been removed from state tests. The researchers worry that teachers, as a result, may start dropping harder math problems from their curriculums.[1]

The *Boys EQUAL Girls* condition was identical to the *Girls EQUAL Boys* condition, except the statements of equality always placed "girls" in the complement position (e.g., "Boys do as well as girls at math").

For comparison, we also included two conditions that expressed equality in a slightly different way. These conditions were identical to the *Girls EQUAL Boys* and *Boys EQUAL Girls* conditions, except the three statements of equality in the paragraphs placed both genders in the subject position. In the *Girls AND Boys* condition, the statements of equality always began with "Girls and boys..." (e.g., "Girls and boys do equally well at math"). In the *Boys AND Girls* condition, the statements of equality always began with "Boys and girls..." (e.g., "Boys and girls do equally well at math"). Since neither gender was framed as the reference point for the other, we expected participants to respond similarly regardless of whether they saw "Girls and boys..." or "Boys and girls...."

***Girls AND Boys* condition** [Underlined again are the three statements of equality.]

Girls and Boys are Equally Good at Math

A recent study has shown that girls and boys do equally well at math. At the University of Wisconsin, a team of researchers analyzed scores from standardized tests taken in 2005, 2006, and 2007 by approximately seven million students in ten different states. Overall, they found that girls and boys performed as well as each other in grades two through eleven. A troubling finding from the study, however, is that many tough math questions seem to have been removed from state tests. The researchers worry that teachers, as a result, may start dropping harder math problems from their curriculums.

No statements were actually underlined for the participants.

In all conditions, after answering the question about either natural ability or effort, participants rated their confidence in their choice using a sliding scale from "Not at all confident" (0) to "Very confident" (100).

2.1.3. Statistical analysis

We conducted statistical analyses in RStudio (version 0.98.1091). Our confirmatory analyses in Experiments 1 and 2 tested our hypothesis that the paragraphs participants read would differentially influence their responses. We analyzed the forced-choice responses using logistic regression models with condition (*Girls EQUAL Boys*, *Boys EQUAL Girls*, *Girls AND Boys*, *Boys AND Girls*, or *Baseline*) as a categorical predictor. We tested two sets of planned contrasts: Our first set compared each experimental condition against the *Baseline* condition. Our second set compared the *Girls EQUAL Boys* condition to the *Boys EQUAL Girls* condition, and the *Girls AND Boys* condition to the *Boys AND Girls* condition. This served as a more direct test of whether the order of "girls" and "boys" in the statements of equality influenced responses. For the Wald tests in the models, we report the odds ratios (OR), their 95% confidence intervals, and their corresponding *p*-values. The odds ratios indicate the relative likelihood of selecting "boys" in each condition. An odds ratio of 3, for example, would mean that participants in the condition of interest were three times as likely to select "boys" than those in the reference condition. An odds ratio of 0.5 would mean that the participants in the condition of interest were half as likely to select "boys."

We also calculated *weighted* responses, by multiplying participants' forced-choice responses (coded as −1 for "girls" and 1 for "boys") by their confidence ratings. The full range of potential weighted responses was thus −100 to 100, with positive values reflecting a preference for "boys" and negative values reflecting a preference for "girls." We conducted the same set of confirmatory analyses on the weighted responses as on the forced-choice responses, using linear regression rather than logistic regression models. For the *t* tests in the models, we report the unstandardized beta coefficients, their 95% confidence intervals, and their corresponding *p*-values.

For our exploratory analyses in each experiment, we added two other categorical predictors to our models: the gender of the participant (female, coded as 0; or male, coded as 1) and the question they answered (about effort, coded as 0; or natural ability, coded as 1). To identify the models that best fit the data, we used information-criterion statistics (Akaike information criterion; Burnham & Anderson, 2002). We did not have specific hypotheses for these effects; rather, the purpose of the exploratory analyses was to determine whether these variables might also influence responses. We provide all exploratory analyses in the Supplementary materials, though we summarize the main findings in the discussion of each experiment.

In each experiment, we reverse-coded the responses to the question about effort. Thus, responses to this question reflected the gender participants believed had to work *less* hard to be skilled. Reverse-coding these responses allowed us to combine the responses to each question to create an overall measure of bias.

2.2. Results

Our confirmatory analyses tested the effects of each experimental condition (*Girls EQUAL Boys*, *Boys EQUAL Girls*, *Girls AND Boys*, and *Boys AND Girls*) on responses.

Forced-choice responses. First, in the *Baseline* condition, when asked, "Who has to work harder to be good at math?" or, "Who is naturally more skilled at math?" 67% of participants ($SE = 4\%$, $n = 128$) attributed more natural math ability or less effort to boys, which was greater than chance, binomial sign test, $p < .001$. This reflects the common stereotype about gender differences in math ability.

To test our hypothesis that participants' inferences would vary according to the paragraph they read, we next compared participants in each experimental condition to those in the *Baseline* condition. As predicted, participants in the *Girls EQUAL Boys* condition

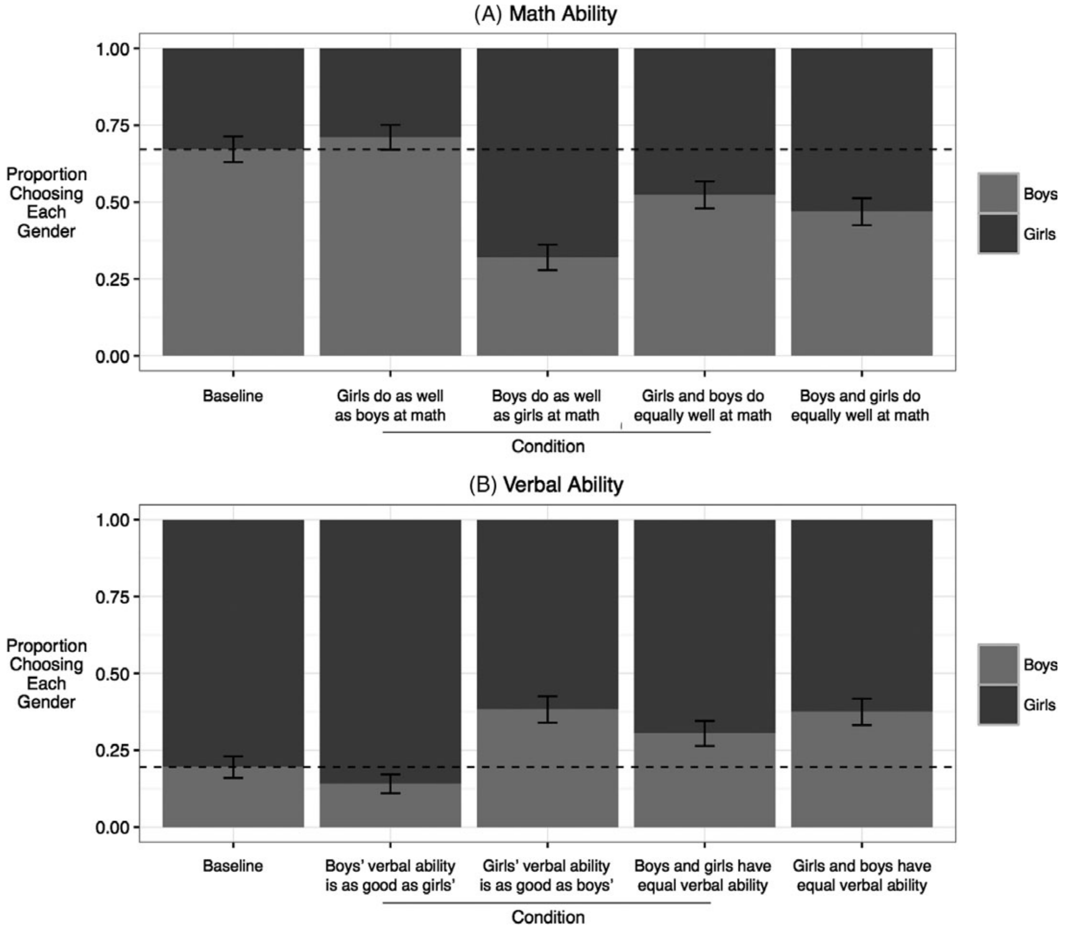

Fig. 1. Participants' forced-choice responses in each condition for (a) math ability in Experiment 1 and (b) verbal ability in Experiment 2. Responses reflect the gender to whom participants attributed more natural ability or less effort. Error bars represent ±1 *SE*.

(who read, e.g., "Girls do as well as boys at math") showed no reduction whatsoever in their gender bias relative to baseline, OR = 1.20 [0.71–2.05], $p > .250$ (see Fig. 1). Instead, 71% (SE = 4%, n = 128; compared to 67% at baseline) stated that boys have more natural math ability or have to work less hard to be good at math.

Strikingly, participants in the *Boys EQUAL Girls* condition (who read, e.g., "Boys do as well as girls at math") showed the opposite bias in their responses. Now, only 32% (SE = 4%, n = 128) attributed greater natural math ability or less effort to boys, OR = 0.23 [0.14–0.39], $p < .001$. This demonstrates the power of the subject-complement sentence structure to influence attributions of natural ability and effort.

Participants in the *Girls AND Boys* and *Boys AND Girls* conditions also showed a reduction in their stereotypic beliefs about math ability. Only 52% (SE = 4%, n = 128) of those in the *Girls AND Boys* condition (who read, e.g., “Girls and boys do equally well at math”) and 53% (SE = 4%, n = 128) of those in the *Boys AND Girls* condition attributed greater natural math ability or less effort to boys, OR = 0.54 [0.32–0.89], p = .016 and OR = 0.55 [0.33–0.92], p = .022, respectively. Egalitarian beliefs about math ability should generate such chance responding.

Our second set of planned contrasts compared the *Girls EQUAL Boys* condition to the *Boys EQUAL Girls* condition and the *Girls AND Boys* condition to the *Boys AND Girls* condition, to directly test whether the order of “girls” and “boys” in the statements of equality influenced participants’ responses. Participants in the *Girls AND Boys* and *Boys AND Girls* conditions responded similarly, OR = 0.98 [0.77–1.26], $p > .250$. Thus, the order of “girls” and “boys” in the subject position did not bias participants to associate more natural ability or less effort with one particular gender. Participants in the *Girls EQUAL Boys* and *Boys EQUAL Girls* conditions, however, showed dramatically different attributions of natural ability and effort, OR = 2.28 [1.76–3.00], $p < .001$. Again, when boys were in the complement position (e.g., “Girls do as well as boys at math”; *Girls EQUAL Boys* condition), participants attributed greater natural math ability or less effort to boys. When girls were in the complement position (e.g., “Boys do as well as girls at math”; *Boys EQUAL Girls* condition), they attributed greater natural math ability or less effort to *girls*.

Weighted responses. Weighted responses were calculated by multiplying participants’ forced-choice responses (coded as −1 for “girls” and 1 for “boys”) by their respective confidence ratings (ranging from 0 to 100, with higher values reflecting greater confidence). Thus, significantly positive averages across participants reflected a preference for “boys,” significantly negative averages reflected a preference for “girls,” and responses close to 0 reflected a preference for neither gender. Overall, we found the same effects for both the forced-choice responses and the weighted responses.

First, at baseline, when asked, “Who is naturally better at math?” or “Who has to work harder to be good at math?” participants associated greater natural math ability or less effort with boys (M = 13.55, SE = 4.90, n = 128), $t(127) = 2.76$, p = .007.

When we compared the weighted responses in each experimental condition to those in the *Baseline* condition, we again found that participants in the *Girls EQUAL Boys* condition (M = 16.18, SE = 4.36, n = 128) were no different from those in the *Baseline* condition, associating greater natural math ability or less effort with boys, b = 2.63 [−9.52–14.78], $p > .250$. Those in the *Boys EQUAL Girls* condition (M = −20.72, SE = 4.67, n = 128), on the other hand, displayed the opposite bias and instead associated greater natural math ability or less effort with *girls*, b = −34.27 −46.43 to −22.12], $p < .001$.

Participants in the *Girls AND Boys* (M = 3.18, SE = 3.84, n = 128) and *Boys AND Girls* conditions (M = 1.86, SE = 4.01, n = 128) were marginally different from those in the *Baseline* condition, b = −10.38 [−22.53–1.78], p = .094 and b = −11.70 [−23.85–0.46], p = .059, respectively. As with the forced-choice responses, the average weighted responses in these two conditions were close to 0, reflecting egalitarian beliefs about math ability.

Our second set of planned contrasts showed that participants in the *Girls EQUAL Boys* and *Boys EQUAL Girls* conditions responded differently, b = 18.45 [12.37–24.53], $p < .001$. Again, when boys were framed as the reference point (e.g., “Girls do as well as boys at math"), participants attributed greater natural math ability or less effort to boys. When girls were framed as the reference point (e.g., "Boys do as well as girls at math"), they attributed greater natural math ability or less effort to girls. Participants in the *Girls AND Boys* and *Boys AND Girls* conditions, however, responded similarly, $b = 0.66$ [−5.42–6.73], $p > .250$.

In addition, in exploratory analyses, which assessed the effects of the participant's gender (female or male) and the question they answered (about natural ability or effort) on responses, we found that overall, participants were biased to attribute more natural ability, but not less effort, to their own gender. We return to this interaction in the General Discussion.

2.3. Discussion

As predicted, participants who read "Girls do as well as boys at math" held onto their stereotypic belief that boys have more natural math talent and have to work less hard to be good at math. Because this statement framed boys as the reference point, it subtly implied that boys set the standard for math ability.

In marked contrast, participants who read statements framing *girls* as the reference point (e.g., "Boys do as well as girls at math") were much less likely to attribute greater natural math ability or less effort to boys. In fact, in this condition, participants actually showed the *opposite* bias and attributed greater natural ability or less effort to *girls*. This is a testament to the power of subject-complement syntactic structure to convey information about the reference point: Even though "Girls do as well as boys at math" and "Boys do as well as girls at math" both express gender equality, they caused participants to associate greater natural ability with different genders.

It is surprising, though, those participants who read "Boys do as well as girls at math" showed a completely reversed bias rather than a bias that was simply reduced. One possible explanation for this is that participants were only moderately committed to their beliefs about math ability in the first place. Recall that in the *Baseline* condition, participants' average weighted response was 13.55. This means that participants did not have strong confidence in their choice—had they been very confident that boys were naturally more talented and had to work less hard, weighted responses would have been closer to 100. Thus, participants in this experiment might have been relatively open to updating their beliefs based on the framing of the paragraph.

Finally, statements with both genders in the subject position (e.g., "Girls and boys do equally well at math") showed the most promise as expressions of equality. Since neither gender was framed as the reference point nor these statements reduced participants' biases without reversing them.

An interesting question is to what extent participants actually changed their beliefs after reading the paragraph, and to what extent they were responding based on what they thought the author of the paragraph believed. Although our experiments cannot speak to this question, recall that participants in each condition were explicitly asked what *they* believed (e.g., "Who do you think is naturally more skilled at math?"). Still, participants might have attributed certain beliefs to the author based on the framing in each condition. When participants read, “Boys do as well as girls at math,” for instance, they might have reasoned that the author must think that girls set the standard. Participants’ responses could have thus reflected a willingness to say what the author believed, rather than what they actually believed themselves. If this was indeed participants’ thought process, then it still speaks to the power of the subject-complement constructions. The explicit content of the paragraph—the evidence for a lack of gender differences in math ability—was also representative of the author’s beliefs, but clearly it was not enough to encourage participants to give egalitarian responses.

In sum, our results show that framing one gender as the reference point for the other is a powerful way of communicating gender differences in raw talent, *even when* the statement explicitly expresses equivalent abilities.

3. Experiment 2

In Experiment 2, we sought to replicate these effects using a domain in which girls are stereotyped as naturally more talented: verbal ability (e.g., Jackson, Hodge, & Ingram, 1994). In this case, placing girls in the complement position (e.g., “Boys’ verbal ability is as good as girls’”) should not substantially change participants’ preexisting beliefs, whereas placing *boys* in the complement position should decrease participants’ tendency to attribute greater natural talent or less effort to girls. Finally, placing both genders in the subject position (e.g., “Boys and girls have equal verbal ability”) should also reduce participants’ biases.

3.1. Method

3.1.1. Participants

Participants were 640 English-speaking adults from the United States ages 18 to 69 (M = 33,368 men) who participated through Amazon Mechanical Turk for a payment of $0.25. A total of 128 participants participated in each of the five conditions (*Girls EQUAL Boys*, *Boys EQUAL Girls*, *Girls AND Boys*, *Boys AND Girls*, and *Baseline*).

3.1.2. Materials and procedure

The paradigm was identical to that used in Experiment 1, except the paragraphs were summaries of Hyde and Linn (1998), a meta-analysis showing a lack of gender differences in verbal abilities.

Boys EQUAL Girls **condition** [Underlined here are the three statements of equality]

Boys = Girls in Verbal Ability

A recent study has indicated that boys’ verbal ability is just as good as girls’. At the University of Wisconsin, a team of researchers reviewed 165 studies on verbal ability, which, all together, reported data collected from over one million participants. They found that boys performed as well as girls on measures of verbal ability, including vocabulary, reading comprehension, speech production, and essay writing.

***Boys AND Girls* condition** [Underlined again are the three statements of equality.]

Boys and Girls Have Equal Verbal Ability

A recent study has indicated that boys' and girls' verbal abilities are equally good. At the University of Wisconsin, a team of researchers reviewed 165 studies on verbal ability, which, all together, reported data collected from over one million participants. They found that boys and girls performed as well as each other on measures of verbal ability, including vocabulary, reading comprehension, speech production, and essay writing.

No statements were actually underlined for the participants.

3.3. Results

Our confirmatory analyses, outlined in Experiment 1, tested the effects of each experimental condition (*Girls EQUAL Boys*, *Boys EQUAL Girls*, *Girls AND Boys*, and *Boys AND Girls*) on responses.

Forced-choice responses. First, in the *Baseline* condition, when asked, "Who has to work harder to be verbally skilled?" or "Who is naturally more verbally skilled?" 80% of participants (SE = 4%, n = 128) attributed greater natural verbal ability or less effort to girls, which was greater than chance, binomial sign test, $p < .001$. This was a stronger bias than that found in the *Baseline* condition for math ability in Experiment 1, where 67% of participants endorsed the traditional gender stereotype.

Our first set of planned contrasts compared participants in each experimental condition to those in the *Baseline* condition. As predicted, participants in the *Boys EQUAL Girls* condition (who read, e.g., "Boys' verbal ability is as good as girls'") were no different from those in the *Baseline* condition, OR = 0.67 [0.34–1.30], p = .244 (see Fig. 1). Eighty-six percent of these participants (SE = 3%, n = 128) attributed greater natural verbal ability or less effort to girls, compared to 80% at baseline. Participants in the *Girls EQUAL Boys* condition (who read, e.g., "Girls' verbal ability is as good as boys'"), on the other hand, displayed a significantly weaker gender bias. Here, only 62% of participants (SE = 4%, n = 128) attributed greater natural verbal ability or less effort to girls, OR = 2.56 [1.47–4.54], p = .001. We likely did not find a complete reversal of biases in this condition as we did for math ability in Experiment 1 because the biases at baseline, described above, were particularly strong.

As in Experiment 1, participants in the *Girls AND Boys* and *Boys AND Girls* conditions also showed weaker stereotypic beliefs about verbal ability. Sixty-three percent of participants in the *Girls AND Boys* condition (who read, e.g., "Girls and boys have equal verbal ability"; SE = 4%, n = 128) and 70% of those in the *Boys AND Girls* condition

(SE = 4%, n = 128) attributed greater natural verbal ability or less effort to girls, OR = 2.47 [1.42–4.40], p = .002 and OR = 1.81 [1.02–3.24], p = .045, respectively.

Our second set of planned contrasts compared the *Girls EQUAL Boys* condition to the *Boys EQUAL Girls* condition, and the *Girls AND Boys* condition to the *Boys AND Girls* condition, to directly test whether the order of "girls" and "boys" in the statements of equality influenced responses. As described above, participants in the *Girls EQUAL Boys* and *Boys EQUAL Girls* conditions had strikingly different responses, OR = 1.95 [1.44–2.67], p < .001. They were more likely to associate greater natural verbal ability or less effort with girls when girls rather than boys were framed as the reference point. Participants in the *Girls AND Boys* and *Boys AND Girls* conditions, in contrast, responded similarly, OR = 1.17 [0.90–1.52], p = .236.

Weighted responses. Weighted responses were calculated by multiplying participants' forced-choice responses (coded as −1 for "girls" and 1 for "boys") by their respective confidence ratings (ranging from 0 to 100, with higher values reflecting greater confidence). Thus, significantly positive averages across participants reflected a preference for "boys", significantly negative averages reflected a preference for "girls," and responses close to 0 reflected a preference for neither gender. Overall, the effects were the same for the forced-choice and weighted responses.

First, in the *Baseline* condition, participants showed a strong bias to associate greater natural verbal ability or less effort with girls (M = −38.61, SE = 4.55, n = 128), $t(127) = -8.48$, p < .001.

In our first set of planned contrasts, we compared each experimental condition to the *Baseline* condition. As predicted, participants in the *Boys EQUAL Girls* condition (who read, e.g., "Boys' verbal ability is as good as girls'"; M = −37.10, SE = 4.04, n = 128) were no different from those in the *Baseline* condition, b = 1.51 [−10.94–13.95], p > .250. Instead, these participants continued to associate greater natural verbal ability or less effort with girls. Participants in the *Girls EQUAL Boys* condition, however, displayed much weaker biases than those in the *Baseline* condition (M = −9.45, SE = 5.00, n = 128), b = 29.27 [16.83–41.72], p < .001.

Participants in the *Girls AND Boys* (M = −12.06, SE = 4.21, n = 128) and *Boys AND Girls* (M = −16.84, SE = 4.54, n = 128) conditions also showed weaker biases relative to baseline, b = 26.55 [14.10–38.99], p < .001 and b = 21.77 [9.32–34.21], p < .001, respectively.

Our second set of planned contrasts compared the *Girls EQUAL Boys* condition to the *Boys EQUAL Girls* condition, and the *Girls AND Boys* condition to the *Boys AND Girls* condition. As with the forced-choice responses, participants in the *Girls EQUAL Boys* condition responded differently from those in the *Boys EQUAL Girls* condition, b = 13.88 [7.66–20.11], p < .001. Participants in the *Girls AND Boys* and *Boys AND Girls* conditions, however, showed similar levels of reduced bias, b = 2.39 [−3.83–8.61], p > .250.

3.4. Discussion

The main findings for verbal ability paralleled those for math ability. When participants read paragraphs about gender equality that framed girls as the reference point (e.g., "Boys' verbal ability is as good as girls'"), they continued to attribute greater natural verbal ability and less effort to girls.

When the paragraphs framed *boys* as the reference point (e.g., "Girls' verbal ability is as good as boys'"), in contrast, participants were much less likely to display this bias, though they still favored girls overall. We likely did not see a complete reversal of biases in this condition as we did for math ability in Experiment 1 because participants' baseline beliefs about verbal ability were substantially stronger than their baseline beliefs about math ability. The average weighted response in the *Baseline* condition for verbal ability was −38.61 (in favor of girls), compared to 13.55 for math ability in Experiment 1 (in favor of boys). Thus, participants might have been less willing to change their beliefs about verbal ability across conditions.

Lastly, when participants saw both genders in the subject position (e.g., "Girls and boys have equal verbal ability"), they also showed weaker biases. This further establishes subject-subject statements of equality as the best candidates for egalitarian expressions.

As with math ability in Experiment 1, exploratory analyses revealed overall differences between women and men for the question about natural ability. Although both women and men attributed greater natural verbal ability to girls across all conditions, women's bias was much stronger. For the question about effort, in contrast, women and men attributed less effort to girls at similar rates. Again, we return to this in the General Discussion.

Overall, these results show that the way we frame statements about equal abilities matters, regardless of whether girls or boys are typically associated with the ability. Thus, the syntax drives these effects, rather than something specific to the gender categories themselves.

4. Experiment 3

In Experiments 1 and 2, we showed that subject-complement statements about equal abilities (e.g., "Girls do as well as boys at math") imply differences in natural talent and effort, despite explicitly expressing similarity. Since these statements *appear* to be egalitarian, they might thus be a particularly insidious way of perpetuating gender stereotypes. To directly assess whether people are aware of these implicit messages—or whether the messages instead bias people in more subtle ways—in Experiment 3 we asked participants directly whether such statements are biased against the gender framed as the variant.

4.1. Method

4.1.1. Participants

Participants were 384 English-speaking adults from the United States ages 19 to 81 (M = 35, 153 men) who participated through Amazon Mechanical Turk for a payment of \$0.25. A total of 128 participants participated in each of the three conditions (*Subject-complement*, *Subject-subject*, or *Inequality*).

4.1.2. Materials and procedure

All participants read a single sentence about either gender equality or gender *in*equality in math. In the *Subject-complement* condition, participants read either "Girls do just as well as boys at math" or "Boys do just as well as girls at math." In the *Subject-subject* condition, participants read either "Girls and boys do equally well at math" or "Boys and girls do equally well at math." Finally, in the *Explicit Inequality* condition, participants read, "Girls do not do as well as boys at math" or "Boys do not do as well as girls at math." After reading the sentence, participants were asked to judge whether the sentence was biased against the first gender mentioned (e.g., "Do you think that this sentence is biased against girls?") using a sliding scale of "Definitely not" (−100) to "Definitely yes" (100).

Our main question was whether people would explicitly recognize that subject-complement statements such as "Girls do as well as boys at math" are biased against the gender in the subject position (here, girls). If people construe these sentences as *un*biased, treating girls' and boys' abilities as equal, then responses should be less than 0 (the neutral midpoint of the scale). The *Explicit Inference* and *Subject-subject* conditions were used to help calibrate responses in the *Subject-complement* condition by setting boundary conditions for explicit awareness of bias. Statements in the *Explicit Inequality* condition (e.g., "Girls do not do as well as boys at math") explicitly convey inequality and should thus yield strong judgments of bias (i.e., responses close to 100). In contrast, statements in the *Subject-subject* condition (e.g., "Girls and boys do equally well at math"), which were shown in Experiments 1 and 2 to be the most egalitarian, should yield strong judgments of *no* bias (i.e., responses close to −100). By comparing responses in the *Subject-complement* condition to those in the *Explicit Inequality* and *Subject-subject* conditions, we could thus get a more accurate sense of the level of bias participants associate with subject-complement statements.

4.1.3. Statistical analysis

We conducted statistical analyses in RStudio (version 0.98.1091). Our confirmatory analyses tested our hypothesis that participants would view subject-complement statements about gender equality as relatively unbiased. First, we compared responses in each condition (*Subject-complement*, *Explicit Inequality*, and *Subject-subject*) to the neutral midpoint on the scale (0) using *t*-tests. With these tests, we could determine whether participants in each condition viewed the statements they read as either biased or unbiased against the first gender mentioned in the sentence. Second, we analyzed responses using a linear regression model with condition (*Subject-complement*, coded as the reference condition; *Subject-subject*; or *Inequality*) as a categorical predictor. This allowed us to assess whether participants judged the subject-complement statements to be more or less biased than the statements in the other two strong comparison conditions. For the *t*-tests in this model, we report the unstandardized beta coefficients, their 95% confidence intervals, and their corresponding *p*-values.

In our exploratory analyses, which are provided in the supplementary materials, we added two other categorical predictors to our models: the gender of the participant (female, coded as 0; or male, coded as 1) and the first gender mentioned in the statement (girls, coded as 0; or boys, coded as 1). To identify the model that best fit the data, we used information-criterion statistics (Akaike information criterion; Burnham & Anderson, 2002). We did not have specific hypotheses for these effects; rather, the purpose of these analyses was to explore whether the differences between the conditions might depend on the gender of the participant and the order of the genders in the statements.

4.2. Results

As expected, participants in the *Explicit Inequality* condition (e.g., "Girls do not do as well as boys at math"; $M = 54.88$, $SE = 5.15$, $n = 128$) rated the statements they read as strongly biased against the first gender mentioned in the sentence [$t(127) = 10.65$, $p < .001$], while participants in the *Subject-subject* condition (e.g., "Girls and boys do equally well at math"; $M = -71.30$, $SE = 4.54$, $n = 128$) rated the statements they read as strongly *un*biased [$t(127) = -15.70$, $p < .001$]. In the critical *Subject-complement* condition (e.g., "Girls do as well as boys at math"; $M = -43.64$, $SE = 6.05$, $n = 128$), participants, in fact, judged the statements they read to be unbiased toward the first gender mentioned, $t(127) = -7.21$, $p < .001$.

When we compared responses in each condition, we found that participants rated the statements in the *Subject-complement* condition as *less* biased than those in the *Explicit Inequality* condition [$b = 98.52$ [83.82, 113.22], $p < .001$], but *more* biased than those in the *Subject-subject* condition [$b = -27.66$ [−42.36, −12.97], $p < .001$]. Thus, although participants judged subject-complement statements to be unbiased, they were aware that subject-subject statements are, in fact, more egalitarian.

4.3. Discussion

As expected, participants in the *Explicit Inequality* condition (e.g., "Girls do not do as well as boys at math") stated that the sentences they read *were* biased against the first gender mentioned, while participants in the *Subject-subject* condition (e.g., "Girls and boys do equally well at math") stated that the sentences were *not* biased against the first gender mentioned. Participants in the *Subject-complement* condition (e.g., "Girls do just as well as boys at math") fell between those in the *Explicit Inequality* and *Subject-subject* conditions, suggesting that on some level participants understood the different implications of subject-subject and subject-complement syntactic structure. Nevertheless, participants in the *Subject-complement* condition were more similar to those in the *Subject-subject* condition than those in the *Inequality* condition. Like participants in the *Subject-subject* condition, participants in the *Subject-complement* condition overall did *not* believe that the statements were biased against the first gender mentioned.

It is important to note that asking participants whether the sentence was biased could have motivated them to scrutinize the sentence and perceive more bias than they otherwise would have. However, even with this limitation, participants reported that subject-complement statements of equality were generally *not* biased against the gender in the subject position. Thus, subject-complement statements of equality are nevertheless a subtle source of information about gender differences.

5. General discussion

In our experiments, participants read summaries of a scientific article arguing for a lack of gender differences in either math or verbal ability. Each summary contained several explicit statements about girls' and boys' equal abilities—the only difference among them was how this equality was framed. Yet participants' attributions of natural ability and effort to each gender varied strikingly according to the syntactic framing.

Participants who read "Girls do as well as boys at math" or "Boys' verbal ability is as good as girls'" (Experiments 1 and 2) were unaffected by the summary of the scientific article. Instead of attributing raw talent and effort to each gender equally, these participants continued to endorse traditional gender stereotypes. By framing the gender typically associated with the ability as the reference point, the statements inadvertently perpetuated boys' and girls' privileged status in math and verbal ability, respectively.

When participants read statements of equality that implicitly *contradicted* preexisting stereotypes in their framing (e.g., "Boys do as well as girls at math"; "Girls' verbal ability is as good as boys'"), on the other hand, they showed weaker (for verbal ability) or even reversed biases (for math ability). Again, these statements expressed the same equality between girls and boys, and yet participants made dramatically different attributions of natural ability and effort.

Finally, statements with both genders in the subject position (e.g., "Girls and boys do equally well at math"; "Boys and girls have equal verbal ability") were the most effective in communicating equality. Since these statements framed neither gender as the reference point, there was no reason to infer gender differences in raw talent or effort.

In each of our experiments, we also found that for the question about natural ability, participants were biased in favor of their own gender (see the Supplementary Materials for these results). For example, even though both women and men attributed more natural ability to girls in Experiment 2, women were more likely to make this attribution. This bias, however, did not extend to the question about effort. By favoring their own gender only for the question about natural ability, participants seemed to consider natural ability a more desirable, prestigious trait (see Rudman & Goodwin, 2004, for a review of in-group gender bias). If natural ability is viewed as more prestigious, then it would make sense for the reference point, the higher-status group, to have this trait.

We would like to emphasize that these sorts of subject-complement statements are natural and common in everyday discourse. We say that girls are as good as boys at math, for example, because we *do* tend to think of boys as the higher-status standard for comparison (e.g., Tiedemann, 2000; see Miller, Taylor, & Buck, 1991; and Hegarty & Pratto, 2001, for explanations of framing preferences)—even the magazine *Science*, the source of our stimuli, titled its 2008 article on gender similarities "Girls = Boys at Math."

However, in addition to *reflecting* stereotypic beliefs about gender, these statements also clearly *reinforce* them.

Moreover, the naturalness of subject-complement statements about equality is precisely what makes them pernicious. In Experiment 3, we asked participants directly whether they believed that subject-complement statements about equality (e.g., "Girls do as well as boys at math") were biased against the gender in the subject position (here, girls). We found that, overall, participants considered these statements to be largely egalitarian. This suggests that when people hear or read subject-complement statements about equality—as in Experiments 1 and 2—they might process and accept the damaging implicit messages without fully realizing it.

Other research in experimental pragmatics has shown, in fact, that is *generally* hard for people to recognize and reject implicit messages in language—a weakness that attorneys and reporters often exploit (e.g., Woodbury, 1984). Asking witnesses of a car accident, "How fast were the cars going when they smashed into each other," for example, elicits higher speed estimates than asking, "How fast were the cars going when they hit each other?" by presupposing that the collision was high-impact (Loftus & Palmer, 1974). Similarly, stating "The woman was abused by the man" rather than "The man abused the woman" causes people to be more accepting of violence against women, because passive voice distances perpetrators from their crimes and consequently makes the crimes seem less severe (Henley, Miller, & Beazley, 1995).

Importantly, when implicit messages like these are made more explicit, people are less likely to accept them. Warning people in advance that the language they hear might be misleading (Greene, Flynn, & Loftus, 1982), attributing language to sources who wish to be persuasive (Dodd & Bradshaw, 1980), and encouraging people to process language more slowly and carefully (Tousignant, Hall, & Loftus, 1986) all improve people's abilities to detect and dismiss potentially biasing or misleading information.

Thus, in some ways, implicit messages might shape beliefs more powerfully than explicit messages that express the same ideas. Consider the generic pronoun *he* (e.g., "To each his own"). Although many people use and accept this pronoun as generic, its masculine form actually biases listeners to think that a typical person is male (Gastil, 1990; Hamilton, 1988; MacKay & Fulkerson, 1979; Martyna, 1978). People who use generic *he*, however, would probably disagree with the explicit assertion "Typical people are male." Generic *he* might thus be more successful at connecting maleness with typicality than any sort of explicit statement. Likewise, when a person states outright that boys are naturally better than girls at math, or that women do not have what it takes to succeed in the tech industry, it is easy to disagree with these claims. A person can say, "No, that's wrong." But when someone says, "Girls are as smart as boys," it is much harder to recognize and reject the implicit message that boys are typically, naturally smarter. Implicit messages, because they are subtle, harder to detect, and often presupposed in conversation, cannot so easily be challenged.

Conversely, in interventions, we might be able to *harness* the power of implicit messages to foster belief change. Recall that participants in Experiment 1 who read statements that implicitly contradicted stereotypes about math ability in their framing (e.g., "Boys do as well as girls at math") were the least likely to state that boys had more raw talent—in fact, they reported that *girls* had more raw talent. Had those same participants instead read, "Girls are naturally better than boys at math," however, they might have been less likely to endorse these counter-stereotypic beliefs. They might have been offended by the explicit claim, or they might not have wanted to believe it was true because it clashed with other beliefs that they held (see Kahan, 2013; Fielding & Hornsey, 2016; and Hornsey, Harris, & Fielding, 2018; for discussions of the relation between motivation and belief change). Indeed, in psychological research, such explicit efforts to change beliefs have often been unsuccessful (for a review, see Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). Explicitly presenting participants with evidence that vaccines are beneficial, for instance, has typically had mixed effects on people's beliefs at best, and sometimes has even caused people to hold on to their prior anti-vaccine beliefs more strongly (Nyhan & Reifler, 2015; Nyhan, Reifler, Richey, & Freed, 2014). Gently "nudging" participants through more implicit messages might thus be a more effective strategy for counteracting gender stereotypes and encouraging more egalitarian behavior (see Thaler & Sunstein, 2008, for a discussion of nudging in psychological research).

To encourage girls to pursue STEM degrees, then, it is ineffective to tell them that they are just as good as boys at math. Indeed, for young children who do not yet have math-related stereotypes, these statements may be not only unsuccessful, but detrimental. Since young children are sensitive to some of the asymmetries created by subject-complement sentence structures (Chestnut & Markman, 2016), hearing someone say, "Girls are as good as boys at math," could actually help instill gender biases about ability. Children as young as 6, for example, begin to believe that a "really, really smart" person is more likely to be a man than a woman (Bian, Leslie, & Cimpian, 2017). And this persists into adulthood: One of the obstacles preventing women from succeeding in fields such as physics and engineering is the stereotype that women lack raw talent (e.g., Leslie, Cimpian, Meyer, & Freeland, 2015; Storage, Horne, Cimpian, & Leslie, 2016). Our findings highlight linguistic framing as one potential source of these beliefs. Although statements such as "Girls are as good as boys at math" seem to express equality, latent in their structure is the subtle, implicit message that the gender framed as the reference point is the one with more raw talent. Instead of promoting equality, these statements could thus backfire and *teach* children that boys are naturally more talented.

Acknowledgments

We thank Kara G. Weisman, David N. Abram, and our anonymous reviewers for their helpful feedback on this work.

Note

1. We replicated this experiment using the identical design, except we deleted the last two sentences in the paragraph. We removed these two sentences in case they might

have suggested that the tests used to assess math ability were too easy to be able to accurately measure gender differences, which could have made participants less likely to accept that girls and boys do have equal math ability. Even if this were true, however, it would not account for the differences we predicted among the four experimental conditions, because each condition had the same last two sentences. Nevertheless, we removed these two sentences in this additional study and replicated the results from Experiment 1. We provide these results in our supplementary materials.

References

Banaji, M. R., & Hardin, C. (1996). Automatic stereotyping. *Psychological Science*, *7*, 136–141. https://doi.org/10.1111/j.1467-9280.1996.tb00346.x.

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*, 272–281. https://doi.org/10.1037/0022-3514.65.2.272.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244. https://doi.org/10.1037/0022-3514.71.2.230.

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391. https://doi.org/10.1126/science.aah6524.

Bowdle, B., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology*, *3*, 244–286. https://doi.org/10.1006/cogp.1997.0670.

Bowdle, B., & Medin, D. (2001). Reference-point Reasoning and Comparison Asymmetries. In J. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 116–121). Hillsdale, NJ: Erlbaum.

Bruckmüller, S., & Abele, A. (2010). Comparison focus in intergroup comparisons: Who we compare to whom influences who we see as powerful and agentic. *Personality and Social Psychology Bulletin*, *36* (10), 1424–1435. https://doi.org/10.1177/0146167210383581.

Bruckmuller, S., Hegarty, P., & Abele, A. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. *European Journal of Social Psychology*, *42*, 210–218. https://doi.org/10.1002/ejsp.858.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach* (2nd ed). New York: Springer.

Chestnut, E. K., & Markman, E. M. (2016). Are horses like zebras, or vice versa? Children's sensitivity to the asymmetries of directional comparisons. *Child Development*, *87*, 568–582. https://doi.org/10.1111/cdev.12476.

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*(5), 461–480. https://doi.org/10.1017/S0140525X13002197.

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, *19*, 695–704. https://doi.org/10.1016/S0022-5371(80)90379-5.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime and visual processing. *Journal of Personality and Social Psychology*, *87*, 876–893. https://doi.org/10.1037/0022-3514.87.6.876.

Fielding, K. S., & Hornsey, M. J. (2016). A social identity analysis of climate change and environmental attitudes and behaviors: Insights and opportunities. *Frontiers in Psychology*, *7*, 1–12. https://doi.org/10.3389/fpsyg.2016.00121.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles*, *23*, 629–643. https://doi.org/10.1007/BF00289252.

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology*, *15*(1), 183–217. https://doi.org/10.1080/10463280440000026.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21.

Girls = Boys at Math. (2008). *Science*. Retrieved from http://www.sciencemag.org/news/2008/07/girls-boys-math

Gleitman, L., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts. *Cognition*, *58*(3), 321–376. https://doi.org/10.1016/0010-0277(95)00686-9.

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior*, *21*, 207–219. https://doi.org/10.1016/S0022-5371(82)90571-0.

Hamilton, M. C. (1988). Using masculine generics: Does generic *he* increase male bias in the user's imagery? *Sex Roles*, *19*, 785–799. https://doi.org/10.1007/BF00288993.

Hegarty, P., & Pratto, F. (2001). The effects of social category norms and stereotypes on explanations of intergroup differences. *Journal of Personality and Social Psychology*, *80*(5), 723–7365. https://doi.org/10.1037//0022-3514.80.5.723.

Henley, N. M., Miller, M. D., & Beazley, J. A. (1995). Syntax, semantics, and sexual violence: Agency and the passive voice. *Journal of Language and Social Psychology*, *14*, 60–84. https://doi.org/10.1177/0261927X95141004.

Heyman, G. D., & Compton, B. J. (2006). Context sensitivity in children's reasoning about ability across the elementary school years. *Developmental Science*, *9*, 616–627. https://doi.org/10.1111/j.1467-7687.2006.00540.x.

Hornsey, M. J., Harris, E. A., & Fielding, K. S. (2018). The psychological roots of anti vaccination attitudes: A 24-nation investigation. *Health Psychology*, *37*, 307–315. https://doi.org/10.1037/hea0000586.

Hyde, J. S., & Linn, M. C. (1998). Gender differences in verbal ability: A meta analysis. *Psychological Bulletin*, *104*, 53–69.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494–495. https://doi.org/10.1126/science.1160364.

Jackson, L. A., Hodge, C. N., & Ingram, J. M. (1994). Gender and self-concept: A reexamination of stereotypic differences and the role of gender attitudes. *Sex Roles*, *30*(9), 615–630. https://doi.org/10.1007/BF01544666.

Jagacinski, C. M., & Nicholls, J. G. (1987). Competence and affect in task involvement and ego involvement: The impact of social comparison information. *Journal of Educational Psychology*, *79*(2), 107–114. https://doi.org/10.1037/0022-0663.79.2.107.

Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *34*(2), 107–117. https://doi.org/10.1037/0003-066X.34.2.107.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*, 407–424. https://doi.org/2013-28010-002.

Karabenick, J. D., & Heller, K. A. (1976). A developmental study of effort and ability attributions. *Developmental Psychology*, *12*(6), 559–690. https://doi.org/10.1037/0012-1649.12.6.559.

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

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, M., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*, 106–131. https://doi.org/10.1177/1529100612451018.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3.

MacKay, D. G., & Fulkerson, D. C. (1979). On the comprehension and production of pronouns. *Journal of Verbal Learning and Verbal Behavior*, *18*, 661–673. https://doi.org/10.1016/S0022-5371(79)90369-4.

Martyna, W. (1978). What does "he" mean? *Journal of Communication*, *28*, 131–138. https://doi.org/10.1111/j.1460-2466.1978.tb01576.x.

Miele, D. B., Son, L. K., & Metcalfe, J. (2013). Children's naïve theories of intelligence influence their metacognitive judgments. *Child Development*, *84*, 1879–1886. https://doi.org/10.1111/cdev.12101.

Miller, D. T., Taylor, B., & Buck, M. L. (1991). Gender gaps: Who needs to be explained? *Journal of Personality and Social Psychology*, *61*(1), 5–12. https://doi.org/10.1037/0022-3514.61.1.5.

Muenks, K., & Miele, D. B. (2017). Students' thinking about effort and ability: The role of developmental, contextual, and individual difference factors. *Review of Educational Research*, *87*, 707–735. https://doi.org/10.3102/0034654316689328.

Nicholls, J. G., Patashnick, M., & Mettetal, G. (1986). Conceptions of ability and intelligence. *Child Development*, *57*(3), 636–645. https://doi.org/10.2307/1130342.

Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, *33*, 459–464. https://doi.org/10.1016/j.vaccine.2014.11.017.

Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics*, *133*, 1–8. https://doi.org/10.1542/peds.2013-2365.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192. https://doi.org/10.1037/0022-3514.81.2.181.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532–547. https://doi.org/10.1016/0010-0285(75)90021-3.

Ross, L. D., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology*, *35*(7), 485–494. https://doi.org/10.1037/0022-3514.35.7.485.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*, 494–509. https://doi.org/10.1037/0022-3514.87.4.494.

Shen, Y. (1989). Symmetric and asymmetric comparisons. *Poetics*, *18*, 517–536. https://doi.org/10.1016/0304-422X(89)90010-7.

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. https://doi.org/10.1371/journal.pone.0150194.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, *92*(1), 144–151. https://doi.org/10.1037/0022-0663.92.1.144.

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading post-event information. *Memory & Cognition*, *14*, 329–338. https://doi.org/10.3758/BF03202511.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352. https://doi.org/10.1037/0033-295X.84.4.327.

Woodbury, H. (1984). The strategic use of questions in court. *Semiotica*, *48*, 197–228. https://doi.org/10.1515/semi.1984.48.3-4.197.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Exploratory analyses for all experiments and a replication of Experiment 1.