Investigating Conflation of Sex and Gender Language in Student Writing About Genetics

Molly A. M. Stuhlsatz, Zoë E. Buck Bracey & Brian M. Donovan

Science & Education
Contributions from History, Philosophy and Sociology of Science and Mathematics

ISSN 0926-7220

Sci & Educ
DOI 10.1007/s11191-020-00177-9

ONLINE FIRST

Springer

Science & Education
https://doi.org/10.1007/s11191-020-00177-9

SI: GENETICS AND IDENTITY

Investigating Conflation of Sex and Gender Language in Student Writing About Genetics

Molly A. M. Stuhlsatz[1] · Zoë E. Buck Bracey[1] · Brian M. Donovan[1]

Accepted: 3 November 2020/Published online: 23 November 2020

Abstract
Conflation of sex and gender is implicated in the development of essentialist thinking, which has been linked to the justification of systems of prejudice in modern society. This exploratory study presents findings from a person randomized control trial conducted with 460 students in 8th–10th grade that investigated the extent to which students conflate sex and gender in their writing about genetics. Students were randomly assigned to one of three short readings that either (1) explained the genetics of sex in plants; (2) explained the genetics of sex in humans; or (3) refuted neuro-genetic essentialism, offering instead a social explanation for why women receive fewer PhDs in science, technology, mathematics, and engineering than men. While previous findings from the authors suggest links between the condition students were assigned to and psychological indicators related to essentialist thinking, no work was done to investigate how students' use of language might implicate cognitive conflation as a possible factor in understanding these results. In this study, student responses to a constructed response writing task given after the reading were analyzed to investigate the use of sex and gender language. Students in all three conditions used both sex and gender language. However, students in the refutational text condition tended to use sex and gender language deliberately in order to explain PhD attainment, while students in the traditional genetics conditions used the terms interchangeably, suggesting subconscious conflation. Students in the genetics of human sex condition were more likely to manifest this conflation than students in the genetics of plant sex condition. Implications for instruction are discussed.

1 Introduction

When someone conflates two words or ideas, they "blend" or "fuze" them (Merriam-Webster 2018) so that their meanings become more difficult to disentangle. Most of the time conflation

✉ Molly A. M. Stuhlsatz
mstuhlsatz@bscs.org

1 BSCS Science Learning, 5415 Mark Dabling Blvd., Colorado Springs, CO 80906, USA

is harmless or even productive for science learning, but it could also lead to fundamental misunderstandings about the natural world that contribute to problematic societal discourses about the nature of human difference. The conflation of sex and gender is not a new topic for scholarship. According to legal scholar Francisco Valdez, there exists a "historic and contemporary confusion and distortion of sex, gender, and sexual orientation as social and legal constructs...[1]This conflation's long history and pervasive presence in our thinking and institutions combine to make it self-perpetuating and often invisible, but never absent" (Valdes 1996). Biological differences that we use to designate babies as either "male" or "female" at birth are "socially meaningless until social practices transform them into social facts" (Lorber 1993). Conflation of biology with gendered traits is one of those social practices. The use of sex and gender words interchangeably may seem like a careless mistake, but this linguistic confusion can contribute to a cognitive conflation of those traits that we have inherited (e.g., sex organs), and those that arise from culture and socialization (e.g., gender identity). This conflation of constructs may manifest as linguistic conflation of the words associated with gender (i.e., man, woman, she, he), and the words associated with sex (i.e., male, female, penis, vagina). This cycle of conceptual conflation and linguistic confusion has not been well studied, but psychological research indicates that the language we use can reify a problematic way of thinking known as essentialism (Rhodes et al. 2012), which is used to justify systems of prejudice in modern society.

Gender essentialism is the idea that differences between men and women are caused by biology and are natural and unworthy of redress (Smiler and Gelman 2008). Essentialist beliefs are quite common and influence how people think about cultural, social, developmental, and physical differences between men and women (Arthur et al. 2008; Eidson and Coley 2014; Henrich et al. 2010; Prentice and Miller 2007; Smiler and Gelman 2008). In science class, students often use vocabulary that refers to distinct concepts interchangeably (i.e., steam and water vapor). This confusion is typically harmless, and can even be productive for science learning (Ash 2008; Brown and Ryoo 2008; Warren et al. 2005). But when students are using vocabulary associated with human identity, it could lead to fundamental misunderstandings about the natural world that contribute to problematic societal discourses about the nature of human difference.

From a sociocultural perspective, the reason for this is that language mediates thought (and vice versa) (Vygotsky 1978). According to Vygotsky, language is a shared social tool that emerges from our interactions with the external world. Furthermore, our use of language varies within and between the complex and overlapping communities of practice to which we belong (Lave and Wenger 1991), including academic communities, social communities, and family communities. Sociocultural theory reminds us that language is not just a social tool, it is also a psychological one. As humans learn, we internalize language over time, enabling cognitive interactions with the self that help to organize thoughts and ideas. In other words, there is no thought without language, and no language without thought. As we navigate the social world, both within the classroom and outside the classroom, we are learning to make sense of a complex language of science and society that is not always clearly defined and may result in psychological biases. If distinctions are not made explicit, then as students construct meaning from the complex amalgam of what they learn in school, from friends, from media, and from family, they may internalize sex and gender language synonymously. Thus, confusion of sex

[1] Note that in this paper, we will look specifically at the conflation of sex and gender, as these constructs are more pertinent to the content in today's K-12 science classrooms.

and gender nouns is not simply a symptom of cognitive conflation, it could feed into a cycle of linguistic confusion *and* cognitive conflation.

While linguistic confusion may appear harmless, cognitive conflation can lead individuals to mistakenly ascribe biological explanations of sex as the causes of complex social problems related to gender, which is implicated in essentializing groups of people (Donovan et al. 2019; Gelman and Taylor 2000; Haslam et al. 2000), stereotyping, prejudice, and the justification of differential social outcomes (Demoulin et al. 2006; Haslam et al. 2002). Previous findings from the authors suggest links between student learning using traditional genetics materials and psychological indicators related to essentialist beliefs and limited growth mindset (Donovan et al. 2019). However, the analyses reported here extend the prior study to investigate quantitatively and qualitatively how students' use of language might implicate cognitive conflation as a possible factor in understanding the prior results.

In the exploratory study reported here, we describe an intervention that asked biology students to grapple with the causes of differences in PhD attainment among women. In the process of revealing the social causes, the reading refuted possible genetic causes for the difference. In the comparison conditions, we asked students to read about either plant or human genetics. What we wanted to know was whether students in the intervention group could clearly differentiate the language of sex and gender to explain PhD attainment in writing and if they used both sex and gender terms to do so. Conversely, we wanted to know whether students that took part in the comparison condition used sex language to explain a phenomenon related exclusively to gender, which would illustrate classical conflation. Finally, we discuss why attending to the language of sex and gender matters for biology education moving forward.

2 Theoretical Framework

2.1 Defining Sex and Gender

Over the last several decades, researchers have called for consensus around a clear distinction between sex and gender in humans (Muehlenhard and Peterson 2011; Unger and Crawford 1993; Gelman et al. 1986; Glasser and Smith 2008). Sex is typically defined by biological characteristics such as gamete size (e.g., sperm versus egg), chromosomes (e.g., XX or XY), reproductive anatomy (e.g., uterus or prostate), and ratio of sex hormones (e.g., testosterone and estrogen) (Richardson 2013). However, while sex is biological, it is not binary (Beldecos et al. 1988). Up to 1.7% of humans do not conform to "male" or "female" sex categories (Fausto-Sterling 2000). Some inherit sex chromosomes that are X, XXX, XXY, or XYY. Some embryos may fuse to produce two different sets of DNA: one XX and the other XY. An embryo may be exposed to chemicals in utero that disrupt the normal conversion of the sex chromosomes into sex hormones, resulting in different hormonal ratios, and perhaps the development of ambiguous outer genitalia or internal organs (Pardue and Wizemann 2001; Krieger 2003).

Gender, on the other hand, is a fluid *social construct* developed through human ideas about attitudes, activities, and beliefs associated with biological sex categories (West and Zimmerman 1987). This construct develops and changes over time and is shaped by overlapping repertoires of practice (Gutiérrez and Rogoff 2003) associated with different communities and societal structures (Bussey and Bandura 1999). Gender often defines one's societal role, and varies across culture and time (Diesendruck et al. 2013). Like sex, gender is also a non-binary category that includes (but is not limited to) such identities as hijra, genderqueer, non-binary, alyha, mahu, transgender, and two-spirit. While "male" and "female" are typically associated with sex categories that extend across species, the words "boy" and "girl" are gender categories that should only be applied to humans. This is not to say that these categories are mutually exclusive. On the contrary, human lives are impacted by both sex and gender categories, and these categories can influence one's lived experience in complex ways. According to Hyde et al. (2019, p 184), "Gender and sex are closely intertwined such that sex cannot be studied without consideration of gender, and studies of gender can often benefit from considering sex as well." While we acknowledge this complexity, for the purposes of this research it is important to clarify and make explicit differences based on biology, differences based on social categories, and the interplay between the two (Krieger 2003).

There is a long history of sex and gender in the sciences, and not simply in human biology. Early European botanists ascribed colorful descriptions of plants that were often in line with descriptions of human biology, and notably traditional gender roles (Schiebinger 2013). In *Nature's Body*, Linda Schiebinger recounts the ways that just as students use their intuitive knowledge and learned biases about the natural world in the classroom, scientists have long done the same. By applying their conceptual understanding of human sexual reproduction, sexuality, and gender norms to plants, these botanists attempted to make sense of a new phenomenon, through using similar language and concepts. Schiebinger (2013) notes that the use of metaphor during this time was particularly focused in conceptions of gender roles, with Linnaeus and Erasmus Darwin characterizing plant reproduction in terms of "nuptials," providing vivid descriptions of "wedding gowns" and "bridal beds." It is not surprising that while not as overt, we continue to see evidence of the interplay between concepts of biological sex and gender in the biology curriculum. Furthermore, botanists do use the words sex and gender to describe distinct characteristics of a plant related to reproduction. According to Borges (1998), "the term plant sex denotes the various mating systems of plants... while the term plant gender refers to the relative representation of male and female function in the entire plant." However, for the purposes of this study, we define gender as a fluid social construct as detailed above. Using this definition, plants do not have gender.

2.2 Gender Essentialism and its Development Through Language

Gender essentialism is the belief that genes inherent in people make people of the same gender physically and behaviorally uniform, and that people of different genders are physically and behaviorally discrete. These ideas may develop at a very young age (Taylor 1996). Evidence suggests that humans tend to divide people into discrete gender categories that are fixed at birth and deterministic of physical ability, intelligence, and aptitude for certain societal roles or careers (Gelman et al. 1986; Smiler and Gelman 2008). Essentialist beliefs are socioculturally situated, deeply ingrained, and can be harmful, as they support social stereotyping and contribute to prejudice (Rhodes et al. 2012). Endorsement of essentialist beliefs has been correlated with gender stereotypes (Bastian and Haslam 2006; Yzerbyt and Rocher 2002; Bigler and Liben 2007), and beliefs about innate academic ability that contribute to persistent disparities in STEM fields (Donovan et al. 2019).

When gender and sex are conflated, it can lead to a particular type of bias called neuro-genetic essentialism, which is the belief that inherent differences in the genes and brains of men and women cause gender disparities. Neuro-genetic essentialism may lead to stereotypes that have detrimental social impacts, in particular, for women (Donovan et al. 2019) and gender diverse individuals. People who hold strong neuro-genetic essentialist beliefs think that differences among men and women are unable to change, because they are innate biological factors connected to the brain, not socially constructed inequities. Those who exhibit strong belief in neuro-genetic essentialism are also more likely to show greater gender bias, including prejudice toward transgendered people (Ching and Xu 2018), and gender stereotyping (Haslam et al. 2000). Stereotyping based on essentialism works to promote and reify the social structures we observe in everyday life, with implications that reach far beyond the classroom. Furthermore, the use of generic noun phrases, subconsciously or consciously, has been shown to perpetuate essentialist thinking (Rhodes et al. 2012; Hyde et al. 2018). A generic noun phrase is a noun phrase that does not refer to a specific individual (or set of individuals), but rather to a kind or class of individuals (Reiter and Frank 2010). For example, "metals are good conductors" is a generic noun phrase applying conductivity to all metals. For humans, generic noun phrases are often applied to social categories like gender (i.e., "men were hunters and women were gatherers"). Rhodes et al. (2012) tested the hypothesis that the use of these generic phrases facilitates the transmission of essentialist beliefs from parents to children. They found that hearing generic language about a social category led both children and adults to develop essentialist beliefs about that social category, and that when parents held essentialist beliefs about a social category, they were more likely to use generic noun phrases when discussing the category with their children. Similarly, Walton and Banaji (2004) found that linguistics matter in essentialist thinking in very subtle ways. They discovered that when preferences are expressed in noun phrases as opposed to verb-phrases, beliefs about people are stronger and less malleable.

These findings suggest the language that we use to talk about social categories like gender matters with regard to the development of essentialist beliefs about those categories. The use of man/woman interchangeably with male/female conflates gender with biological difference, implying that men differ from women in social ways because of some natural cause. Thus, a simple essentialist explanation for a set of very complex social patterns might emerge as a result of conflating sex and gender in texts. This phenomenon played out in a study conducted by Coleman and Hong (2008). In their study, undergraduate students were randomly assigned to read two different versions of a text: one that argued that gender is a biological construct and one that argued that gender is socially constructed. Women in the biological explanation condition were significantly more likely to endorse essentialist beliefs after the intervention than women in the social constructivist condition (Coleman and Hong 2008). Thus, assigning a biological explanation of gender can result in furthering essentialist thinking, particularly in women.

2.3 Language Simplification in the Biology Curriculum

The language of the biology classroom is a complex hybrid of familiar words and ways of speaking that come from students' everyday lives, and new words and ways of speaking introduced in the classroom (Ash 2008; Brown and Ryoo 2008). Each modality serves an important purpose—everyday talk provides a means for students to express their ideas (Warren et al. 2005), while the introduction of scientific vocabulary and discourse supports students in thinking and communicating more like scientists (Lemke 1990). For example, in talking about the reproduction of plants, a teacher might introduce the existence of "male" and "female" trees. As students work in small groups to diagram the way that trees reproduce, they argue over what to draw. "The *girl* tree needs to be closer to the *boy* tree so that it can catch the seeds and get fertilized" says one student. This student is expressing a scientific idea about the importance of proximity in plant reproduction and using the scientific word "fertilization" correctly. But the student also brings in some everyday language: boy and girl, to talk about biological sex. These words are associated with human gender, a complex social identity that is entangled with, but not the same as, biological sex (Beldecos et al. 1988). By using boy/girl instead of male/female, the student is conflating two distinct ways that we learn to talk about human difference. It is reasonable for classroom discussion of human difference to include some conflation of sex and gender, however, discussion of plants and trees as "boy" or "girl" is inappropriate. While it is productive in the science classroom for students bring everyday language to help explain scientific phenomena, ideas that may have important social or scientific implications require additional attention to help students disentangle the meaning of each word.

The connection between language, essentialism, and prejudice suggests that science educators should pay close attention to the way that sex and gender are described in science textbooks (Bianchini 1993). Unfortunately, limited research suggests that science textbooks tend to use sex and gender language interchangeably (Bazzul 2013; Bazzul and Sykes 2011). Bazzul (2013) and Bazzul and Sykes (2011) found that sex/gender binaries were mentioned throughout science textbooks in a variety of distinct topic areas. In a review of high school biology textbooks, Snyder and Broadway (2004) note that representations of heteronormativity are pervasive, privileging and affirming the status quo, and excluding the experiences of those who do not fit that stereotypical framing. The texts tended to essentialize differences between males and females and revealed that the biological basis for both sex and gender was "covertly and overtly interchanged suggesting essentialized identities for each of the two sexes and their corresponding genders" (Bazzul and Sykes 2011, p. 278–279). Furthermore, in a content analysis of four biology textbooks, Bazzul (2013) found multiple examples of gender ascribed to biological contexts. For example, in a text on reproduction, she found examples such as "Use blue dots for boys and red dots for girls. Does there seem to be a gender difference with regard to the age at which infants learn to walk?" or "A couple has two children, one which is a boy. What is the probability that the other child is a girl?"

In an investigation of the use of the term "sex hormones" in secondary biology texts, Nehm and Young (2008) found that the use of the term may lead students to identify hormones as being exclusive to female or male bodies, a concept that was dispelled in the early twentieth century. Not only are these definitions scientifically inaccurate, they could also lead students to see differences in society as scientifically based and fixed, rather than social constructs that are malleable (Bazzul and Sykes 2011; Brescoll and LaFrance 2004).

When the biology curriculum boils human difference down to females as XX and males as XY, the differences between the sexes are reduced to purely genetic components. For instance, during a recent visit to Khan Academy (2020a), a popular resource for many teachers and students that touts 10 million subscribers worldwide, we found an example of both oversimplification and conflation reflecting the findings of previous studies (Bazzul 2013; Bazzul and Sykes 2011; Nehm and Young 2008; Brescoll and LaFrance 2004). The first thing that one encounters in the sex-linked traits section of the high school biology module is a video titled "Sex-linked traits". The video starts by asking the question "How is gender determined in an organism?" The narrator first writes in big yellow letters at the top of the screen "Gender." They go on to walk through sex-determination in terms of XX and XY, bluntly stating to the viewer "And to figure out if something is a male or a female it's a pretty simple system. If you've got a Y chromosome, you are a male." Further into the module, the lesson turns to X-inactivation (including disorders like Klinefelter). The text starts out using appropriate "female" and "male" language, but quickly turns to gender terms, stating "Why doesn't it cause problems for men to have just one copy of the X chromosome, while women have two?" (Khan Academy 2020b).

This reduction and clear oversimplification, while attempting to help students understand the chromosomal nature of human biology, could inadvertently justify belief in an underlying causal essence as has been established in previous studies of gender (Donovan et al. 2019). This explanation does not acknowledge the complexity of internal reproductive anatomy, hormone balance, human relationships, or intersex characteristics. If through a classical biology education students' come to believe that gender categories are essentially discrete because of XX or XY chromosomes, they may also believe that these categories will explain what they observe in the social world. Clément and Castéra (2013) note that simplification of the complex system of genotype and environment interaction as an additive model rather than as an interaction between the two is common in biology texts around the world. They also note that textbooks tend to use phenomena with human representations that emphasize features like twins with similar hairstyles and clothing, which communicates an implicit message to students about the connections between genetics and socially constructed attributes. As students experience the biology curriculum, they amass new information about socially appropriate roles, preference for activities, and social behavior (Maccoby 1988). A mixture of biological sex and social cues within textbooks may appear subtle, yet over time it can lead students to reduce the causes of complex social problems to seemingly simple biological explanations. Research suggests that ultimately this kind of oversimplification is implicated in essentializing groups of people (Donovan et al. 2019; Donovan 2017; Gelman and Taylor 2000; Haslam et al. 2000; Clément and Castéra 2013).

The gendered oversimplification of biological sex in textbooks should not be surprising, given the well-documented history of the inappropriate application of gendered metaphors to science topics ranging from human reproduction to slime molds (Martin 1991; Spanier 1995; Keller 1985; Schiebinger 2013), thus "imposing on nature the very stories we like to hear"(Keller 1985, 187). Luckily in recent years, many members of the scientific community have acknowledged conflation of sex and gender as potentially troublesome (Eveleth 2014; Bond 2013). A new set of recommendations called The Sex and Gender Equity in Research guidelines have been developed to tackle the underrepresentation of issues of sex and gender in peer review of scientific research (Heidari et al. 2016). In particular, the guidelines are intended to address the confusion between sex and gender in peer-reviewed medical studies and to move the editorial staff of scientific journals to attend to issues of underrepresentation in published research (De Castro et al. 2016; Del Boca 2016). However, in education, there is little attention being given to the way sex and gender are addressed in biology classrooms despite the fact that it is clearly problematic. But is simply removing conflation of gender and sex in textbooks going to make a difference?

2.4 Building on Previous Work

Data from the current study was collected as part of a student-level randomized control trial (RCT) that measured the endorsement of gender essentialist beliefs, and beliefs about STEM ability based on gender. To test the effect of how learning about the genetics of sex might impact gender essentialist beliefs, in the prior study, students were randomized into one of three treatment groups (Donovan et al. 2019). One group read about the *genetics of human sex* using only sex language (male/female). The second group read about the *genetics of plant sex* using only sex language (male/female). The third group read a sociocultural text designed to directly address and *refute essentialist ideas about men and women* using both gender and sex language.

The results of the RCT indicated that when students learn about the biology of sex difference through either a human or plant context, they are more likely to endorse belief in neuro-genetic essentialism than their peers who read a sociocultural explanation for gender difference. This finding was consistent in both the plant and human conditions, indicating that neurogenetic essentialism is not necessarily limited to a human context (Donovan et al. 2019). Furthermore, relative to the refutational text condition, students who self-identified as girls and received the plant or human interventions were more likely, on average, to endorse the belief that differences between people cannot change, or is fixed (Dweck 2008). For girls, but not boys, the endorsement of a fixed mindset was mediated by neuro-genetic essentialism. Finally, endorsement of fixed mindset also significantly correlated with future interest in science in girls, but not boys. This means that girls that nominate higher neuro-genetic essentialist beliefs were also more likely to endorse a fixed mindset, and that fixed mindset was implicated in interest in STEM. However, this indirect effect (mediation) could be related to internalized negative gender stereotyping. These findings reinforce the fundamental idea that learning about human difference may not be a socially neutral endeavor and suggest that the way that we present human difference in texts can impact students' ideas about how gender and ability are associated. Furthermore, providing students the opportunity to better understand the complexity of how genetic and social causes interact could counteract reliance on simplistic genetic models of differences between genders (Donovan et al. 2019).

While the findings from our prior study suggest that essentialism was a mediating factor in the effect of the readings on the belief that science ability is innate, the study did not go so far as to investigate the mechanisms associated with these beliefs. Specifically, the prior study did not investigate the linguistic mechanisms linking the content of the treatment texts with the development of essentialist beliefs in the minds of the readers. The current study builds on the prior findings by investigating how students express their understanding of genetics of plant and human sex through the use of sex and gender language, and how sex and gender language are used to explain gender bias in STEM degree attainment (a sociological explanation for gender disparities).

2.5 The Need for this Study

Our goal with this study was to address the mechanisms linking the texts to essentialist ideas, by taking a closer look at the language being used by students in their responses to all three conditions from the RCT. We wondered, were students bringing in everyday gender words like "boy" and "girl," to explain the biology of sex differences? Was use of language different depending on the condition they were randomized into? And in what context were students using gender language as opposed to sex language?

This translated into the following exploratory research questions:

1. With what frequency did students in all three conditions apply gender language or sex language in their response to the writing task?
2. With what frequency did students use both sex and gender in their writing (conflation)?
3. Did the frequency of sex and gender language differ between those reading about the genetics of plant sex (which have no gender), the genetics of human sex, and the refutational text, and were these responses consistent with conflation of sex and gender?

3 Methods

3.1 Study Procedure

This study is an exploratory secondary analysis of a larger study. The intervention (described above as part of the previous RCT) took place in a single class period (approximately 45 min) in high school biology classrooms. The intervention came after learning about autosomal inheritance, but prior to learning about sex-linked inheritance. Each student was randomly assigned to a treatment condition. Random assignment was done first within each classroom and second by self-identified gender using the block-randomization available in Qualtrics, the survey platform used to deliver treatments in our study. This design is an *individually randomized trial with clustering* (IRTC), where the treatment is assigned and delivered to each person and clustering occurs based on how students are nested within classrooms (Weiss et al. 2016). The IRTC allows modeling of the treatment effects at the student level (level 1) which contributes to higher statistical power. Students participated in the experiment using a laptop or tablet in their classroom or in the school computer lab.

To achieve block random assignment within gender groups, we asked students to self-identify their gender. The question asked, "If you had to pick between the categories below, how would you identify your gender?" The options to choose were "male," "female," "other." 49.4% of students identified as "male," 49.8% identified as "female," and .8% identified as "other" (see Table 1) In a follow-up item, we asked students "If you could describe your gender to another person, how would you describe it?" To achieve balance in the study, within each classroom, student's responses to the gender identity question were used for randomization. This randomization method resulted in an equal balance within classrooms and across the three conditions.

Prior to receiving the randomly assigned treatment (genetics of plant sex, genetics of human sex, or the refutational text on PhD attainment), students responded to pretest measures of the dependent variables and covariates. Then, students read the assigned text-based intervention and responded to follow-up reading comprehension questions. Included in the reading comprehension questions was a constructed response question that was used as a compliance measure and is the focus of this study (see measurement section, below). Finally, students again responded to the dependent variables of interest in the prior study, including measures of uniformity and discreteness, inherent genetic causes for trait differences between genders (Donovan et al. 2019), gender essentialism (Rhodes and Gelman 2009), neuro-genetic-essentialism (Donovan et al. 2019), entity theories of science ability based on the field-specific ability beliefs scale (Leslie et al. 2015), future interest in STEM (Kosovich et al. 2015), and a subset of items from the inherence heuristic based on the scale created by Cimpian and Solomon (2014).

3.2 Sample

A total of 460 students in 9th–10th grade from five schools in the San Francisco Bay area participated in the study. Only students with assent and parental consent were enrolled in the study. Fifty-one percent of students self-identified as female. Participating schools included two private schools (13.3% of sample) and one was single sex (6.9% of sample, all boys). Both private schools served high socio-economic status students. Free and reduced-price lunch (FRPL) varied from 6.8% at one school site to 60% at another site (see Table 1). The percent

Table 1 Demographics of sampled schools

School	% FRPL	% American Indian/Alaskan Native	% Asian	% Black	% Hispanic	% Native Hawaiian/Pacific Islander	% White	% 2 or more races	Co-Ed	% of sample
1	6.8	.6	31.1	3.3	9.3	1	48.9	5.4	Yes	46.5
2	68	.7	17.8	15.4	50.2	.8	10.5	4.2	Yes	15
3	40	.2	64.6	1.8	9.9	.9	15	7.1	Yes	25.2
4	NI	NI	NI	NI	NI	NI	NI	NI	No	6.9
5	NI	NI	NI	NI	NI	NI	57	NI	Yes	6.4

Table 2 Frequency of sex and gender language by term

Sex language	%	Gender language	%
Female	70.2		
		Woman	30.70
		Women	29.10
		Girl	5.90
		Wife	0.40
		Her	0.90
		She	2.00
Male	72.0		
		Man	2.80
		Men	16.30
		Boy	2.60
		Guy	0.40
		Husband	0.40
		His	0.20
		He	0.20
Sex	17.4		
		Gender	19.30
		Feminine	1.10
		Masculine	0.90

of white students varied from 57% at one school to 10.5% at another. For the larger study (Donovan et al. 2019), we did not identify variability in treatment effects across schools.

3.3 Treatments

The three treatments in this study were crafted to attend to contemporary thinking around the biological and social scientific consensus about sex and gender differences. The first text led students through a reading on the genetics of human sex that was adapted from a BSCS biology curriculum (BSCS 2007). The second reading on the genetics of plant sex was conceptually the same as the genetics of human sex reading treatment but adapted to take the concepts of the human reading into the plant context. The third reading was a text refuting neuro-genetic essentialism focused on reviews and meta-analyses about sex, gender, and science achievement and ability beliefs (Bian et al. 2017; Ceci et al. 2009; Coley 2001; Flynn 1999; Joel et al. 2015; Voyer and Voyer 2014; Leslie et al. 2015; Storage et al. 2016). For a detailed description of each of the interventions, please see Donovan et al. (2019). Finally, each of the interventions varies in how they attend to essentialist belief structures like uniformity, discreetness, and inherent causation, which should re-enforce essentialist beliefs (genetics of plant and genetics of sex) or refute essentialist beliefs (refutational text).

Each intervention condition was developed to be similar in length, conceptual flow, and difficulty. The genetics of plant sex and genetics of human sex texts emphasized that males and females are discrete genetically based categories, thus triggering students who conflate sex and gender to likely assign this discreteness to the gender categories of "men" and "women." The refutational text, on the other hand, provided a clear refutation of these ideas by disregarding a genetic explanation and supplying students with a research-based explanation that detailed differences in STEM attainment between men and women as socially constructed. We tested the grade level of the text using Compleat Lexical Tutor (2017), and we found the texts were all written at the ninth-grade level. This analysis confirmed that the genetics of human sex and genetics of plant sex materials had 92% of the same words. The refutational text retained 62% of the words from the genetics of human sex text and 65% of the genetics of plant sex reading. Further analysis of the texts investigated the extent to which sex or gender language appeared in each. Each of the texts included sex-based language, but the genetics of human sex and genetics of plant sex texts did not include any gender language like "woman" or "gender." Conversely, the refutational text emphasized gender language through the explanation of the social causes of gender disparities in STEM fields.

3.4 Dependent Variables

After experiencing one of the three interventions, students answered a series of post-test measures, including a reading comprehension task and a constructed response compliance question that asked them to write about what they learned from the reading. The questions were phrased differently depending on the treatment:

- Genetics of plant sex intervention: "How do genes cause differences between male and female plants?"
- Genetics of human sex intervention: "How do genes cause differences between male and female humans?"
- Refutational reading: "How do stereotypes cause differences between male and female humans?"

3.4.1 Sex Language and Gender Language

The first step in our analysis was to identify responses where either sex or gender nouns were used. Using SPSS Text Analytics Version 4.0, we coded responses from the post-test constructed response compliance question for all students. We identified two categories; sex nouns and gender nouns. The coding occurred during one session, by one researcher, with treatment condition blinded and each response classified using the same criteria across treatments (Krippendorff 2018). Thus, the computer coding was completed for all responses at the same time, regardless of experimental condition. Responses were classified as including sex nouns when the words are "female," "male," or "sex." Responses were classified as including gender nouns when the response included some variation of gender language (girl, boy, man, woman, gender, etc.). This coding resulted in two dichotomous variables we titled "sex language" and "gender language." Thus, the presence of the target language was coded as "1" and lack of language was coded as "0". Table 2 lists the words that were included, and the frequency with which each word appeared in the student responses. Fifty-five students (12%) did not use any sex or gender language in their written responses.

3.4.2 Combination of Sex and Gender Language

After coding the individual sex and gender nouns, the next step in exploring conflation was to identify responses where students used both sex and gender language for further analysis. To do this, we calculated a new variable to indicate those responses that were coded for both sex and gender language. In other words, any single response that includes both sex language and gender language is coded "1" for this variable, no matter the number of times the language is used. Responses not meeting the criteria (use of both sex and gender language) were scored as "0". For instance, the following student response is an example from the genetics of human sex treatment that was coded as "1" because the student used both sex and gender language in their response.

> One difference between **male** and **female** humans is that males have one X chromosome and one Y chromosome and **females** have two X chromosomes. This can cause **males** to more likely be color-blind than **women** because the gene for color-blindness is located on the X chromosome and not the Y chromosome.

Responses that were coded this way included both sex and gender language, but were not necessarily conflation of the constructs of sex and gender, so our next step was to review each response that received a 1 to determine if evidence of conflation was present in these responses. We were looking for indications that the student (1) did indeed use both sex and gender language in reference to their scientific writing, and (2) that there was evidence that the student was doing so unintentionally, rather than in an attempt to explain the distinction between the two. This step of the coding process was completed by one researcher and then validated by a second. Thus, the first coding step was intended to simply identify responses that used both types of language and the second step was to consider how the language was used within the responses.

3.5 Statistical Analyses

Each of the outcome variables in this study was categorical (presence or absence of sex or gender language and presence or absence of both). We used logistic regression (Stata version 15) to predict the probability of students using sex or gender language in their response based on their treatment group. Each of the exploratory questions was tested using logistic regressions which models the binary dependent variables are apparent (1) or not (0) as predicted by the treatment variables. For each model, we investigate the probability of the dependent variables (use of gender language, use of sex language, use of a combination of sex and gender language) as an odds ratio (OR) and then convert the odds ratio to a risk ratio (RR) (Sainani 2011). The risk ratio provides an estimate of the increased risk of the dependent variable occurring based on the treatment. Odds ratios are more difficult to interpret because they provide the relative increase or decrease in the odds of an outcome, while the risk ratio provides the relative increase or decrease in risk by dividing the prevalence of the outcome variable in a treatment group by the prevalence of the outcome in the reference (comparison) group (Eq. 1). The odds ratio can be converted to percent of increased (or decreased) risk (Sainani 2011). This relationship between odds ratio and risk ratio is mathematically expressed as:

$$\text{Risk Ratio} = \frac{\text{OR}}{(1-P\,\text{ref}) + (P\,\text{ref}*\text{OR})} \tag{1}$$

3.6 Qualitative Analyses

Following the quantitative analyses, we conducted a textual content analysis of each student response to better understand the patterns of sex and gender language usage. Our prior text analysis identified students who used sex or gender language exclusively. These became our first two categories "sex only" and "gender only." Next, we turned to the responses where both sex and gender language were used together within a response. Finally, we noted that some students used gender only language, which could also be considered a pattern indicative of conflation in each condition. In the results and findings, we report on the frequencies in each category by treatment condition.

4 Findings

We set out to explore the following questions:

1. With what frequency did students in all three conditions apply gender language or sex language in their response to the writing task?
2. With what frequency did students use both sex and gender in their writing (conflation)?
3. Did the frequency of sex and gender language differ between those reading about the genetics of plant sex (which have no gender), the genetics of human sex, and the refutational text, and were these responses consistent with conflation of sex and gender?

4.1 Research Question 1: With What Frequency Did Students in All Three Conditions Apply Gender Language or Sex Language in Their Response to the Writing Task?

Students in each of the treatments used gender language (see Fig. 1 and Table 3). However, there were differences in the frequencies of use of gender language between the treatment groups. Students in the genetics of human sex condition used gender language significantly more than students in the genetics of plant sex condition (OR = 3.35, SE = .92, $p < .001$). Students in the human condition had a 140% greater relative risk of using gender language

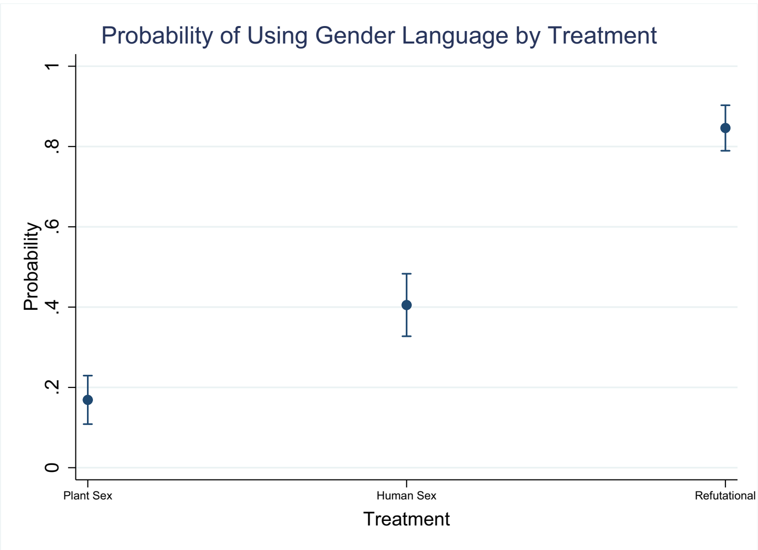

Fig. 1 Probability of use of gender language by treatment group

Table 3 Use of gender language in response by treatment. The genetics of plant sex group serves as a reference group for the genetics of human sex and refutational conditions

Gender language	Risk ratio	Odds ratio	Standard error	*t*-value	*p* value	95% confidence intervals	
Genetics of human sex	2.40	3.352	0.921	4.41	0.000	1.96	5.74
Refutational text	5.01	27.060	8.453	10.56	0.000	14.67	49.92
Genetics of plant sex		0.203	0.045	− 7.25	0.000	0.13	0.31
Mean dependent var			0.479	SD dependent var.			0.500
Pseudo *r*-squared			0.249	Number of obs			457.000
Chi-square			114.893	Prob > chi^2			0.000
Akaike crit. (AIC)			480.955	Bayesian crit. (BIC)			493.329

compared to students in the plant condition (RR = 2.40). Students in the refutational text condition used gendered language significantly more than in the genetics of plant sex condition (OR = 27.06, SE = 8.45, $p < .001$). When converted to a risk ratio, students in the refutational condition have a 401% (RR = 5.01) increased risk of using gender language than students in the genetics of plant condition. Finally, compared to students in the human condition, students in the refutational text condition were 109% more likely to use gender language in their responses (RR = 2.09), and this difference was also statistically significant (χ^2 (1) = 57.0, $p < .001$). In summary, students in the plant condition used gender language, and the proportion of use was significantly greater than zero. Students in the genetics of human sex condition and refutational text conditions were, on average, significantly more likely to use gender language than students in the genetics of plant sex condition.

Students in the genetics of plant sex and the genetics of human sex conditions used sex language at about the same frequency because there was not a significant difference in the odds when comparing these two conditions (OR = .64, SE = .20, $p = .158$) and the associated risk ratio was .93. Students in the refutational text condition used sex language less frequently than students in the plant condition, and the difference was statistically significant (OR = .29, SE = .098, $p < .001$) (Table 4, Fig. 2). Finally, an F-test revealed that students in the refutational condition were 19% less likely to use sex language (RR = 0.81) when compared to the genetics of human sex condition, and this difference is statistically significant (χ^2 (1) = 9.26, $p = .002$). In summary, students in the refutational text conditions were, on average, significantly less likely to use sex language than students in the genetics of plant sex and genetics of human sex conditions.

Table 4 Use of sex language in response by treatment. The genetics of plant sex group serves as a reference group for the genetics of human sex and refutational conditions

Sex language	Risk ratio	Odds ratio	Standard error	*t*-value	*p* value	95% confidence intervals	
Genetics of human sex	0.93	0.641	0.202	− 1.41	0.158	0.35	1.19
Refutational text	0.75	0.287	0.084	− 4.25	0.000	0.16	0.51
Genetics of plant sex		6.400	1.541	7.71	0.000	3.99	10.26
Mean dependent var			0.770	SD dependent var.			0.421
Pseudo *r*-squared			0.044	Number of obs			457.000
Chi-square			20.687	Prob > chi^2			0.000
Akaike crit. (AIC)			477.165	Bayesian crit. (BIC)			489.539

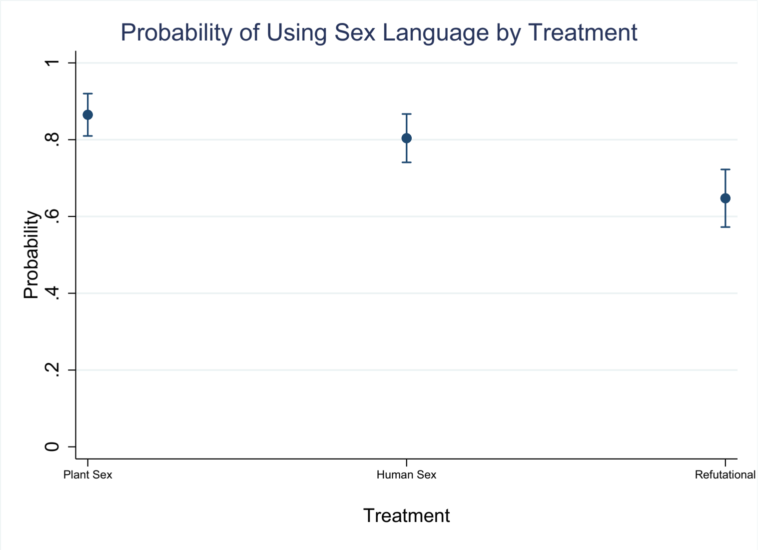

Fig. 2 Probability of use of sex language by treatment group

4.2 Research Question 2: With What Frequency Did Students Use both Sex and Gender in Their Writing?

Investigating the presence of sex and gender language in the same response, students in each of the conditions used combinations of sex and gender language in their responses (Table 5 and Fig. 3). Students in the genetics of human sex condition were 97% more likely (OR = 2.46, SE = 0.69, $p < .001$, RR = 1.97) to use both sex and gender language in their responses to the prompt compared to the plant condition. On average, students in the refutational text condition were 245% more likely to use both sex and gender language in their responses compared to the students in the plant condition (OR = 6.88, SE = 1.88, $p < .001$, RR = 3.45). Finally, an F-test revealed that students in the refutational text condition were 97% more likely than students in the genetics of human sex treatment to use a combination of sex and gender language in their constructed responses and this difference was statistically significant (χ^2 (1) = 18.96, $p < .001$, RR = 1.97). In summary, students in the genetics of human sex condition and refutational text conditions were, on average, significantly more likely to use a combination of sex and gender

Table 5 Use of both sex and gender language (conflation) in response by treatment. The genetics of plant sex group serves as a reference group for the genetics of human sex and refutational conditions

	Risk ratio	Odds ratio	Standard error	t-value	p value	95% confidence intervals	
Genetics of human sex	1.97	2.460	0.686	3.23	0.001	1.42	4.25
Refutational text	3.45	6.888	1.882	7.06	0.000	4.03	11.77
Genetics of plant sex		0.203	0.045	− 7.25	0.000	0.13	0.31
Mean dependent var			0.365	SD dependent var.			0.482
Pseudo r-squared			0.098	Number of obs			457.000
Chi-square			52.455	Prob > chi^2			0.000
Akaike crit. (AIC)			547.115	Bayesian crit. (BIC)			559.489

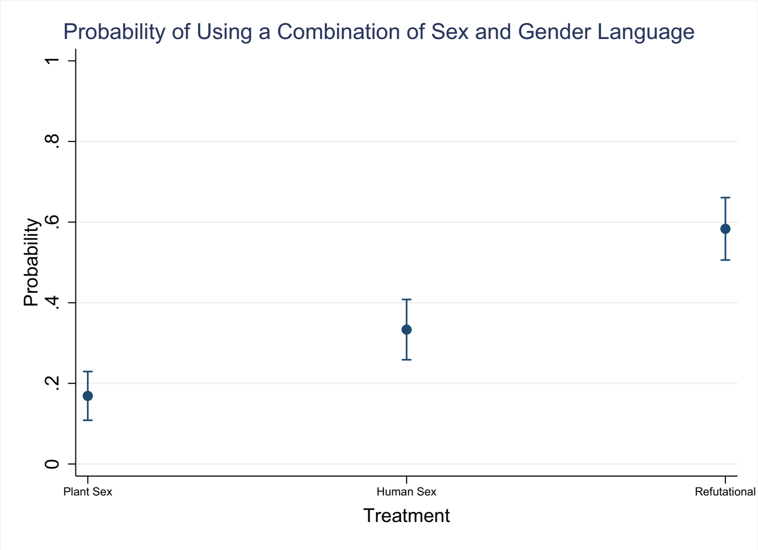

Fig. 3 Probability of conflation of sex and gender language by treatment

language than students in the genetics of plant sex condition, indicating possible conflation in these experimental conditions.

These analyses revealed main effects around the presence and absence of gender, sex, and combinations of gender and sex language that may be indicative of conflation in student responses. Importantly, we observed the use of gender language, sex language, and a combination of the two in each of the treatment conditions. To better identify the characteristics of the responses, we looked more carefully at each student response where both sex and gender language occurred.

4.3 Research Question 3: Did the Frequency and Patterns of Sex and Gender Language Differ Between Those Reading About the Genetics of Plant Sex (Which Have No Gender), the Genetics of Human Sex, and the Refutational Text?

To investigate how students used sex and gender language in their responses, we qualitatively analyzed student responses to further identify evidence of conflation. In the logistic regression analyses, we analyzed response patterns from each treatment group to include sex, gender, or a combination of both. We were particularly interested in the response patterns from students who used both sex and gender language. However, we also wondered if students in the refutational text condition may have used sex and gender language appropriately when they were attempting to explain both sex and gender content from the reading. We also wanted to determine if the trends differed between treatment groups. As we noted above, we identified students with possible conflation using a variable that indicates use of both sex and gender language within a response. Figure 4 illustrates the overall patterns of sex and gender language, sex only language, and gender only language by treatment group.

For responses to the genetics of human sex and genetics of plant sex conditions, we identified the responses that could include conflation as those responses with any gender language. For the refutational text condition, this was a little more complicated. While students

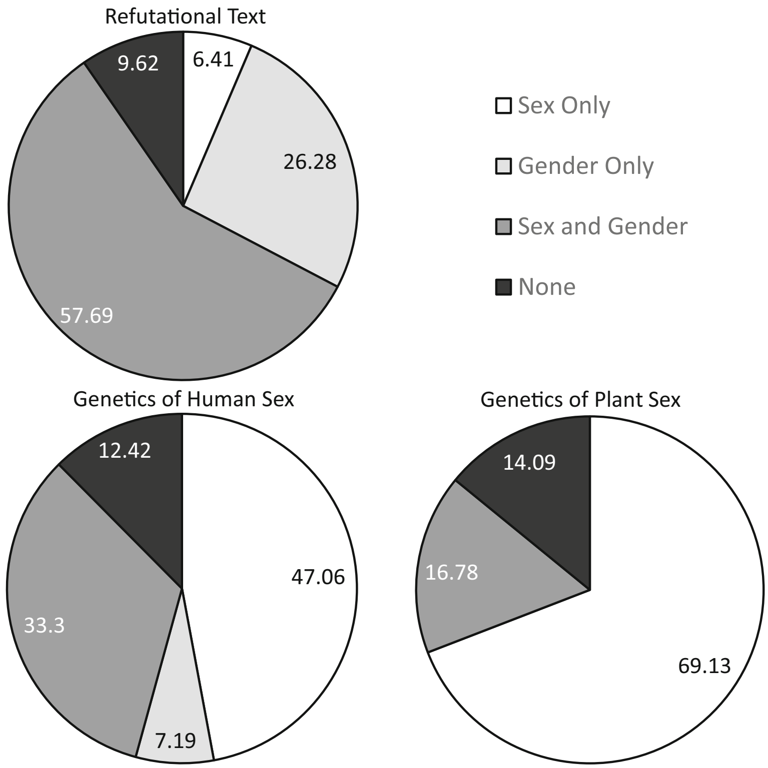

Fig. 4 Frequency of responses by treatment condition

in the genetics of human and plant sex conditions were provided with a reading that was strictly focused on genetic concepts, students in the refutational condition were given a text that included concepts about human brains and used terms like "female brains" and "male brains" as well as information about social phenomena like stereotype threat and PhD attainment, which were associated with gender language like "A stereotype in our society is that men possess intelligence, but women do not." In order to investigate conflation in the refutational text condition, we had to look for *appropriate* use of each of the terms. This resulted in three categories for the responses from students in the refutational text condition.

1) Responses where sex and gender language are used, indicating possible conflation.
2) Responses where sex language is used without gender language, indicating possible conflation.
3) Responses where gender language is used without sex language, indicating possible conflation.

In the first category, responses where sex and gender language are used, indicating possible conflation, we found examples of responses indicating conflation across all three treatment groups. In the second category, sex language without gender language, we found responses indicating conflation in only the refutational condition. In the third category, we found responses where gender language was used without sex language in the genetics of human sex and refutational condition, but not in the genetics of plant sex condition. Table 6 provides examples of each type of response. In the analyses below, we consider the nature of the responses in each category and the evidence that each type of response shows evidence of conflation, either subconsciously or to explain a sociocultural phenomenon.

4.3.1 Evidence of Conflation in the Genetics of Plant Sex Condition

While most students in the genetics of plant sex condition used sex language (69.13%), a statistically significant number of students used both sex and gender language when writing about how genes cause differences between male and female plants (Fig. 4). Only the first category of conflation was present in this treatment group. The students in the genetics of plant sex condition did not use gender language without also using sex language. Additionally, the pattern of student responses indicates that students in the genetics of plant sex condition tended to not assign words like "man" or "woman" to plants, instead they conflated the word "gender" with the word "sex." The sample response in Table 6 illustrates this with the student using sex and gender language interchangeably, mentioning "gender" in the first sentence, then "males" and "females" to discuss the size of the flowers, and then returning to "gender" in the final sentence.

This pattern suggests that students in the plant condition who conflate are subconsciously switching back and forth between sex and gender language. While the rate of conflation in the genetics of plant sex condition was significantly lower than the genetics of human sex and refutational conditions, it is possibly the most instructive finding of this study that some students in the plant condition used both sex and gender language in their responses. If assignment of gender terminology is observed in the genetics of plant sex condition, it is not a surprise that it would also appear in the genetics of human condition at an even higher rate.

4.3.2 Evidence of Conflation in the Genetics of Human Sex Condition

While the many of students in the genetics of human sex condition did indeed use sex language appropriately (47.06%), on average students in the genetics of human sex condition (Table 6, column 2) did use language indicating conflation, and at a higher rate than in the plant condition (Fig. 4). In the example in Table 6, the student uses the terms "female" and "girl," and "sex," and "gender" interchangeably throughout the response.

These examples are different from the examples in the genetics of plant sex condition. In the plant condition, the pattern of conflation was focused on the use of the words "sex" and "gender." In the genetics of human sex condition, student conflation included conflation of "male" and "female" with "man" and "woman" or "boy" and "girl." This is not surprising considering the context. Humans do have gender; however, it is unclear if the students who conflated sex and gender words in our sample understand the distinction between the two. Furthermore, we identified responses from the genetics of human sex condition that used gender language exclusively. This means that despite the question prompt using exclusively sex language, they responded using only gender language.

4.3.3 Evidence of Conflation in the Refutational Text Condition

On average, the students in the refutational text condition were most likely to use both sex and gender nouns in their responses (57.69%). In this category, many of the responses included students starting their response with "male" and "female," and then quickly switching to

Table 6 Examples of conflation by treatment group

	Genetics of plant sex	Genetics of human sex	Refutational text
Sex and gender language indicating conflation	"For some plants the genes are very similar but a certain gene will decide the **gender** of that plant. The genes of a plant are what decide the differences for that plant. If one plant has 20 genes that are the same as a different plant those plants will have certain things that are similar about them. One of the species of plant mentioned in the reading was very interesting. The **males** genes make it produce a larger amount of smaller flowers while the females genes make it produce larger and longer lasting but fewer flowers. The genes of these plants will change the way they grow and survive depending on the **gender** of that plant."	"Gene between **males** and **females** cause differences. An example of a difference is shown with the sex chromosomes of the **genders**, with one being XX for **female** and XY for **male**. This leads to different things in the body (per gender) during processes. **Males** have testosterone and **females** have estrogen in their body during puberty and the endocrine system production. These hormones are unique per each **gender** and affect them in different ways. **Males** go through deepening voice, hair, and growth spurts just to name a few. **Girls** develop breasts, grow hair, and become more mature also to name a few. The genetics behind these differences is what explains how these sexes are different. To conclude, differences in **males** and **females** are caused by genes."	"Stereotypes do indeed cause difference between **male** and **female** humans, however, it is not any biological or genetic factors that really cause the differences. **Although women and men have different body types due to biological factors**, this does not mean their levels of intelligence are different. In fact, studies have shown that a **female** and **male** brain aren't very different and it is quite hard for scientists to tell whether a brain is **female** or **male**. And this is also the same for personalities. But in today's world, there are many stereotypes about **women**. An example would be that **women** are less intelligent than **men**. This stereotype leads **women** to actually believe they aren't smart enough and causes them to doubt they could work in a science or math field, which is considered more of a "**men's**" subject."
Sex language indicating conflation			"Stereotypes cause many differences between **males** and **females**, one being the stereotype that males are better at science and math. This assumption is flawed, as studies show that there is no distinguishable factor in **male** or **female** brains that affects their intelligence. This stereotype lowers **females'** self esteem, which causes them to question their intelligence. Instead of pursuing their desired career pathway, **females** will leave science and math pathways. The stereotype that **males** are better at science and math than **females** causes **females** to stop following a career pathway in the fields of science and math."

Table 6 (continued)

	Genetics of plant sex	Genetics of human sex	Refutational text
Gender language indicating conflation		"**Women** have two X chromosomes and **men** have one X chromosome and one Y chromosome, making it easier for them to have a disorder if the chromosome carries it."	"**Men** and **women** have very similar brains, but not all of the brains are the same. Stereotypes cause test anxiety for **women** and the fields of science with the most stereotypes, has less **women** in it."

discuss "men" and "women" (see example in Table 6). This is not surprising, considering that the question was posed to them using "male" and "female." Students in the refutational text condition were least likely to use only sex language in their responses (6.41%); however, some of these examples were representative of conflation. That said, there were students in the refutational text condition who used sex and gender nouns in their responses, without conflating. For instance:

> In our world, Male and Female is something that divides us. But it's not because of our genetics, it's because of our beliefs. People assume that Men get more PhD's in science because they are somehow born with the skill to succeed in that subject. When the truth is that Men don't usually do better than Women, Women just do worse. That's because of gender stereotyping. When a Women is told that Men do better on standardized tests, and Women do poorly, it stresses them out and gives them anxiety. Maybe they feel like there's no point if men are just going to beat them, or maybe they are trying so hard that they end up failing. The Stereotype Threat is something that changes what we associate with things, especially STEM.

This student was successful in attending to the genetics in the start of their response, using "Male and Female" and then moving to the social context referring to men and women. This example illustrates how a student, when provided with evidence, might construct a response that incorporates both sex and gender language to describe the phenomenon, without conflating terms. However, these examples were rare. Similarly, about a quarter (26.28%) of the responses in the refutational text group included exclusively gender nouns. Some of these responses included both biological and gender-linked content. While it was clear that many of the students in the refutational text condition were attempting to explain the lack of biological evidence along with the social reasons for gender disparities, these responses represent conflation.

5 Discussion and Implications

Students participating in the traditional plant and human sex treatments used gender language less often than students in the refutational text treatment. These instances of appropriate use of sex language are encouraging. However, 40% of the students in the genetics of human sex condition and 16% in the genetics of plant sex condition used gender language in their responses. The patterns associated with students who use gender language in their responses in the genetics of plant or human sex conditions are indicative of conflation. When we look at the responses more closely, students who use gender language to explain the genetics of sex in plants or humans may do so subconsciously. The plant and human readings did not include mention of gender concepts or vocabulary. This suggests that students in the plant and human sex conditions are using gender language that they have grown accustomed to associating with sex. Considering the close association between sex and gender in humans (Krieger 2003), we would expect conflation in the human treatment, but conflation in the plant treatment is particularly telling, considering that plants do not have gender. Students who used gender nouns in the genetics of plant sex treatment also used sex language suggesting that when students use sex and gender nouns in the plant condition, they may be doing so subconsciously. In other words, students are bringing in everyday language to help clarify their ideas, not necessarily to assign other human markers of gender to plants.

The text-based intervention that used sex only language may have assisted in helping *most* students construct responses without the use of gender language, but for some students the availability of prior knowledge about associations between sex and gender may have gotten in the way. Perhaps if these students had a prior experience, before the intervention that helped them to develop a concrete understanding of the difference between sex and gender, the availability would have been counteracted. Conversely, students in the refutational text group were significantly less likely to use sex language in their responses, which is consistent with the treatment they received (Donovan et al. 2019). However, these students did use gender nouns and sex nouns interchangeably and the usage was not always appropriate. Rather than seemingly unconscious conflation of terms, as we observed in the plant and human conditions, students in the refutational text condition attempted to use sex and gender language to *explain* differences between men and women and *refute* essentialist thinking connecting the idea that stereotyping occurs through essentialist beliefs (Rhodes et al. 2012). This supports the findings of the previous study, which established a link between traditional genetics of sex difference text and increased essentialist thinking (Donovan et al. 2019). While the student responses in the refutational text condition included conflation, it indicates that students may attempt to disentangle the concepts of sex and gender in the biology classroom, thus helping students to develop a stronger understanding of how social constructs can be falsely blamed on biological causes. Counteracting this thinking could be a valuable tool in decreasing essentialist thinking.

Students in the plant condition did not use words like man and woman to describe plants, reflecting the typical use of those words in communities of practice outside of the classroom. But the interchangeable use of sex and gender nouns indicates that not enough is being done in these students' biology curricula to make these distinctions clear. The conflation of sex and gender nouns in these students' responses indicates that they likely have little conceptual understanding of the difference between sex and gender. Gendered language where sex language would be more appropriate is common in textbooks (Beldecos et al. 1988; Bazzul and Sykes 2011; Bazzul 2013), as well as in academic writing across disciplines (Borna and White 2003; Johnson et al. 2009; Maccoby 1988), and in popular educational resources like Khan Academy (2020a, b). The use of gendered and sex language in textbooks and in student responses becomes much more significant when viewed through the lens of the availability heuristic (Tversky and Kahneman 1974). The availability heuristic posits that when information comes to mind (e.g., gender categories) at the suggestion of an associated phenomenon (e.g., genetics of sex difference), the recall of information about the associated idea (gender) is used to explain that phenomenon (genetics of sex difference), even if it is not a completely accurate association. Over time, students may associate sex and gender, particularly if they have never considered the difference between the two through either their lived experience or their experience in the classroom. When they attempt to recall information about the genetics of sex difference, they may apply the closely associated gender words, instead. The knowledge construction in the genetics of plant sex and genetics of human sex conditions did not refute the conflation of sex and gender. Unless the strong association between sex and gender is made explicit, they may continue to be conflated in student writing (van den Broek and Kendeou 2008).

Even though it could be argued that students are subconsciously conflating sex and gender nouns, there is good reason to be concerned about the impact this may have, if not clarified. The development of essentialist beliefs about the world is deeply entwined with the use of language and these findings are consistent with prior studies that connect the use of generic noun phrases with psychological essentialism (Walton and Banaji 2004; Reiter and Frank 2010; Rhodes et al. 2012). Language is not socioculturally neutral. Rather, the words that students use in the science classroom to make sense of the world are laden with the context of students' lived reality, and this reality is in turn shaped by the words that students use to make sense of the world (Vygotsky 1978; Lemke 1990; Gelman and Heyman 1999). As students move between communities of practice, special attention to appropriate language usage, especially when there are social implications, needs to be made more explicit. The use of everyday language can become problematic when instruction does not support students in explicitly unpacking the ideas and experiences that they associate with certain words. Take the example of the droplets of water on a soda can. By using the word sweat interchangeably with condensation, students may make sense of the phenomenon using the analogy of human sweat, constructing an understanding of the water droplets they see on the cold soda as having come from inside the can, and not from the water vapor present in the air. This is an unscientific belief about the world, shaped by the unscientific use of everyday language in the science classroom. While conflation of sweat and condensation can lead to unscientific beliefs about the nature of a soda can, it does not have a social consequence. However, conflation of biological sex and gender has been shown to engender unscientific essentialist beliefs about the nature of human difference that could manifest in sexism and transphobia (Gelman and Heyman 1999; Smiler and Gelman 2008; Haslam et al. 2000).

Across the curriculum, the use of everyday language is an important way to help students connect their lived experience to the science classroom. However, in the case of sex and gender language, the social implications of conflation outweigh the benefits of hybrid discourse in the classroom. The findings from Donovan and colleagues (2019) indicate that the traditional way that we present sex and gender could lead to essentialist ideas about the nature of differences between men and women. When combined with the findings reported here, it is apparent that conflation may impact students' understanding of both concepts of genetic attribution and the social world. The pattern that we observed in the refutational condition of students attempting to disentangle sex and gender to explain a social phenomenon and refute a biological explanation provides evidence that they were constructing a conceptual understanding incorporating this new information. But while the refutational text in this study may have helped students to make sense of the interactions implicit in sex and gender, it did not go far enough. When using gendered terms to refer to scientific phenomena, we must define these terms for students.

How can we help adolescents think about "sex" and "gender" in productive and respectful ways in science class? First, an explicit definition of gender as a social construct is essential. This is a complex issue that would be best to tackle with a broad consensus that considers the sociological, psychological, and scientific communities. During this discussion, illustrations of how sex and gender language (and likely imagery) is used in contemporary instructional materials, like we observed in a review of Kahn Academy, could aid in making this issue more visible. Similarly, highlighting examples of students' use of gender language in descriptions of sex-based phenomenon, as we point out above, is important in helping the community to address this issue. Simply taking the step to *notice* instances of conflation in our science classrooms could be an important step forward.

Second, the science education community must work on developing a more comprehensive and nuanced understanding of the historical interplay between science and society as it relates to concepts of biological sex and gender (Schiebinger 2013), and other social identities, including race (Donovan 2017), and the complex interactions between these constructs. Third, the science education community should develop a more comprehensive understanding of the contemporary implications of the problem of conflating sex and gender, including a better understanding of the research focused on essentialism, gender inequality, and how simple definitions of sex, sexuality, and gender lead to even more confusion about the human condition (Snyder and Broadway 2004). Finally, instructional materials developers must acknowledge the challenge and find new ways of integrating social problems into materials development and conduct comprehensive field-testing to ensure that instructional materials impact students in the ways they are intended. This effort should also include considering how to address professional learning for teachers. Without a concerted effort, across the science education community, the pervasiveness of this problem will not be addressed sufficiently.

6 Limitations and Directions for Future Research

First, in our attempt to standardize the way that we phrased the prompt in the refutational text condition to align with the way we asked it in the genetics of plant and human sex conditions, it is possible that students used a combination of sex and gender language to explain the social phenomenon in the refutational condition because they were primed to use both sex language and gender language. However, in the refutational condition, "female" and "male" were used to talk about human brains, and "women" and "men" were used to discuss the social phenomenon. Had we phrased the question using gender terms for the refutational condition, perhaps students would have used gender only language. However, for those students who discuss the biological phenomenon, this use of gender nouns constitutes conflation. Future studies could randomly assign students in the refutational condition so that half of the students receive a prompt that asks about differences between men and women and the other half responds to the male/female prompt. Second, because this study was an exploratory analysis of previously collected student responses, we did not have the opportunity to dig more deeply into the way that students think about the differences between sex and gender. Future studies should attend to the interplay between language and conceptual development of sex and gender, especially as they relate to genetics. Third, we have no way to know if our findings persisted beyond the duration of the study. Future studies should include a time-lag assessment to investigate the extent to which student conceptions persist over time. Finally, the refutational text that was used in this study was very short and would not be appropriate for a normal classroom context. Future materials development in this area should address the complexity of sex and gender comprehensively, be thoroughly field-tested, and revised prior to use in classrooms.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11191-020-00177-9.

Acknowledgments Thank you to Monica Weindling, Dennis Lee, and the anonymous reviewers for their careful reviews of this manuscript.

Authors' Contribution Molly Stuhlsatz co-conceptualized the study, the conflation hypothesis, and the theoretical framework. Stuhlsatz completed all statistical and qualitative analyses, and the writing of the initial manuscript. Zoë Buck Bracey co-conceptualized and contributed to the writing of the theoretical framework and contributed to the integration of reviewer feedback during the revision process. Brian Donovan contributed to the conceptualization of the study and the conflation hypothesis and consulted on statistical analyses.

Funding This material is based upon work supported by the National Science Foundation under Grant No. 1660985. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Arthur, A. E., Bigler, R. S., Liben, L. S., Gelman, S. A., & Ruble, D. N. (2008). Gender stereotyping and prejudice in young children. In *Intergroup Attitudes and Relations in Childhood through Adulthood* (Vol. 66, pp. 1072–1087).

Ash, D. (2008). Thematic continuities: talking and thinking about adaptation in a socially complex classroom. *Journal of Research in Science Teaching, 45*(1), 1–30.

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology, 42*(2), 228–235.

Bazzul, J. (2013). *How discourses of biology textbooks work to constitute subjectivity: From the ethical to the colonial*. Doctoral Dissertation. Univeristy of Toronto.

Bazzul, J., & Sykes, H. (2011). The secret identity of a biology textbook: straight and naturally sexed. *Cultural Studies of Science Education, 6*(2), 265–286.

Beldecos, A., Bailey, S., Gilbert, S., Hicks, K., Kenschaft, L., Niemczyk, N., & Wedel, A. (1988). The importance of feminist critique for contemporary cell biology. *Hypatia, 3*(1), 61–76.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 27*, 389–391.

Bianchini, J. (1993). *The high school biology textbook: a changing mosaic of gender, science, and purpose* (p. 34). Presented at the Annual Meeting of the American Educational Research Asociation, Atlanta, GA.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science, 16*(3), 162–166.

Bond, A. (2013). *Sex, gender, and when "modern" sensibilities go wrong*. Retrieved December 10, 2018 from https://labandfield.wordpress.com/2013/04/06/sex-gender-and-when-modern-sensibilities-go-wrong/.

Borges, R. M. (1998). Gender in plants. *Resonance, 3*(11), 30–39.

Borna, S., & White, G. (2003). "Sex" and "gender": two confused and confusing concepts in the "women in corporate management" literature. *Journal of Business Ethics, 47*(2), 89–99.

Brescoll, V., & LaFrance, M. (2004). The correlates and consequences of newspaper reports of research on sex differences. *Psychological Science, 15*(8), 515–520.

Brown, B. A., & Ryoo, K. (2008). Teaching science as a language: a "content-first" approach to science teaching. *Journal of Research in Science Teaching, 45*(5), 529–553.

BSCS. (2007). *BSCS science: an inquiry approach level 2 Student Edition*. Dubuque: Kendall Hunt Publishing.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676.

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: sociocultural and biological considerations. *Psychological Bulletin, 135*(2), 218–261.

Ching, B. H. H., & Xu, J. T. (2018). The effects of gender neuroessentialism on transprejudice: An experimental study. *Sex Roles, 78*(3–4), 228–241.

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*(05), 461–480.

Clément, P., & Castéra, J. (2013). Multiple representations of human genetics in biology textbooks. In *Multiple representations in biological education* (pp. 147–163). Dordrecht: Springer.

Coleman, J. M., & Hong, Y.-Y. (2008). Beyond nature and nurture: the influence of lay gender theories on self-stereotyping. *Self and Identity, 7*(1), 34–53.

Coley, R. J. (2001). *Differences in the gender gap: comparisons across racial/ethnic groups in education and work*. Educational Testing Service.

Compleat Lexical Tutor. (2017). Retrieved July 12, 2017, from http://www.lextutor.ca/

conflation. (2018). In *Merriam-Webster.com*. Retrieved December 2, 2018, from https://www.merriam-webster.com/dictionary/conflation.

De Castro, P., Heidari, S., & Babor, T. F. (2016). Sex and gender equity in research (SAGER): reporting guidelines as a framework of innovation for an equitable approach to gender medicine. *Annali dell'Istituto Superiore di Sanità, 52*(2), 154–157.

Del Boca, F. K. (2016). Addressing sex and gender inequities in scientific research and publishing. *Addiction, 111*(8), 1323–1325.

Demoulin, S., Leyens, J. P., & Yzerbyt, V. (2006). Lay theories of essentialism. *Group Processes & Intergroup Relations, 9*(1), 25–42.

Diesendruck, G., Goldfein-Elbaz, R., Rhodes, M., Gelman, S., & Neumark, N. (2013). Cross-cultural differences in children's beliefs about the objectivity of social categories. *Child Development, 84*(6), 1906–1917. https://doi.org/10.1111/cdev.12108.

Donovan, B. M. (2017). Learned inequality: Racial labels in the biology curriculum can affect the development of racial prejudice. *Journal of Research in Science Teaching, 54*(3), 379–411.

Donovan, B. M., Stuhlsatz, M., Edelson, D. C., & Buck Bracey, Z. (2019). Gendered genetics: How reading about the genetic basis of sex differences in biology textbooks could affect beliefs associated with science gender disparities. *Science Education, 103*, 719–749.

Dweck, C. S. (2008). *Mindset: The new psychology of success*. Manhattan, NY: Random House Digital, Inc.

Eidson, R. C., & Coley, J. D. (2014). Not so fast: reassessing gender essentialism in young adults. *Journal of Cognition and Development, 15*(2), 382–392.

Eveleth, R. (2014). *Some scientific journals are still confusing sex and gender*. Smithsonian. Retrieved December 5, 2018 from: https://www.smithsonianmag.com/smart-news/some-scientific-journals-are-still-confusing-sex-and-gender-180949424/

Fausto-Sterling, A. (2000). Sexing the body: Gender politics and the construction of sexuality. New York: Basic Books.

Flynn, J. R. (1999). Searching for justice: the discovery of IQ gains over time. *American Psychologist, 54*(1), 5–20.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: the effects of lexicalization on children's inferences about social categories. *Psychological Science, 10*(6), 489–493.

Gelman, S. A., & Taylor, M. G. (2000). Gender essentialism in cognitive development. In *Toward a feminist developmental psychology* (pp. 169–190).

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: the case of gender. *Child Development*, *57*(2), 396–404.

Glasser, H. M., & Smith III, J. P. (2008). On the vague meaning of "gender" in education research: The problem, its sources, and recommendations for practice. *Educational Researcher, 37*(6), 343–350.

Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: individual traits or repertoires of practice. *Educational Researcher, 32*(5), 19–25.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*(1), 113–127.

Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology, 41*(1), 87–100.

Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review, 1*(1), 2.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X.

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: five challenges to the gender binary. *American Psychologist*. Advance online publication.

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., et al. (2015). Sex beyond the genitalia: the human brain mosaic. *Proceedings of the National Academy of Sciences, 112*(50), 15468–15473. https://doi.org/10.1073/pnas.1509654112.

Johnson, J. L., Greaves, L., & Repta, R. (2009). Better science with sex and gender: facilitating the use of a sex and gender-based analysis in health research. *International Journal for Equity in Health, 8*(1), 14.

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

Khan Academy. (2020a) *Sex Linked Traits (video).* Accessed March 10, 2020. https://www.khanacademy.org/science/high-school-biology/hs-classical-genetics/hs-sex-linkage/v/sex-linked-traits.

Khan Academy. (2020b) *X-inactivation*. Accessed March 10, 2020. https://www.khanacademy.org/science/high-school-biology/hs-classical-genetics/hs-sex-linkage/a/x-inactivation

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence, 35*(5–6), 790–816.

Krieger, N. (2003). Genders, sexes, and health: what are the connections—and why does it matter? *International Journal of Epidemiology, 32*(4), 652–657.

Krippendorff, K. (2018). *Content analysis: an introduction to its methodology*. Sage Publications.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lemke, J. L. (1990). *Talking science: language, learning, and values* (p. 07648). Norwood: Ablex Publishing Corporation (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1).

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

Lorber, J. (1993). Believing is seeing: Biology as ideology. G*ender & Society, 7*(4), 568–581.

Maccoby, E. E. (1988). Gender as a social category. *Developmental Psychology, 24*(6), 755.

Martin, E. (1991). The egg and the sperm: how science has constructed a romance based on stereotypical male-female roles. *Signs: Journal of Women in Culture and Society, 16*, 485–501.

Muehlenhard, C. L., & Peterson, Z. D. (2011). Distinguishing between sex and gender: history, current conceptualizations, and implications. *Sex Roles, 64*(11–12), 791–803.

Nehm, R. H., & Young, R. (2008). "Sex hormones" in secondary school biology textbooks. *Science & Education, 17*(10), 1175–1190.

Pardue, M. L., & Wizemann, T. M. (Eds.). (2001). *Exploring the biological contributions to human health: does sex matter?* National Academies Press.

Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science, 16*(4), 202–206. https://doi.org/10.1111/j.1467-8721.2007.00504.x.

Reiter, N., & Frank, A. (2010). Identifying generic noun phrases. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 40–49). Uppsala 11–16 July 2010.

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology, 59*(3), 244–274. https://doi.org/10.1016/j.cogpsych.2009.05.001.

Rhodes, M., Leslie, S. J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, 109*(34), 13526–13531.

Richardson, S. S. (2013). *Sex itself: the search for male and female in the human genome*. University of Chicago Press.

Sainani, K. L. (2011). Understanding odds ratios. *Pm&r, 3*(3), 263–267.

Schiebinger, L. L. (2013). *Nature's body: gender in the making of modern science*. Rutgers University Press.

Smiler, A. P., & Gelman, S. A. (2008). Determinants of gender essentialism in college students. *Sex Roles, 58*(11–12), 864–874.

Snyder, V. L., & Broadway, F. S. (2004). Queering high school biology textbooks. *Journal of Research in Science Teaching, 41*(6), 617–636.

Spanier, B. (1995). *Im/partial science: gender ideology in molecular biology*. Bloomington: Indiana University Press.

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.

Taylor, M. G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development, 67*(4), 1555–1571.

Tversky, A., & Kahneman, D. (1974). Judgements under uncertainty: heuristics and biases. *Science, 185*(415), 1124–1131.

Unger, R. K., & Crawford, M. (1993). Sex and gender—the troubled relationship between terms and concepts. *Psychological Science, 4*(2), 122–124.

Valdes, F. (1996). Unpacking Hetero-Patriarchy: Tracing the Conflation of Sex, Gender & Sexual Orientation to Its Origins. *Yale Journal of Law & the Humanities, 8*(1), 161–211.

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: the role of co-activation in confronting misconceptions. *Applied Cognitive Psychology, 22*, 335–351.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. https://doi.org/10.1037/a0036620.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children, 23*(3), 34–41.

Walton, G. M., & Banaji, M. R. (2004). Being what you say: the effect of essentialist linguistic labels on preferences. *Social Cognition, 22*(2), 193–213.

Warren, B., Ogonowski, M., & Pothier, S. (2005). "Everyday" and "scientific:" rethinking dichotomies in modes of thinking in science learning. In *Everyday matters in science and mathematics: Studies of complex classroom events*. Routledge.

Weiss, M. J., Lockwood, J. R., & McCaffrey, D. F. (2016). Estimating the standard error of the impact estimator in individually randomized trials with clustering. *Journal of Research on Educational Effectiveness, 9*(3), 421–444.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society, 1*(2), 125–151.

Yzerbyt, V. Y., & Rocher, S. (2002). Subjective essentialism and the emergence of stereotypes. In: C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes as explanations: The formation of meaningful beliefs about social groups* (p. 38–66). Cambridge: Cambridge University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.