Gender Stereotypes and Interventions Targeting Gender Stereotypes in STEM

Karly Drabot

Downing College

Department of Psychology

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

October 2022

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any work that I have submitted, or is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the Faculty of Biology Degree Committee.

Gender Stereotypes and Interventions Targeting Gender Stereotypes in STEM

Karly Drabot

SUMMARY

Women continue to be underrepresented in STEM (science, technology, engineering, and mathematics), despite scientific evidence indicating STEM aptitude is not gender differentiated. Various policy and institutional efforts have been made to improve gender diversity in STEM, yet little progress has been made. As psychological research indicates the gender gap in STEM is perpetuated by gender stereotypes, psychological interventions targeting gender stereotypes may be an important and effective approach. Thus, this doctoral research (1) examines the effectiveness of novel interventions designed to increase discernment of gender stereotypes in STEM and (2) aims to improve understanding and measurement of current gender stereotypes.

Using a multi-dimensional approach to gender-STEM stereotyping, I explore my research questions through online studies (Chapters 2, 3, and 5) as well as a field study with university engineering students (Chapter 4). Specifically, I designed and evaluated theory-driven psychological interventions based on optimised debunking and inoculation theory (Chapter 2, 3, and 4). Additionally, I examined the psychological mechanisms involved in improving gender-STEM stereotype discernment (Chapter 3 and Chapter 4). Finally, I address methodological issues in gender-trait stereotype research (Chapter 5). As the conclusions we draw in research can only be as robust as the methods we use to quantify them, I aimed to improve how we conceptualise, define, and measure gender stereotyping and the psychological processes involved.

To the best of my knowledge, this is the first research programme to examine inoculation theory in the context of stereotyping. Findings demonstrate some effectiveness of the interventions in improving discernment of gender stereotypes in STEM, but not endorsement of gender-trait stereotypes, application of gender stereotypes in decision-making, or gender-related attitudes. Additionally, participants in the inoculation-intervention conditions did not demonstrate levels of discernment to the same extent as those in the debunking-intervention conditions. This suggests that exposure to persuasive misinformation about gender stereotypes after the intervention treatment neutralised some of the treatment effects. However, findings show support for the cognitive and affective

mechanisms theorised in inoculation theory (counterarguing and threat), but do not show clear support for the cognitive and affective mechanisms typically associated with stereotype interventions (logical thinking, perspective taking, and empathic concern). Finally, findings indicate that current conceptualisation and measurement of gender-trait stereotypes need to be updated. While the female-communal stereotype persists, it appears the male-agentic stereotype may have reduced in size and shifted in content. Implications, limitations, and suggested future directions are discussed.

TABLE OF CONTENTS

CHAPTER ONE – GENERAL INTRODUCTION

"Wrongful gender stereotyping is a pervasive human rights violation."

UN Office of the High Commissioner for Human Rights (2013, p. 2)

Despite no evidence for an innate gender difference in STEM aptitude (Eagly, 2020; Ellemers, 2018; Hyde, 2014), women remain underrepresented. Law and policy have been in place for decades to protect against discrimination on the basis of gender (see European Commission, 2020; Kaufman, 2021), yet gender disparities that disadvantage women continue to persist in employment rates, pay, and positions (European Union, 2017). The unequal representation and treatment of women and men in STEM has prompted the development of interventions designed to reduce this gender gap. The types of interventions have been varied, ranging from the implementation of educational and mentoring programs at both early stages of education and university-level (Schneider, 2001), to changes in course structure (Brewe et al., 2010), as well as the introduction of workplace initiatives to improve hiring, retention, and promotion of women in these fields (International Labour Organization, 2020). While these efforts have seen some progress in reducing the gender gap in biological fields of STEM, gender disparities in the fields of mathematics, physics, engineering, and computer science persist (National Science Board, 2016).

Gender stereotypes have been identified as playing a key role in perpetuating the gender gap in STEM fields (see Mastari et al., 2019; van Veelen & Derks, 2021; Carli et al., 2016; Cheryan et al., 2015; Moss-Racusin et al., 2012). Gender stereotypes are oversimplified and overgeneralized beliefs about the attributes, characteristics, and roles of each gender (Cook & Cusack, 2011). They are based on the binary assumption that men and women possess oppositional traits, attitudes, and behaviours. Thus, gender stereotypes can lead to inaccurate and unjust beliefs about the abilities, interests, and attributes of others based solely on their gender (Cook & Cusack, 2011). The application of gender stereotypes can result in prejudice and discrimination towards others based on their gender (Council of Europe, 2022; Spence et al., 1975). For clarity purposes, prejudice is a preconceived, often negative, attitude towards individuals based on their group membership, while discrimination is the behaviour, decisions, and treatment of others guided by unjustified beliefs and

attitudes. Therefore, gender stereotypes can contribute to the gender gap in STEM through prejudice and discrimination against women in STEM (or in favour of men in STEM).

Annually, billions of dollars are invested in programmes to reduce prejudice and discrimination (Hansen, 2003; Jayne & Dipboye, 2004). However, the effectiveness of these programmes and, more specifically, which mechanisms work well and which don't, has not been clearly identified (Paluck, 2006; Paluck & Green, 2009; Anand & Winters, 2008). Paluck and Green (2009) summarize their review on prejudice and discrimination reduction procedures with a call for both evidence and theory development. As prejudice and discrimination largely result from negative stereotyped beliefs, efforts may be well spent targeting stereotype endorsement and application. This approach would aim to address the root of the issue. Indeed, Cook and Cusack (2011) conclude, "in order to eliminate all forms of discrimination against women, and indeed to eliminate other violations of women's rights whether or not they constitute a form of discrimination, greater priority and thought need to be given to the elimination of wrongful gender stereotyping" (*p*. 3).

A Note on Studying Gender Stereotypes

Public perception of psychological differences between men and women are often exaggerated and incorrectly linked to inherent aptitude. Take, for example, the widely known book and common phrase, *Men are from Mars, Women are from Venus*. As implied by the book title, gender differences are commonly perceived by society to be innate, immutable differences. On the other hand, another camp of people believe that the notion of any differences between men and women is a patriarchal social illusion. Despite the widely endorsed scientific perspective that gender differences are shaped by biological, social, and cognitive factors, as well as their interactions, researchers investigating gender differences also often favour one view (Eagly & Wood, 2013). Designing and testing models of the multi-faceted, intricate relationships between nature, nurture and self-socialization is challenging, so exploring only one camp allows researchers to focus on providing clear, concise explanations of gendered findings. This separation can facilitate an "or" versus "and" discourse, which perpetuates siloed and polarised positions. This thesis does not aim to compare or evaluate the different approaches to studying gender differences, but rather acknowledge that there are several guiding forces in the development of and perception of gender differences and, consequently, gender stereotypes.

Cognitive Development of Gender Stereotypes

Globally, gender persists as a salient social category (Fiske, 2017; Leaper & Bigler, 2011; Ruble et al., 2006). On a daily basis, people are grouped by gender: in classrooms and sports teams, group activities and social organizations, washrooms and retail stores, and so on. Categorizing people into social groups, like gender, inflates the perception that group members are similar to one another (Quattrone & Jones, 1980). The assumption that members of a social group share the same characteristics and abilities is the cognitive foundation for stereotyping.

Much of stereotype and prejudice literature lies in the theoretical foundation of ingroup bias and outgroup favouritism, resulting in an "us-vs-them" mentality. The outgroup homogeneity effect "implies that outgroup members are perceived to be more similar than ingroup members" (p. 159, Ishii & Kitayama, 2010). For example, a man viewing all women as being more similar to each other than men are to each other. This effect has been demonstrated in numerous studies (e.g., Judd & Park, 1988; Linville et al., 1989; Park & Rothbart, 1982). However, when it comes to one's own stereotyped social identities, people tend to perceive not only outgroup homogeneity, but also ingroup homogeneity - the ingroup homogeneity effect (Rubin & Badea, 2007; Simon & Pettigrew, 1990; Kelly, 1989). This helps to explain why female and male individuals also apply gender stereotypes to their own gender group (Heilman, 2012). Indeed, Goldberg's (1968) influential study found that women, like men, also rated articles by male authors more favourably than female authors, even when the article content was identical (author name was either John Kay or Joan Kay).

Formation and Maintenance

The cognitive psychology theory, Schema Theory, can help to explain the formation of gender stereotypes. A schema is a cognitive template for organising knowledge and interpreting information. The concept was first introduced by Plato and Aristotle (see Marshall, 1995), but is more extensively used by and attributed to Kant (1929). In the field of cognitive psychology, schema theory explains how individuals manage constant inundation of external stimuli, as it would otherwise exceed processing capacities; a primary function of schemas is to select a subset of this stimuli that can be easily and quickly processed according to previously-integrated knowledge structures (Alba & Hasher, 1983; Cantor & Mischel, 1977). Based on past experiences, schemas guide understanding, decision-making, and behaviour.

In other words, schemas help us to simplify the world around us, reduce our cognitive load, and take action. This also means they influence the information we pay attention to and how we interpret information. They help us to make generalisations so that the world around us is less cognitively demanding. This can be incredibly helpful for processing object-related information, but, in the context of human-related information, these cognitive short-cuts can be harmfully inaccurate and consequential. Thus, cognitive psychologists use schemes and schema theory to help explain the formation of gender stereotypes.

More specifically, Gender Schema Theory, originally proposed by Bem (1981), describes how we organise knowledge about men and women into cognitive schemes. These schemes include the attributes, interests, traits, and behaviours associated with the prototype for each gender. Gender schematic processing, then, involves encoding and organising information about ourselves, others, and the world around us based on gender-linked associations (Bem, 1981). This makes both encoding and recalling information, based on gender, to be quicker and more efficient. It also means that when a perceiver does not have access to all relevant information, gender schemas are used to fill in the gaps with information that is consistent with the content of the scheme (Payne et al., 1987).

Gender schemes typically develop according to the culture's definitions of what is masculine and what is feminine, or what is appropriate (and inappropriate) for each gender (Bem, 1981). Thus, gender schemas are salient and used not only in information processing but also in decision-making (Bem, 1981). However, scientific utility has yet to be established for gender schema theory (Spence & Helmreich, 1981; Snell, 2002) and schema theory more broadly (Thorndyke & Yekovich, 1979). For example, Bem (1981) hypothesized and found that gender-typical individuals engage in greater gender schematic processing than androgynous (high in both masculinity and femininity) and undifferentiated (low in both masculinity and femininity) individuals, but this finding has not been replicated (see Markus et al., 1982; Payne et al., 1987). While schema theories can be useful for descriptive purposes, their predictive capacity is limited due to the lack of testable assumptions and constraints. More testable, social theories, Social Role Theory and Role Congruity Theory, will be discussed below.

Content of Gender Stereotypes

Ultimately, gender stereotypes are important to study because many of the attributes that are believed to be gender differentiated are also differentially valued in societies. Indeed, no country has reached gender equality (UNDP, 2020) or equity. Eisend (2019) suggests

gender stereotypes can be organised into four categories of attributes: personality traits, occupational status, physical characteristics, roles/behaviours. For the purposes of this thesis, I will discuss theory and evidence for two of these categories: personality traits and occupational status.

Gender-Trait Stereotypes

Two fundamental and independent dimensions underlie social judgments of the self, others, and groups (Abele et al., 2008; Abele & Wojciszke, 2007; Paulhus & Trapnell, 2008; Beauvois & Dubois, 2009; Fiske et al., 2002; Judd et al., 2005; Entringer et al., 2021); these were first introduced as "agency" and "communion" by Bakan in 1966 (Bakan, 1966). Agency broadly characterises traits to "get ahead" or a concern for oneself, while communion involves traits to "get along" or a concern for others (Eagly & Karau, 2002; Dickman & Eagly, 2000; Abele & Wojcizske, 2007). These two trait dimensions have more recently been termed the "Big Two" (Abele & Wojciszke, 2014; Paulhus & Trapnell, 2008) given their fundamental role in personality as well as social, cross-cultural, and intergroup psychology (e.g., Bruckmüller & Abele, 2013; Gebauer et al., 2012; Abele & Wojciszke, 2014; Paulhus & Trapnell, 2008). Researchers in different fields sometimes refer to agency and communion as instrumentality and expressivity (e.g., Stake et al., 1996; Colley et al., 2009), or competence and warmth (e.g., Cuddy et al., 2008; Fiske et al., 2002; Oldmeadow & Fiske, 2007), respectively. For the purposes of this thesis, the terms 'agency' and 'communality' will be used as they are widely used and understood across different fields of psychology (see Abele & Wojcizske, 2007).

Traditionally, agency and communality have been conceptualised as oppositional, bipolar traits, with the former being male-typed and the latter being female-typed (Deaux & Lewis, 1984; Berndt & Heller, 1986). Accordingly, many researchers refer to (a) 'agency' and 'masculinity' and (b) 'communality' and 'femininity' synonymously (e.g., Boldizar, 1991; Egan & Perry, 2001; Hall & Halberstadt, 1980; Spence, 1984). Communality refers to an "emotional, interpersonal orientation", a core feature of female stereotypes, and agency refers to an "assertive, instrumental orientation", a core feature of male stereotypes (Conway & Vartanian, 2000). Indeed, from the age of about three, boys are rated as more agentic than girls and girls are rated as more communal than boys (Gagne et al., 2013), and this gender stereotype persists across the lifespan (Canetto et al., 1995).

It has been argued that the principles of communality and agency in the context of gender stereotypes are grounded in the difference in status between genders (Kurt & Buckley, 1970; Carrier et al., 2014; Conway et al., 1996). Since men generally have a higher status than women (Lips, 1991), this allows them to behave in a more assertive and competitive way, while women, who generally have a lower status than men, may be required to act in a more submissive, interpersonal, and cooperative manner. Indeed, research shows perceived social status predicts perceived competence (Fiske et al., 2002). Likewise, evaluations of groups as warm or communal tend to implicitly infer the group's low competence (Judd et al., 2005; Fiske et al., 2002). Therefore, gender-trait stereotypes are intricately linked with status and power of gender in society.

Accuracy of Gender-Trait Stereotypes

Ryan (2003) suggests that in order to investigate and understand the process of stereotyping, researchers should not only consider stereotypes from the perspective of their associated negative consequences and biases, but also the possibility that stereotypes may, on occasion, be accurate and useful. However, stereotype accuracy research, which is aimed at determining the actual characteristics of a group, has methodological limitations. One limitation is the absence of an appropriate accuracy criterion, which is a result of research relying on self-reporting by members of the group in question. This can introduce sources of bias, such as self-selection sample bias and social desirability bias (Ryan et al., 1996, Schwarz, 1999). Assessing the accuracy of gender-trait stereotypes using men's and women's 'true' levels of agency and communality is dependent on their self-ratings. Perhaps men overestimate their own agentic traits and women overestimate their communal traits because there are social penalties for violating gender stereotypes - and social rewards for adhering to them (e.g., social desirability bias). This would exaggerate the accuracy of stereotype content. Alternatively, men and women may underreport their gender-stereotypic traits in an effort to push back against traditional gender expectations and appear to not conform to gender stereotypes. This would underestimate the accuracy of stereotype content. Thus, self-evaluations may not sufficiently assess the accuracy of gender stereotypes.

Accuracy research is also limited by the fact that gender stereotypes can change across cultures (Connolly et al., 2020; Sachdev, 2018; Costa et al., 2001; Miner et al., 2018) and over time (Charlesworth & Banaji, 2019, 2021; Twenge, 1997, 2001; Donnelly & Twenge, 2017). Therefore, if it is a moving target, accuracy is not a stable or static metric. Additionally, measurement of the characteristics of a group does not provide insight into, or

take into account, factors that may influence such characteristics. For example, the affirmation that women are more communal or less agentic than men does not provide consideration for the fact that these attributes are influenced and guided by social roles and social rewards (Eagly & Steffen, 1984). Thus, how researchers evaluate stereotype accuracy is an important methodological consideration.

Previous research suggests individuals' accuracy in rating gender differences on traits is low, $r = 0.43$ (as calculated by meta-analyses correlating perceived gender differences with self-reported gender differences on 77 traits; Hall & Carter, 1999). Beliefs about gender differences in traits are much larger than self-reported gender differences in traits (Martin, 1987; Krueger et al., 2003; Hall & Carter, 1999), suggesting the role that gender stereotypes play in perceptions of others (Martin & Ruble, 2004). Even when measured explicitly, most trait differences between men and women are small to moderate (Weisberg et al., 2011). Further, when personality traits are measured implicitly, the gender differences are three times smaller than when measured explicitly (Vianello et al., 2013). This suggests gender stereotypes also influence perceptions of oneself. As a result, our beliefs about gender differences in traits are likely to be inflated due to overlearned stereotypes.

Male-Agentic Stereotype

Although past literature has discussed an evidenced gender difference in agentic traits, as measured by personality questionnaires (Boldizar, 1991; Hall & Halberstadt, 1980), these findings appear to be outdated as they are becoming increasingly inconsistent (e.g., Hentschel et al., 2019). Two meta-analyses found that the gender difference in agency has been shrinking over time (Twenge, 1997; Twenge et al., 2001), with recent studies showing no gender difference in agency (Egan & Perry, 2001; Ramsey, 2017). Indeed, a more recent meta-analysis found that women's self-reported agentic traits increased (d = .23) between 1974 and 1993, and stabilised between 1993-2012 (Donnelly & Twenge; 2017). However, in a comparison analysis of gender stereotypes between 1983 and 2014, Haines and colleagues (2016) found that perceived differences between men's and women's attributes have remained strong and not significantly decreased since 1983. Perhaps, then, agency is no longer a gendered trait, but a gender *stereotyped* trait.

The association of agentic traits with men and not women is thought to explain, in part, the underrepresentation of women in male-dominated positions (Heilmen et al., 2001; Eagly & Karau, 2002). Indeed, many occupations are guided by gender (Hines, 2015).

Someone perceived to have agentic traits is expected to also possess an agentic social role (Deaux and Lewis, 1984; Berndt and Heller, 1986). Broadly, the agentic male stereotype relates to roles and careers of self-interest, whereas the communal female stereotype relates to roles and careers of other-interest (Eagly & Karau, 2002; Dickman & Eagly, 2000; Abele & Wojcizske, 2007). Thus, women in male-dominated fields face a double-bind - either be perceived as competent for the role but unlikeable, or likeable but not competent for the role (Heilmen et al., 2001; Eagly & Karau, 2002; Ellemers, 2018; Rudman & Phelan, 2008).

Gender-Occupation Stereotypes

Gender stereotypes, and the perception of gender differences in agency and communality, are intricately linked with social factors such as different occupations and roles between genders (Echabe, 2010; Eagly & Wood, 2012; Wood & Eagly, 2010; Ridgeway & Erickson, 2000; Jackman, 1994; Glick & Fiske, 1999). Men have traditionally occupied paid employment occupations, and are consequently described with agentic traits, such as "competent, ambitious, and determined" (Eagly & Wood, 2012). Women, on the other hand, have traditionally carried out domestic, care-taking roles and are therefore identified with communal traits, such as "warm, caring, and friendly" (Eagly & Wood, 2012). Conversely, since women have moved into paid employment, they have been perceived as more agentic by both men and women (Diekman & Eagly, 2000) and themselves (Spence & Buckner, 2000). Therefore, gender stereotypes are intricately linked with gender roles in society.

Social Role Theory (Eagly, 1987) proposes that societal roles and responsibilities are assigned based on gender (Eagly & Wood, 1991; Eagly & Wood, 1999; Eagly et al., 2000; Wood & Eagly, 2002). Thus, occupational differences between men and women appear largely as a result of longstanding gender norms. This division of labour in patriarchal systems results in gender inequalities, as men's roles are associated with higher status and privileges (Wood & Eagly, 2010; Eagly & Wood, 2012). Social role theory is well evidenced, particularly in perceived leadership aptitude of men versus women (see Eagly et al., 2000 for review). Therefore, gender-occupation stereotypes, like gender-trait stereotypes, are intricately linked with status and power of gender in society.

Effects of Gender Stereotypes

Role congruity theory, coined by Eagly and Karau (2002), extends social role theory by proposing that a perceived incongruity between gender role (e.g. female) and other roles

(e.g., engineer) leads to prejudice and discrimination. In other words, the greater the perceived overlap between a person's gender (and associated stereotypes) and a job's characteristics, the greater the perceived competence in that role. Likewise, the lesser the overlap, the greater the prejudice and discrimination.

Therefore, the perceived lack of fit between agentic roles and communal applicants can result in a negative evaluative bias that prevents women from attaining agentic positions (Lack-of-Fit Model; Heilman, 1983). Evidence supporting the Lack-of-Fit Model shows that holding an implicit female-communal stereotype results in hiring discrimination (Rudman & Glick, 1999). In this context, although women who present as agentic are perceived as more competent, the violation of the "female = communal" stereotype leads to negative evaluations of them, such as the perception that agentic women are less socially capable (Rudman & Glick, 1999). Further, independent of gender, communal applicants are rated as less hireable than agentic applicants for agentic-typed roles (Rudman & Glick, 1999). Therefore, women face a double-bind situation: present as agentic, and be perceived as less likeable, or present as communal and be perceived less hireable (Rudman, 1998). Meanwhile, for men, when they present as agentic, they are seen as more hireable and more likeable because this is compatible with their gender stereotype.

The negative evaluation for violating the stereotype that women are communal (and therefore not agentic) can result in a backlash effect (Rudman, 1998). The backlash effect refers to retributions experienced when one acts in a counter stereotypical way. For women, this often occurs in the workplace due to the association of female-family/male-career and female-arts/male-science. To be perceived as agentic (competent), women risk the reprisal of violating their gender stereotype, such as decreased sociability, likeability and, ultimately, reduced hireability (Rudman, 1998). Agentic men, like agentic women, are also not seen as particularly nice (Eagly & Mladinic, 1989), but this does not violate a prescriptive (communal) stereotype for men. As such, when male leaders and female leaders behave in identical ways (Eagly et al., 1992) or are presented in identical ways (Rudman & Glick, 1999), men are judged less harshly than their identical woman counterpart.

Taken together, it can be more challenging for women to enter and succeed in male-dominated domains due to the interplay of gender-trait stereotypes and gender-related roles and careers in society.

Gender Stereotypes in STEM

Historical overrepresentation of men in STEM fields has produced a cognitive association between science and male. There is substantial overlap between traits associated with STEM and traits associated with men – both are believed to be high in agency. Indeed, science faculty and students (both men and women) rate agentic traits as more important than communal traits for success in scientific careers (Ramsey, 2017) and show similar levels of gender bias against women in science (Moss-Racusin et al., 2012). Given that men are stereotyped as agentic and women are stereotyped as communal, the perceived importance of these traits for certain jobs and roles may contribute to the gender disparities we see in STEM. The stereotyped perception that 'STEM=agentic' and 'agentic=male' makes it challenging for women to be perceived as equally suitable for STEM.

In the United States, statistics show that women now make up 50% of workers in STEM (Pew Research Center, 2021). While it may appear as though gender equality has been achieved, this percentage is carried by the overrepresentation of women in healthcare roles. Indeed, women make up 74% of the health-related workforce and only 25% in computer science and 15% in engineering. This is not surprising, given that healthcare roles are traditionally female typed (e.g., communal, caring roles).

Indeed, research shows that countries with greater levels of gender-STEM stereotypes have larger gender gaps in STEM (Gygax et al., 2016; Nosek et al., 2009). The fact that gender disparity in STEM varies cross-culturally provides further evidence of the important role societal roles and attitudes play. Although these cross-cultural findings are only correlational, experimental evidence also demonstrates that gender stereotypes significantly contribute to the gender gap in STEM (Glick et al., 1988; Heilman, 1984; Isaac et al., 2009; Moss-Racusin et al., 2012; Rudman et al., 2012).

Among women who pursue careers in STEM, the inaccurate stereotype that men are better suited for STEM contributes to a hiring bias (Glick et al., 1988; Heilman, 1984; Isaac et al., 2009; Moss-Racusin et al., 2012). Steinpreis et al. (1999) found that both male and female academics were more likely to vote to hire a male applicant over a female applicant with the exact same curriculum vitae, which came from a real-life scientist. Moss-Racusin and colleagues (2012) replicated these findings with science faculty, and also found that the male applicant was rated as significantly more competent, hireable, and mentorable, and offered a higher starting salary than the identical female applicant. Similarly, a more recent study by Reuben et al. (2014) found that both men and women were twice as likely to hire a man for a math-based job, as well as offer a higher starting salary for men than women

(Moss-Racusin et al., 2012; Deloitte, 2016). This is in line with role congruity theory (Eagly & Karau, 2002), which proposes that a perceived incongruity between gender roles (e.g. female) and other roles (e.g., scientist) leads to prejudice and discrimination.

When women enter STEM fields, they are often viewed less favourably than men and gender-conforming women, because they violate their gender's prescriptive stereotype that women should be communal, not agentic (Rudman & Glick, 2002). The "double-bind" that women in STEM face - needing to be agentic to be considered fit for the role, but also needing to not express agency in order to avoid backlash - is illustrated in the Supreme Court case of Ann Hopkins v Price Waterhouse. Ann Hopkins was a lawyer who, in 1982, was denied a promotion at her accounting firm despite an outstanding performance record. In her rejection, she was not advised to work harder, but to be more feminine with her make-up, style, walk, and talk. On the basis of gender discrimination, she sued the firm and won. "At each level of decision, the courts recognized the role that gender stereotypes had played in the evaluation of Hopkins" (p. 205; Kite et al., 2008).

The unequal representation and treatment of women and men in STEM has prompted the development of interventions designed to reduce this gender gap. The types of interventions have been varied, ranging from the implementation of educational and mentoring programs at both early stages of education and university-level (Schneider, 2001), to changes in course structure (Brewe et al., 2010), to laws for gender equality (NCWGE, 2017), and the introduction of workplace policies to improve hiring, retention, and promotion of women in these fields (International Labour Organization, 2020). While these efforts have seen some progress in reducing the gender gap in STEM, they are mostly within the field of healthcare – a more 'communal' STEM field. Meanwhile, gender disparities in the fields of mathematics, physics, engineering, and computer science, both in education and the workforce, persist (National Science Board, 2016).

Psychological Interventions Targeting Stereotypes

While organisational-, structural-, and system-level interventions are critical for improving gender diversity in STEM, these may not be sufficient to produce efficient or effective change. Given the substantial literature that finds the gender gap in STEM is perpetuated by gender stereotypes, interventions targeting stereotypes may be an important individual-level approach. Thus, this research aims to develop and test effective psychological interventions aimed at targeting gender-STEM stereotypes.

Targeting Implicit vs Explicit Stereotypes

Stereotype interventions are designed to target explicit, implicit, or both types of stereotypes. Explicit stereotypes are conscious, intentional stereotypes we hold. Implicit stereotypes are the unconscious beliefs we hold based on our experiences and learned associations. For example, people can have low levels of explicit gender stereotypes (e.g., value gender equality and overtly believe women are as capable as men in STEM careers) but have high levels of implicit stereotypes (unconscious associations of men with STEM careers and women with family caretaking). Indeed, research shows implicit stereotypes may be more predictive of stereotypic behaviour than explicit stereotypes (Sinclair et al., 2005). As such, much of stereotype reduction work has focused on targeting implicit beliefs.

However, attempting to change people's core beliefs and attitudes about something as salient and deeply ingrained as gender is complex. Designing interventions to create lasting reductions in implicit gender stereotypes is difficult, and perhaps misguided, given that implicit bias is the accumulation of learned associations from one's social environment. Interventions aimed at changing implicit bias likely cannot permanently change this cognitive social imprint. Instead, implicit bias reduction programmes create a temporary environment in which counter associations are repeatedly exposed and temporarily learned. This would explain why associations learned in a lab or online experiment show immediate change in implicit bias, but no lasting effects a few days later (Lai et al., 2016; Paluck & Green, 2009). Implicit bias may be a better measure of societal-level beliefs, attitudes, and behaviours than individual-level. Through this lens, Forscher and colleagues (2016) suggest efforts would be best spent "[equipping] people with strategies to resist the social environment's biasing influence" (see also Devine et al., 2012; Cohen et al., 2009).

Dual-Process Theories

Some dual-process theories suggest that implicit and explicit routes of processing are distinct (Devine, 1989), while most others suggest they are interdependent (Cunningham & Zelazo, 2007; Gawronski & Bodenhausen, 2006; Petty, Briñol, & DeMarree, 2007). Specifically, interdependent dual process theories propose that cognitive processes can intervene on the application of implicit stereotypes. According to our implicit associations and stereotypes, we immediately categorise people when we meet them or see them. However, if we have the knowledge - and are motivated to - we can consciously interrupt this

process and choose to think more systematically. When people acknowledge their implicit stereotypes, they can then individuate others and reduce the effects of their implicit stereotypes on decisions, judgements, and behaviours (Burgess et al., 2007; Blair, 2002; Sherman et al., 2005). In the absence of motivation and knowledge, implicit stereotypes will guide decisions, judgments, and behaviours.

This would perhaps suggest that targeting implicit stereotypes directly would be better when motivation, cognition, or mental resources are lacking or not readily accessible. However, Forscher and colleagues' (2016) meta-analysis findings did not support this. Additionally, they found no evidence that changes in behaviour as predicted by explicit bias change was mediated by implicit bias. It may be that change in explicit stereotypes and the subsequent impact on behaviour has a feedback effect on implicit bias as our experiences and associations would change as a function of our expressed beliefs and behaviour. Given the above, this research programme aims to target explicit stereotypes and the mental processes associated with reducing stereotype application.

Intervention Approaches

As this thesis develops and examines interventions targeting gender stereotypes, I will now review the literature on widely studied intervention approaches (see Paluck & Green, 2009; Paluck et al., 2021). I have organised this section by cognitive and affective approaches.

Cognitive Approaches

Some approaches to stereotype reduction target cognitive mechanisms for changing beliefs. Examples of these interventions include traditional education, cognitive retraining, and attention or awareness interventions (discussed and cited below). Such an approach typically presents the participants with information about the targeted group, counter stereotypic stimuli, or cognitive practices for engaging with the stereotype differently. As such, this method does not rely on evoking identification with or concern for the member(s) of the target social group. This approach tends to be most effective when the logical arguments are not only strong, but also presented by someone who the observer perceives to be an expert (Petty & Briñol, 2010; Petty & Cacioppo, 1986; Pornpitakpan, 2004; Slater & Rouner, 1996; Tormala, Briñol, & Petty, 2007). I will now briefly review the literature on cognitive approaches.

Instructional Education. Educational or instructional interventions are derived from the assumption that stereotyping is a result of ignorance (Stephan & Stephan, 1984). They aim to increase one's accurate knowledge, either of the target group (e.g., Hill & Augoustinos, 2001; Tompkins et al., 2015) or how to avoid inaccurate group generalisations (Schaller et al., 1996; Gardiner, 1972). On their own, traditional educational interventions involving education of the target group are not very convincing (see Paluck & Green, 2009). For example, a significant drop in prejudice and negative stereotyping towards Aboriginal Australians was found immediately following an educational stereotype-reduction intervention (Hill & Augoustinos, 2001). However, at the 3-month follow-up, negative stereotypes had returned to baseline levels. Moreover, knowledge at 3-months post-intervention remained significantly higher than baseline levels, which indicates that stereotyping is more than just a product of ignorance. Similarly, Tompkins et al. (2015) found that transgender education interventions were not as effective in reducing stigma and changing attitudes as hearing humanising stories and engaging in perspective taking. Thus, while education of the target group can increase historical and cultural knowledge of the stigmatised group, it does not appear to impact beliefs or attitudes toward the target group. Overall, instruction techniques on their own lack theory and impressive findings (see Paluck & Green, 2009 for review).

Cognitive Retraining. Cognitive retraining interventions ask participants to think of counter-stereotypic people like a female mechanic or a Black CEO (Carnes et al., 2015; Levy & Dweck, 1999; Levy et al., 2004). Although short-term laboratory research shows positive results for cognitive retraining, field experiments and longitudinal analyses show weaker effects (e.g., Khan & Katz, 1978; Katz, 2000; Paluck & Green, 2009). This may be explained by the extent, strength, and familiarity of stereotypical stimuli in the real world. Further, a recent study on racial stereotyping found that counter-stereotyping on its own did not reduce stereotype application; however, it significantly reduced stereotype application among those with negative self-directed affect (i.e., guilt) (Burns et al., 2017). These findings build on previous work that shows the significant role of self-regulatory efforts to reduce the application of stereotypes (Amodio et al., 2007; Monteith, 1993; Devine et al., 2012) and suggest that cognitive mechanisms alone may not be sufficient to influence stereotypes. To see sustained change through counter-stereotypic techniques, the counter-stereotypic stimuli would likely need to happen at a systemic level, whereby people are consistently exposed to counter-stereotypic information in their real world. Indeed, schemas are incredibly challenging to change ((Padesky, 1994; Aosved et al., 2009).

Attention Strategies. Alternatively, education programmes aimed at training system 2 thinking to override system 1 thinking (Gardiner et al., 1972) and training in statistical reasoning (Schaller et al., 1996), demonstrate modest effects. These attention strategies have been found to improve judgment (Ashton,1992; Doney & Armstrong, 1996). These techniques are often used in awareness-increasing or consciousness-raising interventions (see Blair, 2002 for review). Consciousness raising strategies focus not on suppressing but rather acknowledging, anticipating, and actively redirecting attention (Goto, 2007; Klein, 2005; Jolls & Sunstein, 2006; Son Hing, Bobocel, et al., 2002; see Blair, 2002 for review). It is proposed that this increase in awareness increases guilt which then increases motivation to mitigate the influence of negative stereotypes (Son Hing, Li, Zanna, 2002; Duguid & Thomas-Hunt, 2015; Amodio, Devine, & Harmon-Jones, 2007). However, increased awareness of one's prejudice could alternatively result in cognitive dissonance and shame among people who hold low explicit stereotypes, which may encourage individuals to come up with explanations to rationalise and defend their attitudes and behaviours.

Additionally, strategies aimed at suppressing stereotypes takes motivation, cognitive effort, and resources (Shelton et al., 2005; Wyer et al., 2000) and can therefore negatively impact decision-making (Richeson & Shelton, 2003). They appear to be very short lived, such that once suppression mechanisms relax, stereotypic beliefs rebound (Macrae et al., 2004). Indeed, instructing people to inhibit stereotyping does not appear to be effective, and can even result in a boomerang effect, in which stereotypic thoughts reappear at a greater rate than if they had never been suppressed (Macrae et al., 2004; Galinsky & Moskowitz, 2000; Kulik et al., 2000). In other words, people instructed to suppress engaged in stereotyping at a greater rate than those who were not instructed to avoid stereotyping.

Altogether, cognitive efforts and resources may be best spent trying to *increase* discernment of stereotypes, not *reduce* or inhibit the stereotypes themselves.

Affective Approaches

Nearly all affective-based stereotype interventions use perspective taking. Perspective taking is one of the most common intervention approaches for stereotypes and prejudice (Hsieh et al., 2022). It is typically evoked through instructions to the participant to take the view of the target individual while reading, watching, or listening to a story (e.g., Stephan & Finlay, 1999) or to take the perspective of a target individual while writing a story (e.g., Galinsky & Moskowitz, 2000; Vescio et al., 2003).

Perspective taking. Perspective taking, in the context of stereotype-reduction interventions, involves imagining how it would feel to be a negatively stereotyped out-group member (Galinsky & Moskowitz, 2000). From around the age of three, individuals start to develop the understanding that others have thoughts, feelings, and experiences distinct from their own, and which inform others' behaviour (Wellman, 1990). Therefore, it can help people to predict and interpret the behaviour of others as well as guide social interactions with others (Singer & Tusche, 2013; Johnson, 1975; Galinsky et al., 2005; Batson et al., 1997). It is considered an important sociocognitive process in improving attitudes towards others (Vescio et al., 2003), reducing stereotyping (Galinsky, 2000), and reducing implicit bias (Todd et al., 2011).

In a field experiment, door-to-door canvassers conducted a 10-minute perspective-taking intervention aimed at reducing transphobia (Broockman & Kalla, 2016). The treatment group reported significantly more support for non-discrimination laws and significantly less transphobia three months after the conversation with the canvasser (Brockman & Kalla, 2016). The long-term success of the perspective taking element in Broockman and Kalla's (2016) "deep canvassing" field experiment may have been more related to the canvasser listening than talking. "Deep canvassing" involves asking sensitive questions, non-judgmental listening, and responding with sincere questions. If a participant expressed prejudice, the canvasser would gently ask them to recall a time when they were treated unfairly for being different and what that felt like. This supportive, psychologically safe conversation may be key to the effectiveness of perspective taking interventions. Indeed, in more recent field experiments, Kalla & Broockman (2020) found that conversations involving only persuasive arguments had no persuasive effects on voters' prejudice, whereas conversations with persuasive arguments and non-judgemental exchange of narratives did. However, upon closer examination, it is worth noting the size of the effect was very small (ds = .04 to .08). Overall, perspective taking interventions do not demonstrate significant effects (see Hsieh et al., 2022 for meta-analysis).

Cognitive vs Affective Approaches

To compare the efficacy of cognitive-based versus affective-based interventions, Pietri et al. (2017) compared two types of intervention videos, expert and narrative, aimed at reducing bias towards women in science. The purpose of the narrative videos was to encourage the participant to relate to women in science through transportation, closely associated with perspective taking. The purpose of the expert videos was to promote logical

thinking about the harmful effects of gender bias. The narrative videos presented scenes of gender bias in science, while the expert videos presented a professor debunking stereotypes and explaining facts related to gender bias in science. The researchers found both types of videos were effective in reducing bias, with the hypothesised indirect effects described above. However, the expert video had some advantages over the narrative video. Specifically, participants in the expert-video condition reported greater perceived increase in knowledge of gender bias and, behaviourally, demonstrated greater responsibility for confronting gender bias compared to the participants in the narrative video condition.

Similarly, in a recent meta-analysis on prejudice-reduction interventions, awareness and perspective taking interventions were named as two of the most popular intervention approaches for prejudice reduction (see Hsieh et al., 2022 for meta-analysis). Awareness interventions, defined as interventions that address the errors associated with stereotypical views and/or the negative consequences of prejudicial views, produce a significant average effect size of 0.442, but perspective taking interventions demonstrate no significant effect (Hsieh et al., 20220). Overall, then, findings indicate the superiority of cognitive approaches that overtly elicit logical thinking over perspective taking approaches that instruct perspective taking.

Psychological Mechanisms

While it is evidenced that cognitive approaches are effective when they elicit logical thinking (cognitive mechanism), it is not known if there are also affective mechanisms involved. Understandably, cognitive psychological mechanisms are assumed in cognitive-based interventions, and affective psychological mechanisms in affective interventions, but it is unlikely that cognitive and affect mechanisms truly operate alone. It is possible that in cognitive interventions, participants also engage in the affective process of empathy towards the target group when learning about the inaccuracy and harmful effects of stereotypes. Similarly, participants may engage in the affective-cognitive process of perspective taking, imagining oneself in the place of the target group when presented with facts about their negative experiences. For clarity purposes, perspective taking is an intervention approach and it is also, separately, a psychological process. Research has not yet empirically evaluated the relative roles of cognitive and affective mechanisms in stereotype-reduction interventions, but rather assumes or evaluates the mechanism related to the approach.

The present research examines three psychological mechanisms associated with popular stereotype-reduction: logical thinking, empathic concern, and perspective taking. These can be broadly represented as cognitive (logical thinking), affective (empathic concern), and affective-cognitive (perspective taking) processes. In this section, I outline logical thinking, empathic concern, and perspective taking and their relevance to the present research, as well as how they have been examined in past research. Lastly, I briefly review the gender differences observed in these variables and implications for the present research.

Logical Thinking

Thinking logically and critically about the facts and evidence presented is evidenced to be an effective mechanism in stereotype-reduction interventions. For example, in Pietri et al.'s (2017) study, participants in the logical-thinking (expert video) condition experienced greater perceived increase in knowledge of gender bias and, behaviourally, demonstrated greater responsibility for confronting gender bias compared to the participants in the transportation (narrative video) condition. Similarly, Cárdaba et al. (2014) found that participants who spent more time thinking about the logical arguments presented in an intervention demonstrated higher resistance to stereotypes against immigrants. Overall, findings indicate the importance of eliciting logical thinking in cognitive-based stereotype reduction interventions.

Empathy and Perspective Taking

Empathy has been interpreted as a psychological mechanism that drives individuals to respond with sensitivity and care to others' hardship (Batson, 2009). People are found to have less empathy towards dissimilar others than similar others (Arroyo, 1996; Plant & Devine, 2003). Perspective taking is sometimes considered the cognitive dimension of empathy (Edgar, 1990). Researchers propose perspective taking is an effective intervention method for stereotyping and prejudice because it elicits empathy towards the target group (Shih et al., 2009). For example, Johnson (2013) found that increased affective empathy for Arab-Muslims, elicited through a fictional novel about a counter stereotypical Arab-Muslin woman, led to prejudice reduction. However, this linear relationship from perspective taking to empathy has largely been assumed, not measured. Empathy on its own has been considered an effective way to improve psychological interventions (Byrnes & Kiger, 1990; McGregor, 1993; Smith, 1990), for example, in cooperative-learning groups (see Johnson & Johnson, 1992 for review); yet, other research does not find an effect of empathy in

stereotype-reduction interventions when it is directly manipulated (e.g., Finlay & Stephan, 2000). Thus, the instruction of engaging in empathy or perspective taking may be less effective than the organic process of experiencing them.

Conceptualising Empathy and Perspective Taking. Because empathy and perspective taking are interrelated concepts, I will clarify here the distinction I use for the purposes of this thesis. While digging into the literature on empathy, it became clear that different researchers use the term empathy to represent different concepts, which is problematic for validity, reliability, and replicability (Batson, 2009; Decety & Cowell, 2014; Levy, 1997). A recent review found forty-three distinct definitions (Cuff et al., 2016). The term is used to identify a wide range of cognitive, affective, and behavioural phenomena (see Lishner et al., 2017; Batson 2011). While some researchers treat affective empathy and perspective taking as distinct (e.g., Davis, 1983; Strayer, 1990), other researchers suggest perspective taking is a dimension within the wider construct of empathy (Decety & Jackson, 2006).

Much of psychological research uses empathy as an umbrella term to encompass how we understand and relate to others, including both affective and cognitive components. In terms of measurement, even some of the most widely used empathy measures, including the Interpersonal Reactivity Index (Davis, 1983), do not actually include the term empathy in any of the items. A recent systematic review of empathy literature examined issues related to the construct's conceptualisation and operationalisation and concluded that the term empathy should be bypassed completely, in favour of using lower levels of abstraction (e.g., emotion sharing, caring feelings, perspective taking, and so on; Hall & Schwartz, 2019). Indeed, research has found that the different empathy dimensions are uniquely evoked (Cummings et al., 2021) and have differential downstream effects (e.g., Wei et al., 2019; Batson et al., 1997). Further research has found affective empathy and perspective taking to be neurologically dissociable (Tusche et al., 2016). Therefore, the dimensions should be examined independently (Cummings et al., 2021).

Some scholars have conceptualised empathy as a two-dimensional construct, consisting of cognitive and affective components (e.g., Strayer, 1990). However, social neuroscientists, cognitive psychologists, social psychologists and developmental psychologists suggest dissociating affective empathy from perspective taking altogether (Stietz et al., 2019; de Vignemont et al., 2006; Kanske, 2018; Brown et al., 2017; Singer &

Lamm, 2009; Decety & Jackson, 2004), suggesting there is no need to capture perspective taking in the term empathy. While this would offer some clarity, it would still merge personal distress, empathic concern, and affect sharing under the term 'empathy'.

Duan and Hill (1996) suggest the affective component should be restricted to experiencing the same emotions experienced by the target individual/group (e.g., distress), also known as emotional mimicry (Decety & Jackson, 2006), emotional contagion (Healey & Grossman, 2018), empathic contagion (Gladstein, 1983), parallel empathy (Davis, 1994), or affect sharing (Stietz et al., 2019; de Vignemont et al., 2006; Kanske, 2018; Brown et al., 2017; Singer & Lamm, 2009; Decety & Jackson, 2004) – to illustrate the issue of inconsistent definitions. Other researchers (e.g., Batson et al., 1997; Finlay & Trafimow, 1998; Batson et al., 1997) measure empathy as empathic concern or empathic reaction, emotions that arise in response to the others' experience (e.g., compassion) – to illustrate the issue of inconsistent operationalisation. This type of empathy, empathic concern, has been found to mediate the effects of interventions on attitudes and behaviours (Finlay, 1996; Finlay & Trafimow, 1998). Conversely, distress, associated with emotional mimicry, has been found to elicit egoistic motivation to reduce one's own unpleasant arousal (Batson et al., 1987). Therefore, empathic concern was assessed in the current research.

Most researchers agree the cognitive dimension of empathy is perspective taking (e.g., Duan & Hill,1996; Decety & Jackson, 2006; Davis, 1983; Stietz et al., 2019). Some argue perspective taking can be further categorised into affective and cognitive components as well (Healey & Grossman, 2018; Cassidy et al., 2003). In this way, cognitive empathy and affective perspective taking would share significant overlap. While more lower levels of abstraction than 'empathy' would be useful, it may be too granular to tease apart both empathy and perspective taking into cognitive and affective components. This level of dissection may not be meaningful. We likely engage in the cognitive and affective processes simultaneously, even if to differing extents, in such a way that they are always interconnected and inform each other. However, research shows that these are dissociable anatomical substrates (Healey & Grossman, 2018). Regardless, it may not be appropriate to extrapolate this sub-perceptual, bidirectional processing through orthogonal, explicit self-report measures – which assume self-awareness of accuracy of experience. That being said, one's perceived experience of relative cognitive and affective states, irrespective of neurobiological-level of accuracy, may be useful for designing social psychological interventions. If, on average, individuals report greater tendency to psychologically shift while in certain (self-perceived) states, interventions can then be designed to evoke these perceived states.

However, the accuracy of one's perceived evoked cognitive and affective states, and their relative weightings, may be more reflective of one's general propensity for perceiving information and psychological states than an effect of specific intervention material. Thus, it may not be that the intervention materials are effective because they evoke affective and/or cognitive state(s), but rather that they are more effective for people who more naturally and easily enter said states. Indeed, individuals demonstrate different propensity for engaging in affective empathy or perspective taking (Tusche et al., 2016). Thus, the present research examines state empathic concern and state perspective taking separately as well as trait empathic concern and trait perspective taking.

Gender Differences in Empathy. A vast literature indicates that, on average, women are higher in empathy than men (see Christov-Moore et al., 2014 for review). Women demonstrate higher empathy than men not only on self-report questionnaires (Baron-Cohen & Wheelwright, 2004; Davis, 1983; Eisenberg & Lennon, 1983; Mehrabian, Young, & Sato, 1988; Rueckert & Naybar, 2008), but also in laboratory experiments assessing emotional contagion response (Kevrekidis et al., 2008; Magen & Konasewich, 2011; Schulte-Rüther et al., 2008), accuracy of facial expression recognition/emotion recognition/emotion perception (Hampson et al., 2006; Thayer & Johnson, 2000; McClure, 2000; Schulte-Rüther et al., 2008), body language (Alaerts et al., 2011; Sokolov et al., 2011), and fMRI neuroimaging studies (Singer et al., 2006; Cheng et al., 2010; Schulte-Rüther et al., 2008). Further, naturalistic experiments demonstrate a gender difference in empathy through gender differences in charitable giving (Mesch et al., 2006; Mesch et al., 2011).

Rueckert & Naybar (2008) suggest that gender differences in empathy may be explained by a neural difference in the activation of the right hemisphere in empathy-related situations. They found that, compared to men, women had increased right hemisphere activation during an empathy-inducing task, along with significantly higher empathy scores on the self-report questionnaire. In support of this neural explanation, patients with anterior temporal atrophy in the right hemisphere report greater empathy deficits than those with lesions in the left hemisphere (Perry et al., 2001; Shamay-Tsoory et al., 2003). Overall, consistent findings from self-report, neuroimaging techniques, and brain lesion studies increase the credibility of the neural explanation of gender differences in empathy.

Additionally, females recruit areas containing mirror neurons to a higher degree than males when engaging in face-to-face interactions in an empathic context. Together with the

reliance on cortical components in the right hemisphere in females as compared to males (Schulte-Rüther et al., 2008), these findings may underlie the gender differences in empathy. The neural explanation can be further explained by early hormonal influence. Specifically, the role testosterone plays in human brain development (Hines, 2006). Indeed, prenatal testosterone exposure appears to be related to empathy (Hines, 2011). For example, levels of testosterone measured in amniotic fluid inversely predict empathy in both males and females (Chapman et al. 2006; Mathews et al. 2009).

Alternatively, other research shows that the gender difference in empathy is negligible (e.g., Kim and Lee 2010; Loffler & Greitemeyer, 2021; Lamm et al. 2011) and largely explained by socialisation and gender roles, not innate differences (Loffler & Greitemeyer, 2021). Gender socialisation, the process by which individuals learn and embody their society's gender norms, provides social rewards – or punishments – for behaving in gender-typical – or atypical – ways (Hyde, 2013). Thus, the expectation that women are more empathic can motivate women to report and/or demonstrate greater levels of empathy than men. Likewise, the expectation that men are not empathic can motivate men to report and engage in less empathy than women. Consequently, women's advantage over men in empathy may be a result of greater reward and, therefore, greater practice. In support of this, research has found the gender difference in emotions is in the expression of them, not the experience (LaFrance & Banaji, 1992; Fischer, 2000). Nonetheless, this could have real consequences on brain correlates for gender differences, as explained by neuroplasticity, the brain's ability to change in response to experiences.

Evidence is mixed as to the relative effects of biological versus social factors in determining the gender difference in empathy, and it is most likely that they influence each other. Irrespective of how the gender difference in empathy originates, the difference may be important for stereotype- and prejudice-reduction interventions given the role empathy is theorised and evidenced to play.

Cognitive and Affective Mechanisms

Cognitive and affective mechanisms of stereotyping may be unrelated (see Gardner, 1973), such that the cognitive process of gender categorisation may be a fundamental, but not sufficient, explanation of gender stereotyping. Likewise, gender stereotype interventions that aim to reduce harmful cognitions of gender may be necessary, but not sufficient without an affective component. Based on the literature presented, interventions may be most effective

when eliciting both affective and cognitive mechanisms. Indeed, a vast literature indicates the advantage of combined cognitive and affective processes in stereotype- and prejudice-reduction (e.g., Son Hing, Li, Zanna, 2002; Duguid & Thomas-Hunt, 2015; Amodio, Devine, & Harmon-Jones, 2007; Brookman & Kalla, 2016; Kalla & Broockman, 2020). However, little is empirically known about the relative roles of each in the effectiveness of interventions.

Reconceptualising Stereotype Interventions

Much of stereotype-reduction research is exactly that – aimed at reducing stereotypes. This aim may be amiss, given the typical short, one-off structure of these interventions. Schemas, and therefore stereotypes, are incredibly difficult to change. They tend to persist even when contradicting evidence is presented (Padesky, 1994). It requires consistent, prolonged exposure to contradicting evidence for schematic content to update (Aosved et al., 2009). Not surprisingly, the positive change from stereotype-reduction interventions typically wears off after a day (Paluck & Green, 2009). Thus, psychological interventions may be better positioned to change one's discernment of stereotypes rather than the stereotypes themselves. For example, the ability to *identify* gender stereotypes and their harm is essential for their eradication (Cook & Cusack, 2011). Indeed, Cook and Cusack (2011) suggest gender stereotypes are like a disease. In this way, one cannot treat a disease if it is not diagnosed. Therefore, first, people must be able to discern gender stereotypes and their harm, and then they can be reduced.

Psychological Interventions Targeting Misleading Information

In Paluck et al.'s (2021) recent review and meta-analysis of prejudice-reduction interventions, they call for researchers to mix ingredients from multiple theoretical perspectives. Given the generally low success rate of typical diversity trainings and stereotype-reduction approaches, it may be fruitful to turn to other disciplines to inform intervention design. Since the foundation of stereotyping is endorsing overgeneralised, unjust, and largely inaccurate beliefs, theory and interventions from persuasion and communication research may be worth exploring. Specifically, this thesis will explore debunking and inoculation theory.

Debunking

Debunking is a psychological intervention aimed at correcting myths or misinformation that have settled into someone's mental model. The "treatment-as-usual" method of debunking typically involves a simple retraction, sometimes with some provision of factual information (Paynter et al., 2019). Simply retracting leaves a gap in knowledge as it is not replaced with alternative, true information; and even when a factual correction is provided to replace the old belief, people don't understand why the previous belief was inaccurate or misleading. Thus, as described in the Debunking Handbook, the debunking process should aim to "unstick" misinformation and, importantly, replace it with alternative information and explain why the misinformation is inaccurate (Lewandowsky et al., 2020). This optimized debunking should include four key components in order to increase the chances of its success:

1. Core Fact

The debunking process should start with a core fact, that can be communicated in a simple and clear way, and that has the same explanatory relevance as the original misinformation. For example, communicating a fact, "The scientific consensus is that women and men have equal aptitude for STEM", rather than a retraction, "The claim that men are better at STEM than women is not true", facilitates the swap between the old misinformation and the new information (Lewandowsky et al., 2020).

2. Warning about the Myth

At this stage, the debunking process should repeat the misinformation a single time before providing the correction. While multiple repetitions of misinformation have been shown to strengthen the misinformation argument (Lewandowsky et al., 2012), a single repetition has been shown to benefit the "swap" between misinformation and the corrected belief (Ecker et al., 2011, Ecker et al., 2017).

3. Explaining the Fallacy

Following repetition of the misinformation the next step is to rebut the misinformation, clearly and directly. The rebuttal should not simply deny the misinformation but also provide clear arguments against why those arguments are false. Specifically, the Debunking Handbook (Lewandowsky et al., 2020) advises explaining: (1) why the mistaken information was thought to be correct in the first place, (2) why it is now clear it is wrong and

(3) why the alternative is correct". By providing such detail, the individuals become aware of the inconsistencies of the misinformation, which facilitates the updating of the belief and its overtime protection (Ecker et al., 2017, Rich & Zaragoza, 2020).

4. Repetition of Core Fact

Finally, the debunking process should end with a repetition of the core fact, so that the last thing the subject hears is the correct information (Lewandowsky et al., 2020).

Paynter et al. (2019) compared the effectiveness of a typical debunking intervention and an optimised debunking intervention, based on the process and factors outlined above, and validated the importance of not just saying a piece of information is incorrect, but why it is incorrect. Likewise, in their meta-analysis, Chan et al. (2017) showed that the use of a detailed debunking message explaining the reasons why the misinformation is wrong, rather than just stating it as incorrect, improved the debunking effect.

A number of meta-analyses have been conducted on the efficacy of debunking as a psychological intervention to combat misinformation-based beliefs and attitudes (e.g., Chan et al., 2017), though most studies are often in the context of health misinformation. While there have been several calls to debunk gender stereotypes (Agu, 2021; Kang & Kaplan, 2021; Heilman & Caleo, 2018; Krišová, 2020; Alzaabi, 2021; Roberts & Hill, 2003; Amnesty, 2021), the few researchers who have attempted to do so acknowledge that debunking gender stereotypes is challenging but promising (Krišová, 2020; Breda et al., 2021) and more effective than traditional perspective-taking approaches (Pietri et al., 2017).

Inoculation Theory

While debunking interventions aim to expose the falseness of information in order to persuade people to change their beliefs, attitudes, or behaviours, inoculation interventions aim to *protect* beliefs, attitudes, and behaviours from persuasion -- or *prebunk* false information. Inoculation theory, coined by social psychologist McGuire (1964), is considered the grandparent theory of resistance to change (Eagly & Chaiken, 1993). Like a medical vaccination, the analogy is to attitudinally 'vaccinate' people against persuasion or misinformation. Specifically, inoculation involves two key components: (1) an explicit warning about misinformation and the danger of being misled and (2) pre-emptive refutation

to the misinformation, explaining the flaws in the misinformation message (Banas & Rains, 2010). This can be conceptualized as the process of (1) receiving a weakened version of the 'virus' (misinformation warning) and (2) the production of 'building antibodies' to protect oneself from being infected by the virus (counterarguments addressing the flaws).

Inoculation theory has demonstrated success in protecting attitudes and maintaining behaviours in several arenas, including climate change, public relations, health, marketing, and political campaigning (see Compton & Ivanov, 2013; Banas & Rains, 2010; Godbold & Pfau, 2000; Veil & Kent, 2008; van der Linden et al., 2017). Banas and Rains (2010) found an overall effect size of d = .43 in their meta-analysis of interventions based on inoculation theory.

Therapeutic Inoculation

Although inoculation theory is a strategy originally and traditionally used to protect people from change, Wood (2007) found that their inoculation message not only protected beliefs but also changed the beliefs of those initially opposed to the inoculation message in favour of it: *therapeutic inoculation*. Similarly, van der Linden et al. (2017) found that their inoculation intervention effectively changed climate change perceptions in favour of the scientific consensus. Compton (2019) and Wood et al. (2007) argue that therapeutic inoculation is not inconsistent with the medical analogy, as therapeutic vaccines can be administered after a disease or infection has occurred to activate the immune system of the patient. Indeed, people with chronic viral infections are given therapeutic vaccine as their immune systems do not produce enough antibodies (Shimasaki, 2014).

As we can assume everyone has had substantial prior exposure to gender stereotypes (Paechter, 2007; Paludi et al., 2004; Plant et al., 2000; Rheingold & Cook, 1975; Seavey et al., 1975; Condry & Condry, 1976), there may be people with a chronic infection who do not produce enough antibodies. With differing levels of gender beliefs and attitudes (individual differences in levels of immunity and infection), we can imagine that an inoculation intervention targeting gender stereotypes would be therapeutic for many. Given that therapeutic inoculating messages are most effective when they (1) explain the flawed argument or technique used in the misinformation and (2) highlight the scientific consensus (Cook et al., 2017), therapeutic inoculation is closely aligned with optimised debunking; both aim to change beliefs and attitudes informed by false information. The relationship and distinctions between debunking and therapeutic inoculation will be discussed in greater detail

in Chapter 2. In sum, inoculation theory could be a promising avenue for stereotype reduction.

Psychological Mechanisms

The two theoretical pillars of traditional (prophylactic) inoculation theory are threat and counterarguing. The vulnerability of being misled is meant to elicit threat and, in turn, heighten resistance to change through counterarguing. Within the medical inoculation analogy, this can be conceptualized as the process of (1) receiving an infection of the weakened virus (threat) and (2) building antibodies to protect against and attack future viral invasion (counterarguing) (Compton, 2013). Thus, the psychological mechanisms underpinning inoculation-conferred resistance are affective (threat) and cognitive (counterarguing).

However, this linear separation of affective and cognitive components may not be productive or accurate at a psychological level of assessment, as discussed throughout this thesis, given that these processes are likely intertwined. Likewise, Compton (2013) identifies that there are conflicting findings for the linear assumption that threat leads to counterarguing and, consequently, counterarguing leads to resistance. Moreover, this relationship has largely been investigated in prophylactic inoculation. While the psychological processes involved in either strengthening (prophylactic inoculation) or changing (therapeutic inoculation) attitudes and beliefs may be similar, it is important to investigate and clarify for therapeutic inoculation theory building (Compton & Pfau, 2005; Compton, 2020).

Threat. In developing an understanding of the process of inoculation, McGuire (1964) identified implicit threat and explicit threat as the affective elements in the generation of resistance. Implicit (or inherent) threat was brought about by the counterarguments that challenge an existing position or belief; that is the pre-emptive refutation in the inoculation message. Meanwhile, explicit threat was elicited by a direct forewarning of a persuasive attack; that is a message which directly warns the recipient that, although they currently hold the right belief or attitude about a specific issue, they will encounter future arguments that have the potential of challenging or even changing that belief or attitude. More contemporary research has conceptualised threat as apprehensive (fear of belief or attitude being attacked) and motivational (motivated to protect belief or attitude).

The level of explicit threat in the inoculation message can impact the effectiveness of an inoculation intervention (Compton, 2020). A level of threat that is too high risks overwhelming the participant and even lowering attitude certainty (Pfau et al., 2004), while a

not-high-enough level of threat risks being insufficient for the generation of resistance, which is key for successful inoculation. Further, threat may affect individuals differently. Indeed, treatment-elicited threat is found to be affected by factors such as gender and pre-existing beliefs and attitudes (Wood, 2007; Pfau et al., 2008). When considering therapeutic inoculation, given that threat elicits psychological protection, which is associated with inhibiting processes like withdrawal and avoidance (Shen & Dillard, 2007; Dillard & Anderson, 2004), a sense of psychological safety may be important to prevent such psychological reactance. This would then allow for the motivational component of threat to be activated – e.g., the motivation to hold correct beliefs in line with the scientific consensus (see the Gateway Belief Model, van der Linden & Leiserowitz, 2019).

Much of inoculation scholarship has regarded threat as the more important pillar of the two (McGuire & Papageorgis, 1962; Pfau, 1997; Compton & Pfau, 2005; Compton & Ivanov, 2012). However, Banas and Rains' (2010) meta-analysis did not find an effect of threat on resistance in inoculation studies. This may be because most studies have measured threat as fear-based and not motivation-based. Indeed, in a follow-up paper, Banas and Richard (2017) found that motivation-based threat was more predictive of resistance in inoculation than the traditional fear-based measure. This is perhaps not surprising given fear-based psychological protection is associated with inhibiting processes like withdrawal and avoidance (Shen & Dillard, 2007; Dillard & Anderson, 2004).

Further, as threat is typically measured as the threat elicited by the treatment message, there is little research on the threat or psychological reactance actually experienced in response to future attacks. Improving understanding of the psychological reactance (i.e., 'immune response') to the misinformation attack ('virus') would be fruitful for continued inoculation theory development.

Counterarguing. McGuire (1964) speculated that counterarguments had greater cognitive impact than just reinforcing beliefs with supportive evidence. Inoculated individuals could then use these counterarguments ('mental antibodies') in the future when they encounter an 'attack'. In addition to this passive process of inoculation, through having counterarguing modelled, inoculated individuals can also begin to elaborate and develop counterarguments of their own, 'active' inoculation (Compton & Pfau, 2005; Roozenbeek & van der Linden, 2018).

As such, for successful refutational preemption, messages do not need to address and refute all potential counterarguments (Pfau, 1995). Further, protection can be conferred by

both refutational-different (i.e., the treatment provides refutations to falsehoods that are not the same falsehoods in future attacks) and refutational-same (i.e., the treatment provides refutations to falsehoods that are the same falsehoods raised in future attacks; McGuire, 1964; Banas & Rains, 2010).

As well as the content of refutations, the strength of counterarguments and refutations can also impact the degree of intervention efficacy. Compton and Pfau (2004) tested four inoculation conditions: strong counterarguments with strong refutations, weak counterarguments with weak refutations, strong counterarguments with weak refutations, and weak counterarguments with strong refutations. While all four conditions elicited threat, they found that matched argument strength was most effective.

In contrast to threat, counterarguing has been defined and measured more consistently. Most research has measured it following the future 'attack' and operationalised it as the extent to which participants self-perceive engaging in counterarguments against the misinformation. Counterarguing, then, is conceptualised as the activity of generating counterarguments in the time between exposure to the inoculation treatment and the subsequent attack message (Compton, 2013; Compton & Pfau, 2005).

Integrating Debunking and Inoculation Interventions

Taken together, there is significant overlap in optimised debunking and inoculation interventions (see Figure 1.1), which will be discussed in more detail in Chapter 2. In general, debunking interventions and therapeutic inoculation interventions both aim to *change* beliefs and attitudes informed by false information, while prophylactic inoculation interventions aim to protect beliefs and attitudes from being misled by false information. The vast literature of research targeting stereotypes tends to, understandably, focus solely on stereotype reduction (debunking). However, it may be important to also consider how to effectively protect the beliefs and attitudes of those who do not currently endorse gender stereotypes. Thus, in tandem with debunking, inoculation theory may be a promising strategy to explore in stereotype interventions.

Psychological Considerations and Limitations

I will now discuss psychological factors that influence the effectiveness of debunking and inoculation interventions together.

Decay Effects. While debunking and inoculation treatments have been shown to successfully replace misinformation with updated true information, the effects are limited and can wear off over time (Paynter et al., 2019, Swire et al., 2017; Maertens et al., 2020). For example, some interventions which appear to be successful when assessed immediately post-intervention, show no effect maintained at 1-week (Paynter et al., 2019) or 6-week (Smith et al., 2011) follow ups. Further, in a meta-analysis by Banas and Rains (2010), inoculation effects were found to decay after two weeks. This has practical and economic implications for real-world application, as repeated interventions seem to be important. Indeed, 'top-up' effects have been suggested as promising additions to inoculation interventions (Compton & Pfau, 2009; Ivanov et al., 2015). Just as some vaccines require a second dose, the concept is that the effect of the inoculation could become more robust if given a top-up of the treatment.

Backfire Effect. Another limitation of debunking and inoculation is the backfire effect. The backfire effect occurs when presenting factual information to correct previous incorrect beliefs actually increases the individual's belief in the misinformation (Ecker et al., 2017). For example, after being presented with facts about the absence of weapons of mass destruction in Iraq, a group of conservatives displayed increased belief that such weapons had indeed been found (Nyhan & Reifler, 2010). Similarly, Pluviano et al., (2017) found that their "myth vs. fact" and fear-inducing pro-vaccination strategies resulted in increased anti-vaccination beliefs.

An intertwined affective-and-cognitive element involved in the process of "backfiring" is psychological reactance (Dillard & Shen, 2005), a negative reaction that people experience when they perceive their freedom to be eliminated or threatened (Brehm, 1966). In the context of misinformation, confronting someone with an argument that contradicts their prior belief and tells them that something they held as true is wrong may evoke such feelings. This phenomenon has been described extensively in the context of courtrooms, with research showing that jurors are more likely to take into account evidence that has been considered "inadmissible" after being specifically ordered by the judge to disregard it (reviewed in Lieberman & Arndt, 2000).

Another phenomenon that is closely linked with backfiring is belief polarisation, which occurs when two parties with opposing views are presented with the same information and this leads to increased attitudinal divergence (Lord et al., 1979). In a study including both a group of individuals favouring nuclear power and a group of individuals opposing nuclear power, after being presented with the same evidence about a non-catastrophic near-

meltdown at a nuclear power plant, the attitudes of both groups became more polarised; supporters focused on the fact that the safeguards had worked and prevented a total meltdown, while opposers focused on the fact that there had been an accident in the first place (Plous, 1991).

Despite the evidence presented above, more recent research indicates the backfire effect may be more negligible than previously thought. A large-scale meta-analysis conducted by Wood & Porter (2018), which included over 10,000 subjects and 52 polarised issues with potential for backfiring, found that correction did not trigger backfiring in any case. Indeed, it seems the backfire effect is largely related to repeating the myth, which would be circumvented with optimised bunking approaches (Lewandowsky et al., 2020).

Resistance to Belief Change. While the backfire effect may not be as worrisome as previously thought (van der Linden, 2022), there is still the less-extreme issue of *resistance* to corrections of misinformation. A number of cognitive explanations have been identified to explain when and why this happens. Among them is the principle of *fluency* and *familiarity* (Lewandowsky et al., 2012, Schwarz et al., 2007), which relies on two cognitive processes: i) information to which we have been previously exposed, meaning it is familiar, is processed more fluently and ii) information that our brains process more fluently, in other words more easily and faster, is assumed to be more likely to be true. This phenomenon of fluency and familiarity has been described as the "illusory truth effect", by which even a single exposure to misinformation can increase the perceived truthfulness of that argument in future exposures (Pennycook et al., 2018).

Additionally, repetition of information leads to a strengthening and settling of that information in one's memory, and it becomes associated with other information that we know to be true (Lewandowsky et al, 2012). In other words, how the brain decides if something is true largely depends on how often we are exposed to it. The implications of this poses challenges for efforts to rectify old, familiar messages with new and correct information. In the context of gender, for example, continuous exposure to gender stereotypes, which are ubiquitous in society and propagated by media, literature, schools, workplaces, and so on, makes interventions to reduce gender stereotypes challenging. Through constant exposure to gender stereotypes, we have developed strong neural connections, or associations, for them; and our brains interpret this frequency as an indicator of truth.

In the context of gender, for example, continuous exposure to gender stereotypes, which are ubiquitous in society and propagated by media, literature, schools, workplaces, and

so on, makes interventions to reduce gender stereotypes challenging. Through constant exposure to gender stereotypes, we have developed strong neural connections, or automatic associations, for them, increasing their fluency and therefore perceived truth.

Other cognitive factors linked with resistance to belief update are *confirmation bias* and *motivated reasoning* or *motivated cognition*. Confirmation bias, first described by Peter Wason in 1960 (Wason, 1960), refers to the tendency to interpret or recall information in a way that supports prior beliefs. It plays a key cognitive role in the perpetuation and exacerbation of gender stereotypes. For example, a common gender stereotype is that men are better suited for STEM. If presented with a situation in which a woman is in STEM, we are more likely to notice elements indicating her unsuitability for the role, which provides support our prior belief, while ignoring the elements that contradict such belief. Meanwhile, motivated reasoning, refers to our tendency to evaluate and process arguments in a way that favours our current views and opinions by, for example, quickly accepting facts that agree with our current opinion but more critically analysing those that do not. In other words, we are motivated to protect our worldviews, behaviours, etc.

Continued Influence Effect. A final limitation of debunking, and other psychological interventions is that, even when someone has successfully intellectually updated their beliefs, the previously held false information may still influence behaviour (Lewandowsky et al, 2012). The "continued influence effect" (see Chan et al., 2017; Cook et al., 2017; Lewandowsky et al, 2012; Ecker et al., 2010; Schwarz et al., 2016; Pluviano et al., 2017) describes this phenomena: how, when people learn a prior belief is false, and they intellectually accept the updated information, they still reason and behave from the old information.

Putting the above together, many debunking strategies end up amplifying the familiarity of the myth, making it easier to process (Schwarz et al., 2016), which then leads to the illusory truth effect (Dechêne et al., 2009; Gigerenzer, 1984; Hasher et al., 1977). This effect happens even for myths that are initially rated as questionable, meaning the myth seems more valid and true after repeated exposure regardless of initial belief (Arkes et al., 1989). Once a belief is formed and settled, people find explanations and arguments that reinforce the belief, regardless of their validity (Nyhan et al., 2010).

Additionally, people are motivated to seek information that is congruent with deeply held beliefs and attitudes. As explained by cognitive consistency theories (e.g., Festinger, 1957), changing a core belief generates inconsistencies that can threaten one's self concept.

Therefore, information that challenges an individual's worldview or personal ideology can threaten one's self-concept, increasing the likelihood of backfiring (Prasad et al., 2009). Given the saliency of gender beliefs, when gender biases are challenged, a system of core beliefs and identities may be threatened too. To protect oneself from this uncomfortable experience, people may experience psychological reactance and be motivated to reject the true but incongruent information.

Intervention Design Considerations

When designing a psychological intervention, a number of factors can influence effectiveness. *Perceived credibility* of a source is important (Traberg & van der Linden, 2022). For example, if a message comes from a person or institution that is high in perceived credibility, it is much more likely to be effective (Guillory & Geraci, 2013). However, this effect may be driven by the source's perceived trustworthiness more than expertise. *Visual communication*, like imagery or graphs, can also help boost the impact of a correction (see van der Linden et al., 2017; Dixon et al., 2015) likely through increasing attention, processing, and retention. Enhancing the salience of a correction message, for example, through bold font or coloured boxes, can also help improve attention, processing, and retention (Schwarz et al., 2016; Paynter et al., 2019). *Detailed refutations* are more effective than general ones or the provision of factual information alone (van der Linden et al., 2017; Bensley & Lilienfeld, 2017; Swire et al., 2017; Kowalski et al., 2009). Finally, *accessible language*, with an empathetic and non-confrontational tone are also important (Lewandowsky et al., 2012). These were all taken into consideration in the present research design.

Present Research

The present research contributes to theoretical, methodological, and empirical advancements in gender-stereotype intervention literature. It presents the first examination of inoculation theory in the context of stereotyping, and the first integrated inoculation theory and optimised debunking intervention design targeting gender-STEM stereotypes. Responding to Paluck et al.'s (2021) and Paluck and Green's (2009) call for theoretical and empirical advancements, and greater use of field experiments, the intervention design examined the psychological mechanisms involved in improving stereotype discernment and

was tested using both online (Chapter 2 and 3) and field (Chapter 4) studies. Throughout, this research programme emphasises a multi-dimensional approach to stereotype reduction; it focuses primarily on improving discernment of gender-STEM stereotypes and, secondarily, reduction in gender-trait stereotype endorsement (Chapter 2 – 4), application of gender stereotypes (Chapter 4), and gender attitudes (Chapter 3 – 4). Finally, this thesis offers a more up-to-date conceptualisation and measurement of gender-trait stereotypes. Ultimately, gender stereotypes and their effects are important to study because many of the behaviours, abilities, and traits that are believed to be gender differentiated are also differentially valued in society.

CHAPTER TWO – DESIGNING AND TESTING VIDEO INTERVENTION MATERIALS TARGETING GENDER STEREOTYPES IN STEM: AN INTEGRATIVE APPROACH TO DEBUNKING AND INOCUALTION THEORY

In this chapter, I sought to apply the principles of optimized debunking and inoculation theory to design a novel intervention for targeting gender stereotypes in STEM. For this purpose, I will now provide a methodological and theoretical comparison of debunking and inoculation interventions and their application to this research. Further, I will discuss design and gender considerations relevant to the current research.

Introduction

Methodological Considerations

Although many debunking interventions do not follow it, a debunking message should include the same elements as an inoculation treatment message; that is: a statement of the facts, a warning of the myth, a detailed explanation of why the myth is incorrect, and finally, a repetition of the true fact (Lewandowsky et al., 2020). Thus, methodologically, the distinction between (optimised) debunking and inoculation is that, in an inoculation intervention, participants are exposed to a counter message attack after the treatment message (Figure 2.0). The purpose of such attack is to measure the inoculative effects of the intervention (i.e. the ability of the inoculation message to protect beliefs). Conversely, debunking interventions tend to measure outcomes directly following the treatment message(s), in order to measure endorsement and/or change in endorsement.

Theoretical Considerations

Theoretically, there is greater distinction between debunking and inoculation. In particular, in relation to the "infection status" of the person; while debunking aims to correct a prior (incorrect) belief that is already engrained into people's mental model (i.e. that person has become infected, in inoculation terms), prophylactic inoculation aims to prevent "infection" and *protect* from persuasion. Meanwhile, *therapeutic* inoculation aims to act during the "incubation period", that is the time after exposure to the "pathogen" (i.e. misinformation) and before "infection" (van der Linden, 2023).

It should be noted, however, that while this theoretical distinction is important for contemporary misinformation campaigns, such as the coronavirus pandemic or fake news headlines, it becomes less relevant in the context of more historically pervasive myths, such as gender stereotypes. Indeed, one could assume that everyone has, to a greater or lesser extent, been exposed to gender stereotypes, likely since they were born (Paechter, 2007; Paludi, Paludi, & DeFour, 2004; Plant et al., 2000; Rheingold & Cook, 1975; Seavey, Katz, & Zalk, 1975; Condry & Condry, 1976). Thus, theoretically, it is unlikely that individuals would be in an incubation period past early childhood.

Indeed, once enough time has passed since first exposure, therapeutic inoculation essentially becomes debunking, as the belief has settled and they have become "infected". Little is known about how often someone needs to be exposed to a piece of misinformation before they become infected themselves, and so the cross-over point between therapeutic inoculation and debunking is unclear. In the present research, the distinction between debunking and (therapeutic) inoculation interventions is a methodological one – the addition of the countermessage attack after the treatment to assess inoculative effects.

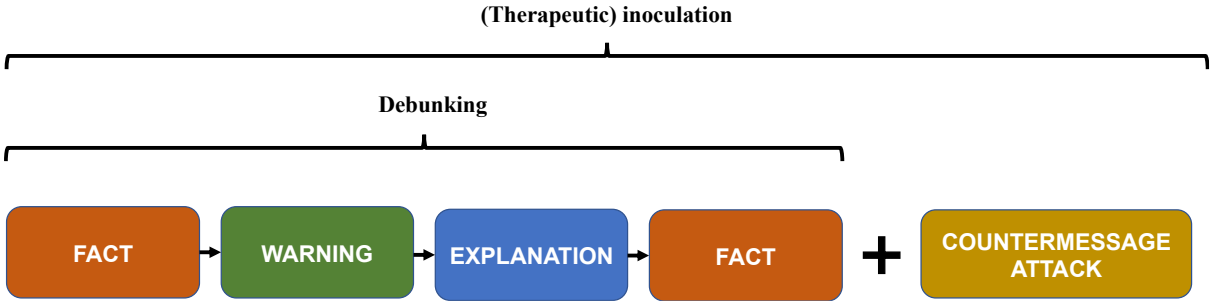
Further, the distinction between traditional inoculation and therapeutic inoculation is only possible if and when prior beliefs are measured. Again, the same intervention material would serve as traditional inoculation for someone with the "correct" belief and as therapeutic inoculation for someone with the "incorrect" belief.

In sum, debunking aims to change or update prior beliefs, therapeutic inoculation aims to change prior beliefs *and protect* updated beliefs, and traditional inoculation aims to protect prior beliefs. Thus, the purpose of this thesis is not to compare debunking with therapeutic inoculation by positioning them against each other, but rather to take an integrative approach and look at the effectiveness of the treatment message, which follows the same process model in both debunking and inoculation, in terms of improving discernment of gender stereotypes in STEM. Debunking/therapeutic inoculation and traditional inoculation are distinguished only by the prior belief held (myth vs accurate information), not the treatment message.

Based on the aforementioned methodological considerations, I will use the term "debunking" to refer to the treatment-only condition, "misinformation" to refer to the counter message attack (or myth), and "inoculation" to refer to the treatment + counter message condition. Thus, this thesis aims to examine i) the persuasive effects through the debunking conditions and ii) the protective effects of the treatment message against misinformation in the inoculation conditions.

Figure 2.0

Schematic showing the methodological process for (therapeutic) inoculation and debunking



Inoculation theory has previously demonstrated prophylactic and therapeutic effects on belief in the scientific consensus on climate change (van der Linden et al., 2017). This arena could be considered similar to gender stereotypes in that it is an important social issue and is fraught with misinformation. However, climate change is likely less part of one's core identity, on average, and people are less likely to be constantly exposed to climate change misinformation from a young age. Further, one may understand the scientific consensus, but this does not mean they necessarily agree with it or believe it. Therefore, as well as assessing acknowledgment of the scientific consensus, participants provided their discernment of gender-STEM myths, their negative effects, and misleading tactics.

In summary, the principle of debunking is to expose the falseness of information in order to change a belief, attitude, or behaviour. On the other hand, inoculation theory (McGuire, 1961) is a strategy used to *protect* attitudes, beliefs, and behaviours from persuasion or change. However, van der Linden et al. (2017) and others have found that inoculation can also effectively change perceptions in favour of the scientific consensus: *therapeutic* inoculation. Thus, therapeutic inoculation and debunking share the same theoretical intent. Methodologically, debunking, therapeutic inoculation, and prophylactic inoculation treatments are the same. The distinction lies in the participants' prior beliefs.

This intervention, from the perspective of inoculation theory, involved two key components: (1) a warning about the harmful effects of gender stereotypes ('the weakened virus') and (2) counterarguments to the gender stereotypes, scientifically explaining the flaws and inaccuracies ('building antibodies'). From the debunking framework, the treatment involved (1) a warning about the myth and (2) an explanation of the fallacies. Importantly,

these components were book-ended with the scientific fact that there is no reliable evidence for inborn gender differences in STEM aptitude.

Design Considerations

Media interventions aimed at reducing the effects of stereotypes may be more effective than traditionally designed interventions, which often rely on text and/or figures. Examples of successful media interventions for negatively stereotyped groups include the use of radio (Paluck, 2009), storybooks (Cameron, Rutland, Hossain, & Petley, 2011), television and film (Schiappa et al., 2005; Riggle et al., 1996), and most recently videos (Pietri et al., 2017; Pietri et al., 2018). Media interventions like videos may be effective at influencing perceptions and attitudes through direct and/or indirect mechanisms. Direct approaches occur through exposing participants to strong arguments communicated by a reliable source. Indirect mechanisms work through narratives, which attempt to engage the viewers in a story through transportation and identification with the characters (Chung & Slater, 2013). Inoculation theory and debunking align with the direct approach, whereas traditional stereotype-reduction interventions tend to align with the indirect approach.

Gender Differences in Gender-Targeted Interventions

Past research on gender bias training in STEM has reported participant gender can moderate the effectiveness of the intervention, such that men can be more impacted than women (e.g., Jackson et al., 2014). Moss-Racusin et al. (2018) cites this research and suggests this may be because, on average, men demonstrate more scepticism about gender bias evidence than women do (see Moss-Racusin et al., 2015). Building on this, men may be able to experience greater psychological change due to their different starting point.

Alternatively, other gender bias interventions in STEM have found no moderating effect of gender (Pietri et al., 2017; Moss-Racusin et al., 2018; Zawadski et al., 2012). These interventions used videos (Pietri et al., 2017; Moss-Racusin et al., 2018) and games (Zawadski et al., 2012) versus a presentation (Jackson et al., 2014), suggesting that more engaging and media-based interventions may be equally effective for men and women. Further, as well as different modes of delivery, these two camps of interventions also differed in their operationalisation and outcome assessment. Upon examining Jackson et al.'s (2014) findings, the moderating effect of gender was for *implicit* attitudes only - when *explicit* attitudes were measured, there was no moderating effect of gender. Likewise, the research

that did not find a moderating effect of gender (Pietri et al. 2017; Moss-Racusin et al., 2018; Zawadski et al., 2012) measured *explicit* gender bias or gender bias *literacy* (i.e., knowledge and understanding, not attitudes), which is also measured explicitly. Therefore, based on our explicit measures, we do not expect to see a moderating effect of gender on intervention effectiveness.

Present Study

In this study I pilot tested a treatment video, misinformation video, and control video as materials for future studies examining gender stereotypes in STEM (see Appendix A; Intervention Videos). As the intervention materials demonstrated satisfactory levels of comprehension, and the treatment and control videos demonstrated satisfactory levels of credibility, preliminary analyses were conducted on the intervention's effectiveness in debunking and inoculating against gender stereotypes in STEM. Additionally, exploratory analyses were conducted on treatment-elicited psychological mechanisms as well as gender differences.

Method

Participants

Participants (N = 119) aged between 18 to 75 (M = 35.45, SD= 12.64) were recruited through Prolific Academic (www.prolific.co), an online participant recruitment platform. It is considered superior to alternative online platforms, including CrowdFlower and MTURK, as its participants are more diverse, produce higher quality data, and provide more honest responses (Peer et al., 2017). Eligibility was limited to North American residents over 18 years of age and fluent in English. Six participants were excluded from analyses due to not completing the study, failure on the attention check and/or viewing less than 50% of the video. After data cleaning procedures, a final total of 113 participants remained. (Analyses were re-run with the inclusion of all participants and results did not meaningfully change.) Overall, participants were liberal (M = 4.95, SD = 1.40). Approximately half of the participants identified as female and only 12.4% of participants identified as non-White. Nearly half of the participants had an undergraduate university degree or higher (47.8%). Further demographic information is provided in Table 2.1. Participants received a remuneration of £0.75 for participating.

Procedure

Participants were invited to participate in the study through Prolific Academic. The recruitment link directed them to the online survey, hosted on Qualtrics Survey Platform. After reading the information sheet and providing informed consent, participants were randomly assigned to one of four conditions: treatment video, misinformation video, treatment video + misinformation video, or control video. After watching their assigned video (participants in the treatment + misinformation condition had watched only the treatment video at this stage), participants answered a series of questions about the video they had just watched. Those in the treatment + misinformation group then watched the misinformation video and again answered a series of questions about the video they watched. Finally, all participants completed a series of questionnaires on gender bias literacy and gender stereotype endorsement as well as demographic information. At the end of the survey, participants were invited to give feedback if they so wished. To conclude, participants were debriefed (Appendix A; Debrief form) and financially compensated for their participation. Ethical approval was provided by the Psychology Department at the University of Cambridge.

Table 2.1

Sample Demographics (N = 113)

	n	%
Ethnicity		
White	99	87.6
Black	3	2.7
Asian	6	5.3
Chinese	2	1.8
Mixed	3	2.7
Gender		
Female	60	53.1
Male	52	46.0
Prefer not to say	1	0.9
Age[a]		
18-25	23	20.4
26-30	27	23.9
31-35	16	14.2
36-40	12	10.7
41-45	6	5.4
46-50	6	5.4
51-55	15	13.3

56-60	2	1.8
61-65	5	4.5
66-70	0	0
71-75	1	0.9
Education		
No university degree	59	52.2
Undergraduate university degree	33	29.2
Graduate university degree	21	18.6

Note. [a] Age was collected as a continuous variable, but is presented as categorised data here for ease of interpretation.

Materials and Measures

Responses for each item were optional, and all items were randomised within measures. The order of presentation of measures was not randomised as this would compromise the study design by increasing the likelihood of priming effects and demand characteristics. All materials and measures were administered in Qualtrics Survey Platform and are included in Appendix A.

Intervention Videos

Three videos were created for the purposes of this research: a (debunking/inoculation) treatment video, a misinformation video, and a control video (see Appendix A; Intervention Videos). These videos were developed in iMovie software by utilising relevant clips from existing online videos and combining them with experiment-specific scripts and layered background music. All three videos went through several rounds of feedback and revisions. Feedback and revisions were provided by five University of Cambridge scientists, with expertise in social psychology, gender development, or STEM, as well as five lay-persons without academic or STEM backgrounds.

Treatment Video. The treatment video (4 minutes 7 seconds) was designed with the aim to 'treat' the inaccurate stereotype that men are innately better suited for STEM than women. Following the principles of debunking and inoculation theory, the video begins with an expert presenting the scientific consensus ("It's not a lack of ability causing women to not pursue careers in [STEM] fields.") and debunking the myth that women perform worse than men in STEM. This is followed by a 'warning' message in text with ominous background instrumental music: "Some people use misleading tactics to try to convince others that men

are innately better at STEM." The video continues with clips of animation, expert interviews, narrative scenes, and text to specifically explain the flaws in two misleading tactics and provide refutational scientific facts:

Misleading tactic 1: Exaggerate the size of biological differences between men and women and incorrectly link them to STEM performance

Scientific fact 1: There are few, specific mental tasks that show small or negligible gender differences, and these have not been found to relate to STEM suitability

Misleading tactic 2: Argue STEM ability is inherent and related to one's gender

Scientific fact 2: The brain changes in response to experiences. When people do STEM, their brains change to make them better at STEM. There is no reliable scientific evidence that men are naturally better at STEM.

Banas & Rains (2010) meta-analysis demonstrates that passive refutations (provided by the messenger) are no less effective than active refutations (produced by the participant).

Figure 2.1

Four Screengrabs from the Treatment Video

Misinformation Video. The misinformation video (30 seconds) was designed with the aim to promote inaccurate myths about innate gender differences in STEM ability. It is narrated by a female to improve source credibility and representation of the targeted gender

group. This ensures that any potential confounding effects of a male narrator, such as perceived male privilege or outgroup bias, were eliminated. Further, as women are less gender biased than men (United Nations Development Programme, 2020), they may be seen as more credible and less biased when presenting the information.

Figure 2.2

Two Screengrabs from the Misinformation Video

Photo of screengrab removed for copyright reasons.	Photo of screengrab removed for copyright reasons.

Control Video. The control video (3 minutes 39 seconds) was designed with the aim to expose participants to STEM material without gender-stereotype information. As such, this was a no-treatment control video. The video discusses protein synthesis and enzymes in humans. It includes both male and female scientists as narrators to ensure there is no confounding effect of expert gender. However, this exposure to a female scientist may have primed participants and influenced results. Given the gender disparity in STEM, a male-only narration may have been more reflective of the real-world and elicited more ecologically valid responses.

Figure 2.3

Two Screengrabs from the Control Video

Photo of screengrab removed for copyright reasons.	Photo of screengrab removed for copyright reasons.

Measures

Attention Check. Participants completed a multiple-choice question, "What was the topic of the video you watched?" with the following response options, (a) racism in film, (b) organic farming in Europe, (c) protein synthesis in humans, and (d) gender stereotypes in STEM.

Active Inoculation. After viewing the video, participants were asked to give a free-response answer to the question, "In 1-2 sentences, what will you remember most about the video you watched?" This activity was used to encourage active inoculation, specifically active recall of refutation. Responses were not qualitatively analysed in this thesis. Active inoculation interventions, versus passive ones, are hypothesized to be more effective against novel counterarguments as they involve greater cognitive processing and participation *(M*cGuire, 1961; Pfau, 1997; Roozenbeek & van der Linden, 2018).

Video Comprehension. Comprehension of the videos was assessed with five face-valid semantic differential items anchored by bipolar adjectives on a 7-point scale: *not too slow - too slow; not too fast - too fast; not too much information - too much information; easy to follow - confusing to follow; understandable wording - complicated wording.* Therefore, contrary to most measures, lower scores indicate greater comprehension. Because these items demonstrated poor internal consistency when evaluated together ($\alpha = 0.56$), they were analysed separately rather than combined into a composite variable.

Video Credibility. Following research by Miller et al. (2013), the credibility of the videos was measured by five semantic differential items anchored by bipolar adjectives on a 7-point scale: *not scientific - scientific; unpersuasive - persuasive; inaccurate - accurate; not trustworthy - trustworthy; not credible - credible.* The composite score demonstrated excellent internal consistency, $\alpha = 0.95$.

Engaging. To measure how engaging the video was, two semantic differential items bound by bipolar adjectives on a 7-point scale were assessed: *not at all interesting - interesting; not at all entertaining - entertaining.* The composite score demonstrated excellent internal reliability, $\alpha = 0.90$.

Counterarguing. Participants' level of counterarguing in response to the message in the video was indicated on a 1-item 7-point response, from 1 (*I accepted a lot of the arguments offered)* to 7 (*I thought of arguments against) (M*iller et al., 2013).

Treatment-Elicited Threat. Following inoculation research (e.g., Miller et al., 2013) treatment-elicited threat was assessed using a 7-point bipolar semantic item anchored by bipolar adjectives, *non-threatening* (1) to *threatening* (7), in response to "Please rate your feelings towards persuasive arguments designed to mislead you about women's competence in STEM."

State Logical Thinking. Logical thinking was assessed using an adapted version of Pietri and colleagues (2017) 4-item Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Participants responded to statements, "This video made me think logically about the evidence it presented" and "This video made me think critically about the evidence it presented". The scale demonstrated good internal consistency, $\alpha = .82$.

Perceived Increase in Knowledge of Gender stereotypes in STEM. Four items captured participants' *perceived* increase in knowledge of gender stereotypes in STEM (adapted from Pietri et al., 2017). Items were statements rated on a Likert scale from 1 (s*trongly disagree*) to 7 (s*trongly agree*). Examples of items include, "This video increased my knowledge of the facts and evidence about gender stereotypes in STEM" and "After watching this video, I feel like I know more about research on gender stereotypes in STEM". The scale demonstrated excellent internal consistency, $\alpha = .95$.

State Perspective Taking. Face-valid items on a Likert scale ranging from 1 (*strongly disagree*) to 7 (s*trongly agree*) comprised the perspective taking measure. Items included, "The video made me think about what it would be like to be negatively stereotyped" and "The video made me think about how I would feel being negatively stereotyped". The scale demonstrated excellent internal consistency, $\alpha = 0.93$.

State Empathy. State empathy was measured with two items, "The video made me feel empathy for women in STEM" and "The video made me feel concerned about gender stereotypes in STEM", on a Likert-scale ranging from 1 (*strongly disagree*) to 7 (s*trongly agree*). The composite score demonstrated very good internal consistency, $\alpha = .91$.

Discernment of Gender-STEM Stereotypes (DOGSS). To measure discernment of gender stereotypes in STEM, six face-valid 7-point Likert-items, from 1 (*strongly disagree*) to 7 (s*trongly agree*), were derived from the messages in the treatment video. The items demonstrated good bivariate correlations (|.69| <*r*s < |.72|) and internal consistency (α = .76).

The set of items were constructed such that three were 'correct' statements that measured evidence-based discernment of gender stereotypes in STEM (e.g., "The scientific consensus is that there are no innate gender differences in STEM ability") and the other three were 'incorrect' statements, such that they targeted endorsement of gender stereotypes in STEM (e.g., "On average, men's and women's brains function differently"). To test the internal validity of the factor structure, a Confirmatory Factor Analysis (CFA) was run. Findings revealed a very good model fit, χ^2 (8) = 18.13, p = .02 (CFI = .95, TLI = .91, RMSEA = .10, SRMR = 0.07). The factor correlation was .46, p < .001. Due to good internal validity and reliability, a total composite score of discernment was created (with the three 'incorrect' items reverse-coded). When tests of main effects are broken out at item-level, corrections for multiple comparisons are applied.

Male-agentic Stereotype. Participants completed agentic traits from the Gender Stereotype Index (Rudman & Glick, 2001) to indicate their belief in agency as a gendered trait. Four agentic traits ("individualistic", "competitive", "independent", "self-sufficient") were rated on a scale from –3 (*more true of women*) to 3 (*more true of men*), such that higher scores indicate more traditional gender-trait beliefs. The scale demonstrated very poor internal consistency, α = 0.31.

A low Cronbach's alpha can be due to too few items, a lack of unidimensionality, and/or poor inter-relatedness between items (Gardner, 1995; Tavakol & Dennick, 2011). The issue of too few items is unlikely as the scale has four items, and three items is the minimum number recommended to yield a reliable Cronbach's alpha for an underlying construct (Eisinga et al., 2012). To test the inter-relatedness between items, the correlations between each item and the total scale score were computed. Traditionally, items with low scale correlations (close to 0) are removed to improve the internal consistency of the total scale (Tavakol & Dennick, 2001). However, all item-total bootstrapped correlations demonstrated *rs* between .44, 95% BCa CI [0.20, 0.63], and .68, 95% BCa CI [0.55, 0.79], with all *p*s <.001.

Researchers such as Schmitt (1996) and Cortina (1993) have argued that scales with a low alpha value can still be useful in some cases. For instance, if researchers believe the items have 'face equivalence' (i.e., all target the same underlying construct), they may choose to accept a lower internal reliability (Taber, 2017). However, this is typically in reference to Cronbach's alphas closer to .5. Further, the original publication of the scale did not report internal reliability or scale validation (Rudman & Glick, 2001). Therefore, the four agentic items were examined separately, not as a composite score.

Political Orientation. In line with previous social psychological research (e.g., Roozenbeek & van der Lindeen, 2019), political orientation was measured on a 7-point scale from 1 (*very conservative)* to 7 (*very liberal)*.

Analyses

While it would provide interesting data to measure pre-post differences within each intervention condition, this could create some methodological concerns. These include priming, anchoring, and fatigue effects, which can impact psychological change. As such, differences in outcome measures were analysed between conditions. When tests of main effects are broken out at item-level, corrections for multiple comparisons are applied. Given the small sample size in each condition (n = 27 to 30), the following analyses are preliminary and exploratory in nature.

To overcome violations of the assumption of homogeneity of variance, the more robust Welch's test was run and interpreted for all ANOVAs and t-tests. Field (2017) advises using the Welch's test in all situations, regardless of homogeneity of variance, as it is a more robust test. For non-normally distributed data, non-parametric t-tests Mann-Whitney U and Wilcoxon Signed-Rank tests are reported for between-group and within-group comparisons, respectively. Welch's ANOVAs are fairly robust against violations of normality (Field, 2017); however, as a robustness check, non-parametric Kruskal-Wallis tests were run when the data were skewed, and there were no meaningful differences between the tests. Thus, Welch's ANOVAs are reported throughout. For the MANOVA, the Wilks' lambda test statistic is reported as it is most appropriate for heterogeneous variance. As a robustness check, Pillai's trace was also run, and no differences were found between Wilks' lambda and Pillai's trace.

For post-hoc procedures, the Tukey method is reported as it is considered more powerful than the Bonferroni method when there are a larger number of comparisons made (Field, 2017). However, as the Bonferroni method is considered slightly more conservative, it was examined as a robustness check. No differences were found between Tukey and Bonferroni for any post-hoc analysis. To further increase confidence in findings, the Games-Howell procedure was run, due to the uncertainty of knowing whether the population variances were equal (Field, 2017). Again, the Games-Howell procedures provided the same results. Thus, Tukey is reported throughout.

Lastly, bootstrapping, a common robust estimation method (Efron & Tibshirani, 1993), was applied to all analyses. Bootstrap parameter estimates help to ensure the model parameters are robust and provide 95% confidence intervals of the parameter.

All analyses were conducted in SPSS (versions 25, 26, and 27) and JASP (versions 0.12 and 0.16)

Results

First, demographic variables were investigated to ensure there were no significant between-groups differences. Welch's *t*-tests and a chi-square test confirmed that the four experimental conditions did not differ on levels of educational attainment ($F(3, 109) = 0.13$, $p=.943$), age ($F(3, 109) = 2.04$, $p=.113$), political orientation ($F(3,109) = .21$, $p = .887$) or gender identity ($\chi^2(3)=4.38$, $p=.223$). Therefore, in all further analyses, parsimonious models without these variables were estimated.

Intervention Videos

Levels of comprehension, perceived credibility, and engagingness of the intervention videos were examined to ensure suitability of the videos as intervention materials for future studies. Next, preliminary analyses on the psychological mechanisms evoked by the treatment video were examined.

Comprehension

Video comprehension was acceptable across videos (see Table 2.2). With the exception of the misinformation video on *not too fast - too fast* and the control video on *not too much information - too much information*, all other comprehension ratings were below 4 (midpoint on the scale). Between-group comparisons are provided in Table 2.2.

Table 2.2

Descriptive Statistics for Comprehension Items by Experimental Video

Comprehension items	Treatment (n=56)[o]		Misinformation (n=30)		Control (n =27)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Comprehension						
Not too slow - Too slow	3.39_{a}	1.61	2.67_{a}	1.40	3.48_{a}	1.67
Not too fast - Too fast	2.95_{a}	1.31	4.53_{b}	1.70	3.67_{a}	1.52
Not too much information - Too much information	3.29_{b}	1.39	2.20_{a}	1.16	4.48_{c}	1.34
Easy to follow - Confusing	2.75_{a}	1.38	2.93_{a}	1.87	3.56_{a}	1.74
Understandable - Complicated	2.14_{a}	1.21	2.60_{ab}	1.45	3.26_{b}	1.66

Note. [a] The treatment video values includes ratings from both the debunking condition and the inoculation condition, as participants in these two conditions were exposed to the same intervention material (i.e., only the treatment video) prior to ratings. Shared subscripts indicate no significant difference between intervention conditions based on Tukey's post-hoc tests.

Credibility

The control video *(M* = 5.70, *SD*= 0.97) and treatment video *(M* = 5.19, *SD*= 1.06) were perceived as credible, however, the misinformation video *(M* = 3.02, *SD*= 1.58) was not. A one-way ANOVA revealed the misinformation video was rated significantly less credible than both the treatment video *(M_{diff}* = -2.17, 95% CI [-2.770, -1.482], *t(*56.21) = -7.96, p_{tukey} < .001, *d* = -1.711) and control video *(M_{diff}* = -2.67, 95% CI [-3.323, -1.967], *t(*56.21) = -8.37, p_{tukey} < .001, *d* = -2.017), *F(*2, 56.21) = 30.87, *p* <.001, η2 = .439. The misinformation video's low credibility ratings may be in part due to the participants' pre-intervention resistance to inaccurate gender stereotypes in STEM, suggested by the low levels of gender-STEM stereotype endorsement found across all conditions, as discussed later on.

Engaging

Similarly, the misinformation video *(M* = 3.45, *SD*= 1.70) was rated as significantly less engaging than the treatment video *(M_{diff}* = -1.29, 95% CI [-2.023, -0.579], *t(*56.08) = -3.70, p_{tukey} < .001, *d* = -0.841) and control video *(M_{diff}* = -1.118, 95% CI [-1.957, -0.227], *t(*56.08) = -2.73, p_{tukey} = .020, *d* = -0.686), *F(*2, 56.08) = 6.35, *p* = .003, η2 = .116. The means of the control video *(M* = 4.57, *SD*= 1.56) and treatment video *(M* = 4.75, *SD*= 1.46), along with their non-significant group differences *(p* = .880) suggest they were both suitably engaging intervention videos. Again, the misinformation video's low engaging rating

may be related to the sample's overall low levels of gender-trait stereotype endorsement, which are discussed in more detail further on.

Psychological Mechanisms

Preliminary exploratory analyses on the cognitive and affective psychological mechanisms evoked in each intervention video were examined. As the debunking (treatment-only) condition and inoculation (treatment + misinformation) condition were exposed to the same treatment material and measures for the first phase of the study, these two conditions were collapsed for the following analyses. To note, however, the inoculation condition engaged in significantly less logical thinking than the debunking condition (M_{diff} = -.58, p=.014, d = -.683). Additionally, differences in ratings of credibility approached significance, with the inoculation condition perceiving the video as less credible than the debunking condition (M_{diff} = -.50, p = .07, d = -.485). Nonetheless, these are considered random group differences, as the two groups were randomly assigned and had been exposed to identical treatment processes thus far.

State Logical Thinking. A one-way Welch's ANOVA showed significant group differences in state logical thinking, F (2, 51.83) = 14.78, p<.001. Post hoc comparisons using the Tukey HSD test indicated that the average score for logical thinking in the misinformation video (M = 3.94, SD= 1.36) was significantly lower than in the treatment video (M = 5.44, SD= 0.905) and control video (M = 5.01, SD= 1.12). However, the treatment video and control video did not differ in state logical thinking (p =.221). This may be because the logical thinking questions were worded in a general way (e.g., "made me think logically"), not specific to gender-STEM stereotypes (e.g., "made me think logically about gender stereotypes in STEM").

State Perspective Taking. A one-way Welch's ANOVA revealed significant group differences in state perspective taking F (2, 57.53) = 90.71, p<.001. The Tukey HSD post hoc procedure indicated significant differences between all video conditions (see Table 2.3), with the control video eliciting the least perspective taking and the treatment video eliciting the most.

State Empathic Concern. A one-way Welch's ANOVA revealed significant group differences in state empathic concern F (2, 57.53) = 90.71, p<.001. The Tukey HSD post hoc procedure indicated significant differences between all video conditions (see Table 2.3), with the control video eliciting the least empathy and the treatment video eliciting the most.

Perceived Increase in Knowledge of Gender Stereotypes. A one-way Welch's ANOVA revealed significant group differences in perceived increase in knowledge of gender stereotypes in STEM, $F(2, 57.10) = 52.38$, $p < .001$, $\eta 2 = .488$. Post hoc procedures indicated significant differences between all video conditions: the treatment video ($M = 5.21$, $SD = 1.35$) produced the greatest perceived increase in knowledge of gender stereotypes, followed by the misinformation video ($M_{diff} = 1.52$, 95% CI [0.773, 2.280], $t(57.10) = 4.81$, $p_{tukey} < .001$, $d = 1.093$) and lastly the control video ($M_{diff} = 3.330$, 95% CI [2.549, 4.111], $t(57.10) = 10.13$, $p_{tukey} < .001$, $d = 2.421$). Although participants in the misinformation condition perceived a significantly greater increase in knowledge of gender stereotypes than the control condition ($M_{diff} = 1.80$, 95% CI [0.920, 2.688], $t(57.10) = 4.848$, $p_{tukey} < .001$, $d = 1.244$), the mean value ($M = 3.68$, $SD = 1.47$) still fell below the midpoint on the scale, between 'slightly disagree' and 'neither agree nor disagree'.

Relative to the control video, with the exception of logical thinking, the findings suggest the treatment video elicited typical stereotype-reduction psychological mechanisms. The lack of group differences between the control and intervention video on logical thinking is likely due to the cognitively demanding and scientific nature of the control video's content. Alternatively, relative to the misinformation video, the treatment video elicited significantly higher logical thinking. Additionally, the misinformation video, like the treatment video, elicited greater state empathy, perceived increase in knowledge, and perspective taking than the control video. However, participants who watched the misinformation video engaged in less logical thinking than those who watched the control video. This is discussed further in the discussion section.

Inoculation Checks

Next, group comparisons were conducted between the inoculation condition and misinformation condition on ratings of perceived credibility and counterarguing in response to the misinformation video, as well as the role of threat in motivating counterarguing. These analyses test the efficacy of the treatment video as an inoculation treatment against misinformation.

Credibility

The credibility of the misinformation video was not rated significantly lower among those who watched it after the treatment video *(M* = 3.01, *SD*= 1.50) compared to those who watched only the misinformation video *(M* = 3.02, *SD*= 1.58), p = .962. The average rating for both conditions was 'slightly disagree'. This indicates the sample may have been resistant to the misinformation messages, even without the treatment message. This may be explained in part by the highly-educated (close to 50% had a university degree) and politically liberal *(M* = 4.95, *SD*= 1.40) demographics of the sample – and is further suggested by the sample's overall low endorsement of gender stereotypes as discussed below in Intervention Effects and the Discussion.

It could also indicate that the treatment video was successful in inoculating, i.e., protecting, prior 'correct' beliefs, but was not successful in increasing resistance in the "right" direction, as the average credibility rating of the misinformation video could have been lower among the inoculation group (i.e., there was no issue of floor effects).

Counterarguing

Those who watched the treatment video before the misinformation video (inoculation condition) engaged in significantly greater counterarguing *(M* = 5.10, *SD*= 1.59) in response to the misinformation video than those who watched the misinformation video only *(M_{diff}* = 0.89, 95% CI [-0.007, 1.732], *t(*55.93) = 2.04, p_{tukey} = .045, d = 0.538). This group difference suggests the treatment video successfully elicited counterarguing, an important element in psychological inoculation.

Treatment-Elicited Threat

To test the role of traditional treatment-elicited threat in predicting counterarguing, a linear regression was run on participants in the inoculation condition. As expected, treatment-elicited threat did not demonstrate a linear relationship with counterarguing (p = .852). Additionally, an ANOVA demonstrated no differences between conditions on treatment-elicited threat (p = .924). This supports more recent inoculation literature, suggesting scholars move away from the original conceptualisation of fear-based threat and towards a motivation-based threat (e.g., Compton & Ivanov, 2014; Banas & Richards, 2017; Compton & Pfau, 2005).

Intervention Effects

Given the intervention videos' comprehension ratings and successful inoculation checks of the treatment video, analyses on the intervention's main effects were performed. Due to the misinformation video's relatively low credibility ratings, findings and interpretations should be taken with caution.

To ensure there were no significant differences between the debunking and inoculation groups on relevant measures after viewing the treatment video, independent Welch's *t*-tests were run. No between-group differences were found on levels of comprehension or engaging (*p*s> .08). However, compared to the debunking condition, the inoculation condition engaged in significantly less logical thinking (M_{diff} = -.58, p=.018, d = -.683) and state empathy (M_{diff} = -.96, p=.008, d = -.735). Additionally, differences in ratings of credibility approached significance, with the inoculation condition perceiving the video as less credible than the debunking condition (M_{diff} = -.50, p = .074, d = -.485). Each of these ratings were provided immediately after watching the treatment video, meaning the two groups of participants had been exposed to the exact same study material thus far. As such, it appears there were some random group differences between the two conditions. This may help to explain some of the findings below.

Discernment of Gender Stereotypes in STEM

Intervention effects on the main outcome variable, discernment of gender stereotypes in STEM, were examined first. As an initial omnibus test, a one-way MANOVA on the effect of intervention condition on the six DOGSS demonstrated significance, F (18, 291.81) = 3.38, Wilk's Λ = 0.585, partial η^2 = .164, p<.001. To gain a deeper understanding of the discernment scores, follow-up one-way ANOVAs were conducted and presented below. Three of the six DOGSS items demonstrated group differences (see descriptive statistics in Table 2.3).

Table 2.3

Descriptive Statistics for Measures by Intervention Condition

	Inoculation (n=28)		Debunking (n=28)		Misinformation (n=30)		Control (n=27)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
DOGSS								

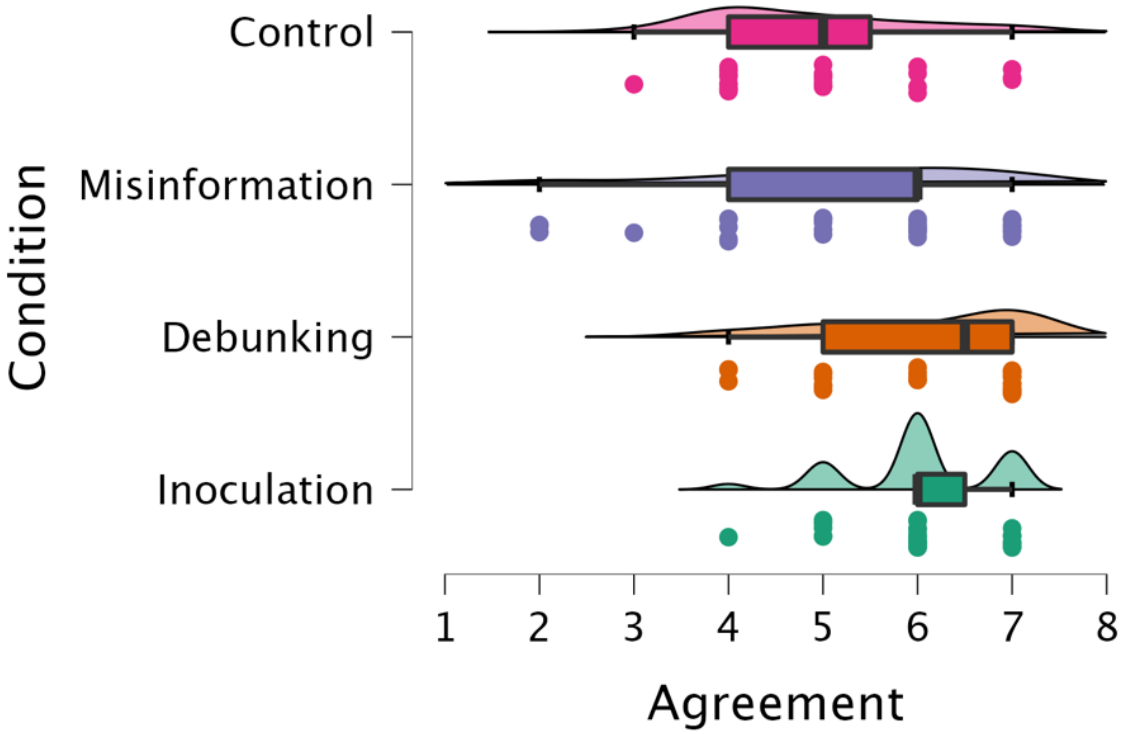
Scientific consensus	5.85_{ab}	1.11	6.14_{a}	1.11	4.60_{c}	1.38	5.00_{bc}	1.49
Not naturally interested	3.46_{a}	1.89	2.92_{a}	1.69	3.53_{a}	1.54	2.96_{a}	1.34
Brains function differently	3.28_{ab}	1.71	2.78_{a}	1.59	4.20_{b}	1.97	4.03_{b}	1.62
Men innately better	3.00_{a}	1.54	2.42_{a}	1.37	3.26_{a}	1.50	2.96_{a}	1.40
Use misleading tactics	6.00_{ab}	0.78	6.07_{a}	1.08	5.23_{bc}	1.56	4.85_{c}	1.09
Hire/promotion effects	5.85_{a}	1.49	6.10_{a}	1.13	5.60_{a}	1.38	5.33_{a}	1.46
Discernment total score	5.32_{ab}	0.81	5.70_{a}	0.99	4.74_{b}	1.04	4.87_{b}	0.83
Male-Agentic								
Individualistic	-0.07_{a}	0.94	-0.07_{a}	0.76	-0.13_{a}	0.81	0.14_{a}	0.71
Competitive	0.57_{a}	0.92	1.00_{a}	0.94	0.63_{a}	0.85	0.88_{a}	1.01
Independent	0.32_{a}	0.90	0.03_{a}	0.88	0.03_{a}	0.92	0.07_{a}	0.78
Self-sufficient	0.03_{a}	0.79	-0.28_{a}	1.18	-0.10_{a}	1.24	0.07_{a}	0.82
Knowledge Increase	4.98_{a}	1.40	5.43_{a}	1.29	3.68_{b}	1.47	1.88_{c}	1.42
State Logical Thinking	5.15_{a}	0.91	5.74_{a}	0.80	3.94_{b}	1.36	5.01_{a}	1.11
State Perspective Taking	5.10_{ab}	1.34	5.28_{a}	1.19	4.34_{b}	1.57	1.74_{c}	1.07
State Empathy	4.96_{ab}	1.47	5.92_{a}	1.12	4.10_{b}	1.95	1.77_{c}	1.12

Note. Shared subscripts indicate no significant difference between intervention conditions based on Tukey's post-hoc tests. Inoculation condition = treatment video + misinformation video; Debunking condition = treatment video only. DOGSS = Discernment of Gender-STEM Stereotypes; Knowledge Increase = Perceived Increase in Knowledge of Gender Stereotypes in STEM

A Welch's ANOVA on "Some people use misleading tactics to falsely claim that men are better at STEM than women" revealed a significant main effect of condition ($F(3, 59.13) = 8.43$, $p < .001$, $\eta^2 = .162$; see Figure 2.4). Bootstrapped post hoc comparisons indicated that both the inoculation condition ($M_{diff} = 1.15$, 95% CI [0.552, 1.676], $t(59.13) = 3.57$, $p_{tukey} = .003$, $d = 1.202$) and the debunking condition ($M_{diff} = 1.208$, 95% CI [0.604, 1.768], $t(59.13) = 3.82$, $p_{tukey} < .001$, $d = 1.116$) agreed significantly more with the statement than the control group. Additionally, the debunking condition agreed significantly more with the statement than the misinformation condition ($M_{diff} = 0.82$, 95% CI [0.194, 1.525], $t(59.13) = 2.70$, $p_{tukey} = .040$, $d = 0.617$). The inoculation condition demonstrated a non-significant trend in the same direction with the misinformation condition, ($p_{tukey} = .074$, $d = 0.608$). All other group comparisons (inoculation – debunking; inoculation – misinformation; misinformation – control) were non-significant (all $ps > .617$). In sum, the treatment video had a significant persuasive effect (increase in the 'correct' direction in the debunking condition) and, to a lesser extent, inoculative effect, as the misinformation video attenuated some of the treatment video's persuasive effect, but not to the level of the control group.

Figure 2.4

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "Some people use misleading tactics to falsely claim that men are better at STEM than women"

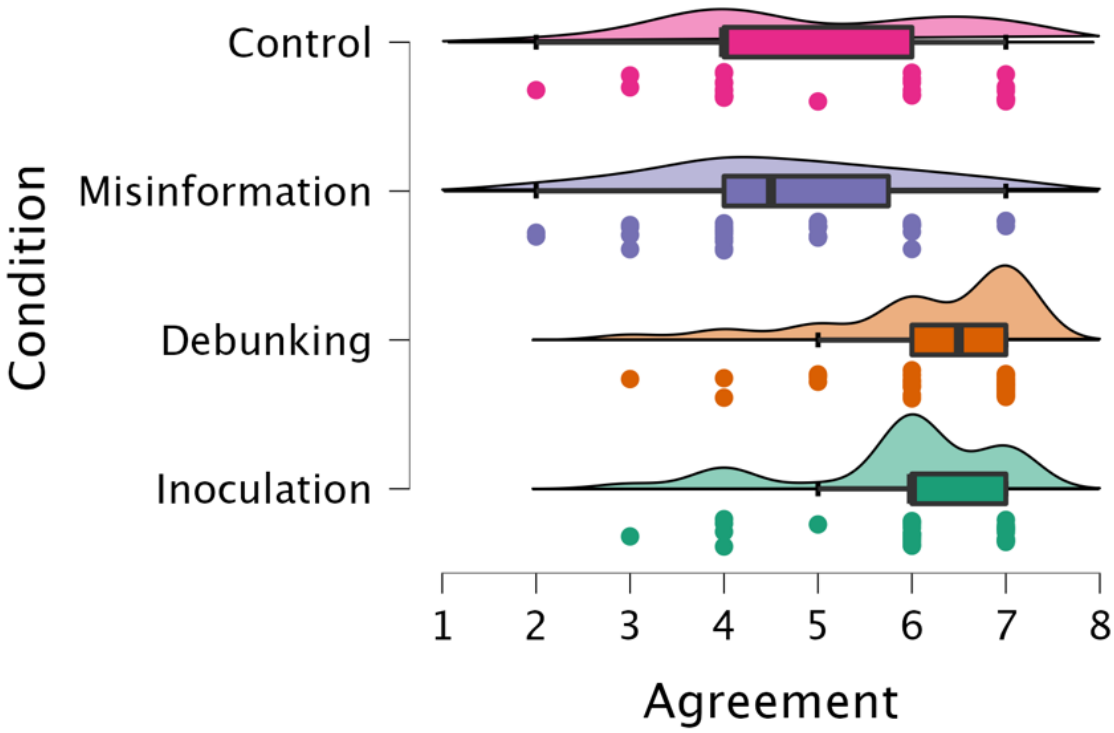


Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges. The possible range of scores is 1 to 7.

A Welch's ANOVA on "The scientific consensus is that there are no innate gender differences in STEM ability" revealed a significant main effect of condition ($F(3, 60.00) = 9.11$, $p < .001$, $\eta^2 = .199$l; see Figure 2.5). Bootstrapped post hoc comparisons indicated that both the inoculation condition ($M_{diff} = 1.25$, 95% CI [0.591, 1.906], $t(60.00) = 3.72$, $p_{tukey} = .002$, $d = 0.999$) and the debunking condition ($M_{diff} = 1.52$, 95% CI [0.904, 2.201], $t(60.00) = 4.56$, $p_{tukey} < .001$, $d = 1.226$) agreed significantly more than the misinformation condition. Further, the debunking condition agreed significantly more with the statement than the control condition ($M_{diff} = 1.15$, 95% CI [0.406, 1.822], $t(60.00) = 3.29$, $p_{tukey} = .007$, $d = 0.870$). The inoculation condition demonstrated a non-significant trend in the same direction, ($p_{tukey} = 0.07$, $d = 0.653$). All other group comparisons were non-significant ($ps > .645$). Once again, there was a significant persuasive effect (higher scores in the debunking condition) and, to a lesser extent, inoculative effect, as the misinformation video reduced some of the treatment video's persuasiveness, but not to the level of the control group.

Figure 2.5

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "The scientific consensus is that there are no innate gender differences in STEM ability"

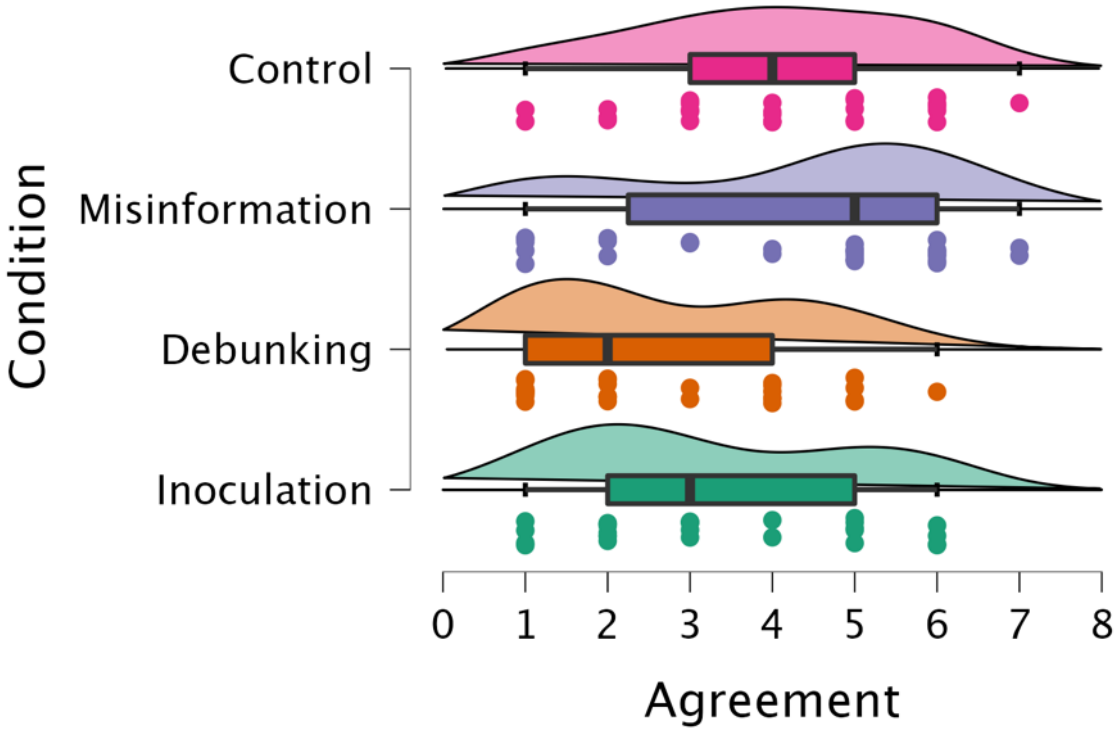


Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges. The possible range of scores is 1 to 7.

A Welch's ANOVA on "On average, men's and women's brains function differently" revealed a significant main effect of condition ($F(3, 60.51) = 4.10$, $p = .01$, $\eta^2 = .101$; see Figure 2.6). Bootstrapped post hoc comparisons indicated that the debunking condition agreed significantly less with the statement than the misinformation condition ($M_{diff} = -1.44$, 95% CI [-2.277, -0.428], $t(60.51) = -3.092$, $p_{tukey} = .013$, $d = -0.78$) and control condition ($M_{diff} = -1.25$, 95% CI [-2.079, -0.418], $t(60.51) = -2.66$, $p_{tukey} = .043$, $d = -0.776$). Meanwhile, the inoculation condition did not significantly differ from any other condition (all *ps*>.195). The misinformation and control conditions also did not significantly differ from each other ($p = .985$). Therefore, the treatment video provided a successful persuasive effect (significantly lower agreement in the debunking condition), but not an inoculative effect, as the misinformation video seemed to 'neutralise' the treatment video's effect.

Figure 2.6

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "On average, men's and women's brains function differently"



Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges. The possible range of scores is 1 to 7.

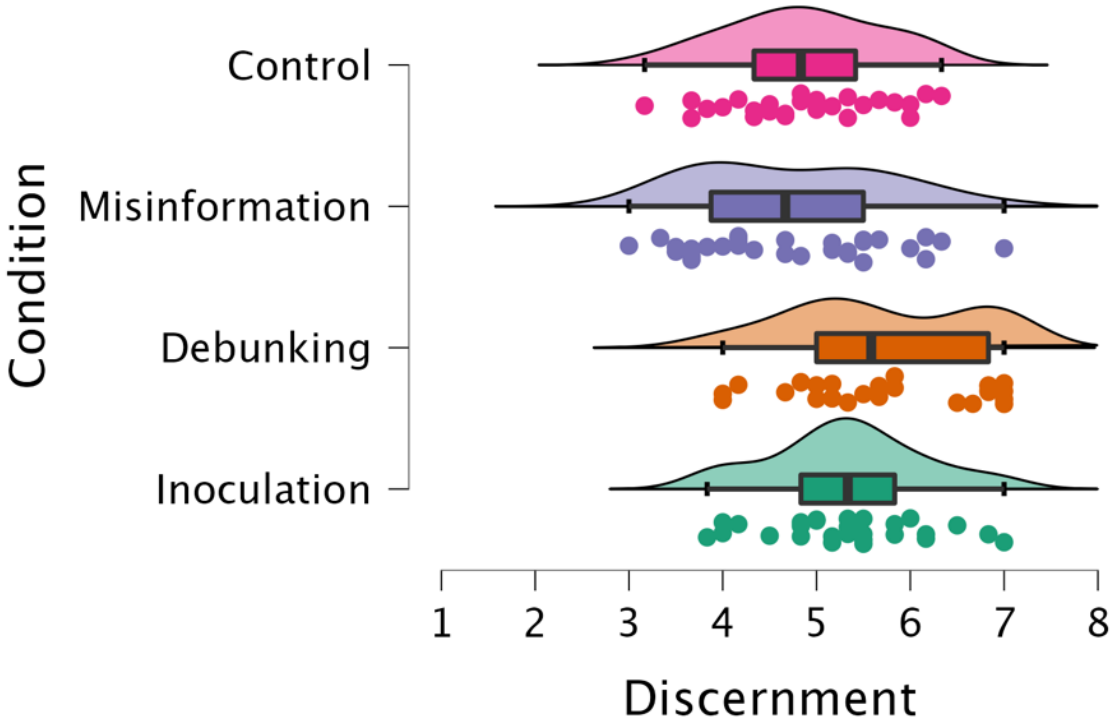
The remaining three discernment items did not reveal a significant main effect of condition: "Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM" ($F(3, 60.07) = 1.78$, $p = .159$, $\eta^2 = .049$); "More females are not in STEM because they are naturally interested in other things" ($F(3, 60.11) = 1.13$, $p = .341$, $\eta^2 = .029$); and "Men are innately better at STEM" ($F(3, 60.43) = 1.71$, $p = .174$, $\eta^2 = .044$).

Finally, as a total composite score of discernment, a Welch's ANOVA demonstrated significance, $F(3, 60.41) = 5.72$, $p = .002$, $\eta^2 = .149$ (see Figure 2.7). Bootstrapped post hoc comparisons indicated that the inoculation condition and debunking condition did not differ ($p_{tukey} = .440$). Additionally, the inoculation condition did not differ from the control condition ($p_{tukey} = .273$) or the misinformation condition ($p_{tukey} = .082$). The misinformation condition and control condition also did not significantly differ from each other ($p_{tukey} =$

.952). However, the debunking condition demonstrated significantly more discernment than the control condition (M_{diff} = 0.826, 95% CI [0.174, 1.478], t(60.51) = 3.30, p_{tukey} = .007, d = .892) and the misinformation condition (M_{diff} = 0.958, 95% CI [0.322, 1.593], t(60.51) = 3.93, p_{tukey} < .001, d = 1.034). Therefore, overall, the treatment provided seemed to provide a successful persuasive effect (significantly higher discernment in the debunking condition), but not an inoculative effect on discernment of gender-STEM stereotypes. This may be because the misinformation video 'neutralised' the treatment video's effect.

Figure 2.7

Raincloud Plot showing Between-Condition Comparison of Discernment of Gender Stereotypes in STEM



Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges. The possible range of scores is 1 to 7.

Male-Agentic Stereotype

Next, we examined the generalised effect, also known as 'umbrella protection' or 'blanket of protection' (Papageorgis & McGuire, 1961; Parker et al., 2012), of the intervention on the male-agentic stereotype. As an initial omnibus test, a one-way MANOVA was conducted with experimental condition as the predictor variable and the four agentic

traits as the outcome variables. There was no main effect of intervention condition on endorsement of agentic gender stereotype traits, F (12, 280.74) = .90, Wilk's Λ = 0.904, partial η^2 = .033, p = .534. The sample mean of the composite score *(M* = 0.19, *SD*= .52) suggests participants had low gender stereotype endorsement in general. Indeed, the means and standard deviations for each condition suggest very little to no endorsement of agentic gender stereotype beliefs across all conditions (see Table 2.3). Due to the poor performance of the Gender Stereotype Index, a different measure of agentic traits will be used for future studies.

Gender x Condition Interaction Effects and Gender Differences

Finally, participant gender was explored for interactive and main effects.

Gender x Condition Interaction

In line with previous intervention research on explicit gender bias, gender did not interact with video condition to predict male-agentic trait endorsement *(p* = .075) or discernment of gender stereotypes in STEM (all items *ps*>.074).

Gender Differences

Contrary to Rudman and Glick (2001), we did not find that men scored higher on the GSI than women *(p* = .171). This may reflect a change in the male-agentic stereotype over the past 20 years and/or the GSI not being a valid construct of agency (as suggested by the poor internal consistency). However, women agreed significantly more than men with the statements, "Some people use misleading tactics to falsely claim that men are better at STEM than women" *(M_{diff}* = 0.80, 95% CI [0.336, 1.250]; Welch's *t(*106.63) = 3.47, *d* = 0.662, p_{tukey} < .001; Mann-Whitney *U* = 2109, *r* = .375, *p* < .001) and "Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM careers" *(M_{diff}* = 0.57, 95% CI [0.112, 1.034]; Welch's *t(*101.39) = 2.98, *d* = 0.469, p_{tukey} = .03; Mann-Whitney *U* = 1956, *r* = .254, *p* = .01). Overall, this suggests women tend to demonstrate greater understanding of how gender-STEM stereotypes perpetuate and their consequences on women in the workplace. The other DOGSS items did not demonstrate a main effect of gender (all *ps*>.114). Thus, women and men did not differ on discernment items related to the content of gender-STEM stereotypes (e.g., On average, men's and women's brains function differently) or the scientific consensus.

Discussion

Intervention Videos

Treatment Video

Overall, the inoculation video was rated as credible, comprehensible, and engaging. Importantly, it evoked greater empathy, perspective taking, and perceived increase in knowledge than the control and misinformation videos. Although it evoked greater logical thinking than the misinformation video, it did not differ from the control video, which may be because the logical thinking statements were general (e.g., "made me think logically about the evidence it presented") versus specific to gender-STEM stereotypes, and the control video was a scientific video. Overall, these findings suggest the treatment video is acceptable to use in future research as treatment material. Further, the treatment video was tested and validated as an inoculation treatment specifically.

Misinformation Video

The misinformation video received generally high comprehension ratings, but low credibility and engaging ratings. Additionally, participants who watched the misinformation video reported less state logical thinking than participants who watched the treatment video or control video. These findings suggest the video needs improvements in order to be acceptable as convincing myth material for future studies.

Of consideration, the misinformation video was quite short in length (~30 seconds) compared to the treatment video (~4 minutes) and was rated as 'too fast'. It is possible that the differing video lengths between the treatment video and the misinformation video may have impacted the misinformation video's lower ratings on credibility, engaging, and logical thinking and, thus, overall persuasiveness. Therefore, lengthening the misinformation video may improve credibility, logical thinking, and effectiveness of the video. Additionally, McGuire (1961) hypothesized that lengthening the exposure to the misinformation, to allow both active and passive inoculation, should produce greater resistance (in the inoculation condition). Increasing the length of the misinformation video by elaborating with supporting evidence should both provide more credibility to the claims being made in the video, as well as provide more time for the inoculation participants to process the information, develop counter arguments, and build resistance. Future research may benefit from testing different

and matched lengths of treatment and attack messages. To the best of my knowledge, this has not been previously studied or addressed in inoculation research.

Alternatively, the misinformation video's lower ratings on logical thinking, credibility, and engagingness may be in part due to pre-intervention resistance to gender stereotypes in STEM, as suggested by the sample's low levels of gender-STEM stereotype endorsement found across all conditions (including control).

Control Video

While the control video's comprehension ratings were on average lower than the inoculation and misinformation videos, the control video was rated as the most credible video. The control video's comprehension is not relevant to this research as the intention of the control video is not to increase STEM knowledge but to expose participants to STEM material without gender-related information. Together, these findings suggest the control video is appropriate for future studies.

Intervention Effects

Overall, results indicate that participants in the debunking condition demonstrated the greatest discernment of gender stereotypes in STEM. Additionally, there were no significant differences between the debunking condition and inoculation condition on the total DOGSS score or any of the individual items. However, the inoculation group did not differ from the misinformation and control groups to the extent that the debunking group did. As post-hoc comparisons and corrections were performed across all conditions (not only those of interest), and not all post-hoc comparisons were relevant or hypothesized, this impacted the reported statistical performance of the treatment conditions. This statistically conservative approach was taken due to not pre-registering specific contrasts of interest.

The debunking group demonstrating the greatest gender-STEM stereotype literacy is in line with stereotype-reduction research that finds significant change immediately following an intervention treatment. As participants in psychological intervention research most likely face conflicting information post-intervention, the inoculation group may have been more representative of the treatment video's real-world effects.

These findings may also be due in part to the random difference found between the debunking and inoculation groups. In particular, the inoculation group perceived the treatment video as less credible and less evidence-based than the debunking group. (These ratings were provided immediately after watching the treatment video, *prior* to being exposed

to the misinformation, so the inoculation and debunking conditions had been exposed to identical material thus far.) Credibility is strongly associated with perceiving information in a message as more valid (Briñol et al., 2004; Kaufman et al., 1999). Therefore, the random group differences between the debunking and inoculation conditions may contribute to some of the findings.

Conversely, there were no treatment effects on endorsement of the male-agentic gender stereotype, which may be due in part to the floor effects on the Gender Stereotype Index, as suggested by the control and misinformation groups' values. This may be interpreted as the continued influence effect, in which new discernment does not translate to related attitudes and behaviours (Chan et al., 2017; Cook et al., 2017; Lewandowsky et al, 2012; Ecker et al., 2010; Schwarz et al., 2016; Pluviano et al., 2017). Alternatively, these findings may be related to the sample being overall highly educated (nearly half had a university degree) and liberal in political orientation *(M* = 4.95, *SD*= 1.40). These sample demographics reduce the representativeness and therefore generalisability of the findings.

The lack of intervention effects on endorsement of the male-agentic stereotype, alongside some significant intervention effects between the control group and treatment groups on discernment of gender stereotypes in STEM, suggests that gender stereotypes and their effects on beliefs and attitudes may be multi-dimensional and differentially influenced.

CHAPTER THREE – EXAMINING THE PERSUASIVE AND INOCULATIVE EFFECTS OF A VIDEO INTERVENTION TARGETING GENDER STEREOTYPES IN STEM

In this chapter, I sought to investigate the effects of the novel gender-STEM stereotype intervention developed and tested in Chapter 2. Specifically, I examined the persuasive and inoculative effects of the intervention on gender-STEM stereotype discernment as well as more distal, attitudinal, and behavioural measures of gender stereotypes. Based on the findings in Chapter 2, the same treatment and control videos were used in the present study, and the misinformation video was updated to improve credibility.

Introduction

Scientific evidence suggests that gender stereotypes significantly contribute to the gender disparity in STEM (Eagly, 2020; Ellemers, 2018; Hyde, 2014), including hiring bias against women (Moss-Racusin et al., 2012; Glick et al., 1988; Heilman, 1984; Lindberg et al., 2010) To establish greater gender diversity in domains such as STEM, identifying and understanding effective strategies to reduce the application of negative gender stereotypes is needed. Current intervention literature suggests the combination of cognitive and affective mechanisms is more effective than cognitive or affective alone in stereotype change. However, the relative contributions of each, and which specific mechanisms are effective, is not well understood.

Gender-STEM Hiring Bias

The inaccurate stereotype that men are better suited for STEM contributes to a hiring bias in favour of men (Glick et al., 1988; Heilman, 1984; Isaac et al., 2009; Moss-Racusin et al., 2012). Steinpreis et al. (1999) found that both male and female academics were more likely to vote to hire a male applicant over a female applicant with the exact same curriculum vitae, which came from a real-life scientist. Moss-Racusin and colleagues (2012) replicated these findings with science faculty, and also found that the male applicant was rated as significantly more competent, hireable, and mentorable, and offered a higher starting salary than the identical female applicant. Similarly, a more recent study by Reuben et al. (2014) found that both men and women were twice as likely to hire a man for a math-based job, as

well as offer a higher starting salary for men than women (Moss-Racusin et al., 2012; Deloitte, 2016). This is in line with role congruity theory (Eagly & Karau, 2002), which proposes that a perceived incongruity between gender roles (e.g. female) and other roles (e.g., scientist) leads to prejudice and discrimination.

Disentangling Gender Stereotypes and Sexism

Gender stereotyping broadly refers to ascribing specific traits, aptitudes, or roles to an individual based solely on gender group membership, whereas sexism typically refers to prescribing behaviours and social roles based on traditional gender roles and status (Brown & Gaertner, 2001; Rudman & Glick, 2008; Mastari et al. 2019). As such, gender stereotypes often go hand in hand with sexism. Further, different types of sexism have been identified. The most overt form, hostile sexism, is a form of sexism that is expressed in a more aggressive, resentful attitude towards women who violate traditional gender roles (Glick & Fiske, 1994). More current, subtle gender attitudes are measured by the Modern Sexism Scale (Swim et al., 1995). The Modern Sexism Scale assesses attitudes toward policies and practices that aim to reduce societal gender inequalities. The Modern Sexism scale has acceptable psychometric properties (Campbell et al., 1997). Finally, neosexism, measured by the Neosexism Scale (Tougas et al., 1999) measures beliefs that gender equality has already been achieved.

Present Study

The primary aim of this study was to investigate the effectiveness of a novel gender-STEM stereotype intervention, tested and validated in Chapter 2. The same intervention design and conditions were used. Once again, the intervention's main effects were examined as levels of gender-STEM stereotype discernment. Given the short, one-off nature of the intervention, it was not expected that significant shifts in core gender-stereotypic beliefs would shift. Rather, the intention was to improve one's ability to discern gender stereotypes. The ability to identify gender stereotypes and their harm is essential for their eradication (Cook & Cusack, 2011). Indeed, Cook and Cusack (2011) suggest gender stereotypes are like a disease. In this way, one cannot treat a disease if it is not diagnosed. Therefore, first, people must be able to discern gender stereotypes and their harm and then they can be 'treated' and reduced.

Psychological Mechanisms

The role of empathy in stereotype reduction is largely assumed and, when it is empirically examined, it is often treated as an outcome variable (e.g., Hill & Augoustinos, 2001), indicating a logical fallacy in discussing it as an explanatory variable in stereotype reduction. Participants may experience an increase in empathy and a decrease in stereotyped beliefs, but this does not necessarily mean empathy predicates the stereotype reduction. The relationship between the two may be associative rather than predictive. Indeed, other intervention research which manipulates empathy does not find that inducing greater empathy leads to greater stereotype reduction (Finlay & Stephan, 2000).

Thus, the intervention *approach* is often conflated with the related psychological *mechanism.* Cognitive programmes are assumed to be effective because they elicit cognitive mechanisms. While this may be true in part, it is most likely that both cognitive and affective mechanisms are activated and contribute to the stereotype reduction process. Indeed, inoculation theory, an overtly cognitive approach, acknowledges the crucial role of affective processes in intervention success (Compton & Ivanov, 2012; Banas & Richards, 2017; Compton & Pfau, 2005; Compton, 2021; Compton, 2022).

As such, the common cognitive and affective mechanisms assumed in typical stereotype reduction research were examined. Specifically, the influence of state empathic concern, perspective taking, and logical thinking were examined for their relative effects on intervention effects of discernment of gender-STEM stereotypes. Likewise, cognitive and affective mechanisms theorised and evidenced in inoculation research were examined. Specifically, counterarguing and psychological reactance to misinformation were tested.

Generalised Effects

Some inoculation interventions find a spill-over effect, such that the intervention has positive effects on attitudes and behaviours peripheral to the target belief or attitude in the intervention (Parker et al., 2012; Wong & Harrison, 2014). This is referred to as ‘umbrella’ or ‘blanket’ protection in inoculation research (Ivanov et al., 2012; Parker et al., 2012; Parker et al., 2016). Conversely, the continued influence effect is demonstrated when people’s intervention belief update does not translate to attitudes or behaviour change, because they continue to be influenced by the previous, incorrect belief (Chan et al., 2017; Cook et al., 2017; Lewandowsky et al, 2012; Ecker et al., 2010; Schwarz et al., 2016; Pluviano et al., 2017). To test the generalised effect of the intervention, male-agentic stereotype endorsement was examined, like in Chapter 2; additionally, this study examined the generalised effect of

the intervention on *attitudinal* and *behavioural* domains of gender stereotypes. Specifically, we examined the intervention effects on sexism and hiring evaluations of a male/female STEM applicant. We did not expect to find a generalised effect, given the saliency of gender beliefs. Beliefs that are intrinsically tied to one's self-concept, like gender, are more reluctant to change (Johnson, 1989; Eagly & Chaiken, 1993; Sherman, 2002).

Gender Differences

Finally, I examined gender differences in the intervention effects and across the outcome measures. Chapter 2 findings and previous literature do not suggest a moderating effect of gender on gender bias interventions (Pietri et al. 2017; Moss-Racusin et al., 2018; Zawadski et al., 2012). Alternatively, Chapter 2 and past research finds that men express greater sexist attitudes than women (Swim et al., 1995; Swim & Cohen, 1997; Glick & Fiske, 1996; Glick et al., 2000; Moya et al., 2007; Sibley et al., 2007; García-Sánchez et al., 2019; Baber & Tucker, 2006; Moss-Racusin et al., 2015; Handley et al., 2015; United States Bureau of Labor Statistics, 2019Parry, 2014; Zucker & Bay-Cheng, 2010). Additionally, past research has shown that both men and women demonstrate preference for male applicants over female applicants in STEM fields (Rueben et al., 2014; Glick et al., 1988; Heilman, 1984; Isaac et al., 2009; Moss-Racusin et al., 2012; Steinpreis et al., 1999). Thus, we did not expect to find a moderating effect of participant gender, or a main effect of participant gender on applicant evaluations, but we did expect a main effect of participant gender on gender stereotyped beliefs and attitudes.

Method

Participants

In line with previous research (e.g., Moss-Racusin et al., 2018; Pietri et al., 2017; Hennes et al., 2018), an effect size of $d = 0.31$ was used to conduct an a priori power analysis with 90% power and an error probability of 0.05 (G*Power; Faul et al., 2007; Faul et al., 2009). This revealed that a total of 596 participants would be required to detect a main effect of condition on the outcome measures.

Participants (N=616) were recruited through Prolific Academic (www.prolific.co), an online participant recruitment platform. As previously mentioned, it is considered superior to alternative online platforms, including CrowdFlower and MTURK, as its participants are more diverse, naive, and engaged (Peer et al., 2017). Participants received a remuneration of

£1.25 for participating. Eligibility was limited to individuals in North America over 18 years of age and fluent in English. Data cleaning procedures resulted in a total of 592 participants to analyse for main effects (four less than required to detect an effect size of $d = .31$).

In order to detect an effect of $\eta^2_{p=}.03$ (informed by Eaton et al., 2020 findings) with 90% power and 0.05 probability in the 4 (condition) x 2 (applicant gender) between-subjects analysis, an a priori power analysis revealed that a total sample size of 463 participants was necessary. This was achieved ($N = 552$) after excluding participants who failed the manipulation check for applicant gender and the previous data cleaning procedures.

Approximately half of the participants identified as female and around 30% identified as non-White. More than half of the participants had an undergraduate university degree or higher. Roughly one third of participants (31.2%; $n = 172$) indicated that they study or work in STEM, and 35.3% of participants ($n = 195$) indicated that they have experience in making hiring decisions. Further demographic information is provided in Table 3.1. The sample demographics question whether the sample is sufficiently representative of the general population, which is addressed further in the discussion section and Chapter 6 – General Discussion.

Table 3.1

Sample Demographics (N = 552)

	n	%
Ethnicity		
White	389	70.6
Black	42	7.6
Asian	46	8.3
Chinese	17	3.1
Latino/Hispanic	21	3.8
Mixed	26	4.7
Other	10	1.8
Gender		
Female	250	45.3
Male	286	51.8
Transgender male	2	0.4
Transgender female	2	0.4
Non-conforming/queer	9	1.6
Prefer not to say	3	0.5
Age[a]		
18-25	174	31.4
26-30	98	17.8
31-35	97	17.6
36-40	66	12
41-45	35	6.3
46-50	20	3.7
51-55	23	4.2
56-60	16	3
61-65	11	2
66-70	8	1.5
71-75	4	0.8
Education		
No university degree	259	46.9
Undergraduate university degree	204	37
Graduate university degree	89	16.1

Note. [a] Age was collected as a continuous variable, but is presented categorically here for ease of interpretation.

Procedure

Participants were recruited through Prolific Academic and directed to the online survey hosted on Qualtrics Survey Platform. Participants provided informed consent prior to participation and were then randomly assigned to one of four conditions: debunking (treatment video), inoculation (treatment video + misinformation video), misinformation video, or control video. After watching their assigned video (participants in the inoculation condition had watched only the treatment video at this stage), participants answered a series of questions about the video they had just watched. Those in the inoculation group then watched the misinformation video and again answered a series of questions about the video they watched.

Next, all participants viewed a job description for a vacant Research Associate position in the Department of Engineering at the University of Cambridge. This was followed by a completed application for the advertised position. Within each condition, half of the participants viewed a female applicant, and the other half viewed a male applicant. After reading the application, participants completed a series of questions about the candidate.

Finally, all participants completed a series of questionnaires on gender bias literacy and stereotype endorsement as well as demographic information. At the end of the survey, participants were invited to give feedback on the survey experience. To conclude, participants were debriefed (Appendix B; Debrief Form) and financially compensated for their participation. Ethics approval was obtained from the Psychology Department at the University of Cambridge.

It was hypothesised that:

H1: The treatment video would elicit logical thinking, empathic concern, and perspective taking relative to the misinformation and control videos

H2: Participants in the inoculation condition would experience greater counterarguing and psychological reactance to the misinformation video than participants in the misinformation condition

H3: Participants in the treatment conditions would demonstrate greater gender-STEM stereotype discernment than participants in the control and misinformation conditions

$H_0$4: There would be no effect of treatment on the more distal measures of gender stereotypes (i.e., no blanket of protection)

H5: Counterarguing and psychological reactance to the misinformation video would mediate the relationship between the treatment video and discernment scores among inoculation participants

H6: The intervention treatment would have an indirect effect on gender-STEM stereotype discernment through state logical thinking, state empathic concern, and/or perspective taking

H7: Participants in the treatment conditions would rate the female applicant more favourably than participants in the control and misinformation conditions

H8: The female applicant would be rated as more agentic than the male applicant across conditions

$H_0$9: There would not be a moderating effect of gender on intervention outcomes

H10: Women would demonstrate greater gender stereotype literacy than men across conditions

Materials and Measures

All items were randomised within measures and optional. To minimise priming effects and demand characteristics, the presentation of measures was not randomised. All materials and measures were administered in Qualtrics Survey Platform and are included in Appendix B.

Intervention Videos

The misinformation video was edited, based on results from Chapter 2, to improve effectiveness. The control video and treatment video tested in Chapter 2 were used without changes. The treatment and control videos were considered acceptable to use without edits given the ratings on comprehension, credibility, engagingness, and perceived increase in knowledge of gender-STEM stereotypes, as well as the successful inoculation check (the treatment video successfully increased counterarguing to misinformation). While the inoculative performance was not strong in Chapter 2, and it would have been preferred to test further iterations of all videos, this was not possible due to resource constraints. As discussed in Chapter 2, the inoculation intervention's weak effects may have been related to the random group differences between the inoculation condition and debunking condition, the sample's pre-existing resistance to gender-STEM stereotypes, and/or the statistically conservative approach of multiple post-hoc comparisons and corrections across all conditions (versus only those of interest).

Misinformation Video. The 1-minute misinformation video was designed with the aim to promote inaccurate information about innate gender differences in STEM ability. It is narrated by a female to improve source credibility and representation of the targeted gender group. This reduces any potential confounding effects of a male narrator, such as perceived male privilege or outgroup bias. Further, women are on average less gender biased than men, so may be seen as less biased and more credible when presenting the information.

The misinformation video was doubled in length from the pilot study version, from 30 seconds to 1 minute, and edited to include more "scientific" information supporting its

arguments. In particular, the video included greater discussion of evidence from the scientific paper it cites about men's and women's brains being wired differently, as well as data on the gender distribution across STEM degrees (presented visually as a graph and pie chart) to support that the gender gap in STEM is explained by men being naturally more interested in studying STEM. Increasing the length of the video by elaborating with supporting evidence should provide more credibility to the claims being made in the video, as well as provide more time for the participants to process the information and build resistance.

Figure 3.1

Four Screengrabs from the Misinformation Video

Photo of screengrab removed for copyright reasons.	Photo of screengrab removed for copyright reasons.
Photo of screengrab removed for copyright reasons.	Photo of screengrab removed for copyright reasons.

Applicant Assessment

Job Description. The job description, including salary range, was taken from a real job posting on the University of Cambridge careers website. Some of the detail regarding the position was removed for accessibility of language and to reduce fatigue effects.

Job Application. A completed job application form was used instead of a curriculum vitae, in order to avoid any confounding effects of name. Names have been used in previous research to signal applicant gender. Instead, with a job application form, the "name" was redacted, while still providing information about gender and ethnicity (which is information

commonly collected by employers for the purposes of meeting equality and diversity commitments). Protected characteristics, such as gender and ethnicity, can be used by interviews if they can prove they are using the information to improve underrepresentation, also known as 'positive action' (Equality and Human Rights Commission, 2019). The personal details provided were balanced to provide male-typical, female-typical, and gender neutral traits, activities, and experiences.

Measures

The following measures were repeated from Chapter 2: Active Inoculation, Attention Check, Video Credibility (α = 0.96), Counterarguing, Perceived Increase in Knowledge of Gender Stereotypes in STEM (α = .90), State Logical Thinking (α = .74), State Empathy (α = .89), State Perspective Taking, Political Orientation, Awareness of Gender Bias in STEM (α = 0.88), and Discernment of Gender Stereotypes in STEM (α = 0.83). Internal consistencies reported are from the current sample.

Discernment of Gender-STEM Stereotypes (DOGSS). To measure discernment of gender stereotypes in STEM, the six 7-point Likert-items tested in the pilot study were used.

CFA findings from Chapter 2 indicated very good model fit. The factor structure was re-tested with the current sample, and again revealed excellent model fit, χ^2 (8) = 42.98, p < .001 (CFI = .99, TLI = .98, RMSEA = .10, SRMR = 0.04). However, the factor correlation was extremely high (.93, p < .001), so a second CFA was conducted with one factor to see if it produced a better model fit. The bootstrapped one-factor model demonstrated very similar results, χ^2 (9) = 45.49, p < .001 (CFI = .99, TLI = .98, RMSEA = .10, SRMR = 0.05), and so the items were retained as one index, not two.

The index demonstrated good internal consistency as a single composite score (α = 0.83), and the internal reliability did not improve with the dropping of any items, indicating they should all be maintained. However, the items demonstrated poor inter-item bivariate correlations *(rs* ranging from |.38| to |.58|). Therefore, the items were first analysed independently for main effects and then together as one index in subsequent analyses. When analyses are performed at item-level, corrections for multiple comparisons are applied.

Manipulation Check. Participants completed a multiple choice question, "What was the gender of the applicant you reviewed in this study?" with the following response options, (a) Female, (b) Male, (c) Gender minority, and (d) I don't remember.

Psychological Reactance. Psychological reactance was assessed using an adapted version of the Salzburger State Reactance (SSR) Scale (Brehm, 1966). Participants responded to 4-items on a 7-point scale, ranging from 1 (*Not at all*) to 6 (*Very much*), including questions such as, “Did you perceive the message in the video as a restriction of freedom?” and “To what extent are you offended/disturbed by the message in the video?”. The scale demonstrated excellent internal consistency, $\alpha = 0.90$.

Male-Agentic Stereotype. Due to the poor performance of the Gender Stereotype Index (GSI) in the pilot study, participants completed the revised agentic subscale of the Personality Attributes Questionnaire (revised PAQ; Curran & Warber, 2011; original PAQ; Spence et al., 1973) on a scale from –3 (more true of women) to +3 (more true of men), with higher scores indicating more gender-stereotypic beliefs. Importantly, the PAQ measures only desirable instrumental traits (Frable, 1989; Spence, 1991). The valence of traits is important to control for in our study as research shows desirable agentic traits (e.g., independent) are found to relate to perceptions of STEM suitability (Ramsey, 2017), not undesirable agentic traits (e.g., hostile).

The masculine (agentic) subscale consists of eight agentic traits: independent, active, competitive, can make decisions easily, never gives up easily, self-confident, feels superior, stands up well under pressure. The scale demonstrated much better internal consistency than the GSI, however, it is still not considered acceptable, $\alpha = 0.69$. Taking a closer look at the inter-item and item-total correlations, the scale’s internal reliability improved to $\alpha = .73$ when dropping “feels superior”. This theoretically makes sense as “feels superior” may not be perceived as a desirable agentic trait, even though it is categorised as one by the PAQ scale authors. These findings are replicated in Chapter 4 as well. Therefore, for both empirical and theoretical reasons, the item was removed from the composite score and analyses. Analyses were re-run with the item included, and the results did not meaningfully change. Both the composite score and individual items are reported.

Modern Sexism. Modern sexism was measured using Swim et al.’s (1995) 8-item Likert scale, ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*). Participants responded to statements such as, “Discrimination against women is no longer a problem in our society" and “It is easy to understand why women's groups are still concerned about societal

limitations of women's opportunities". The scale demonstrated excellent consistency α = 0.92.

Hostile Sexism. Hostile sexism was measured using the Glick & Fiske (1997) 4-item Likert hostile sexism subscale of the Ambivalent Sexism Scale, ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*). Participants responded to statements such as "Women exaggerate problems they have at work" and "Many women are actually seeking special favours, under the guise of asking for "equality". The scale demonstrated excellent consistency, α = 0.93.

Affirmative Action. Support for affirmative action was measured using the affirmative action subscale of Tougas et al. (1999) neosexism scale. The affirmative action subscale consists of a 3-item Likert scale, ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*). Participants responded to statements such as "If there are no affirmative action programs helping women in STEM, they will continue to be unfairly treated" and "After years of discrimination, it is only fair to set up special programs to make sure women are given a fair and equitable treatment". The scale demonstrated excellent consistency, α = 0.94.

Evaluation of Job Applicant. Items evaluating the applicant's competence, likeability, hireability, and starting salary were adapted from Moss-Racusin et al.'s (2012) research. The suggested starting salary was measured using a sliding scale from £31,604 to £38,83; the range was informed by the University of Cambridge's job postings at the time of the study. The competence (α = 0.89) and hireability (α = 0.83) subscales were each comprised of three Likert items, and likeability was a single Likert-item. All Likert-items were measured on a scale from 1 (*Strongly disagree*) to 7 (*Strongly agree*).

Perceived Agency of Job Applicant. Participants completed the agentic subscale of the Personality Attributes Questionnaire (PAQ; Spence et al., 1973) on a 5-point bipolar semantic scale: not at all independent – very independent, not at all active – very active, not at all competitive – very competitive, cannot make decisions easily – can make decisions easily, gives up easily – never gives up easily, not at all self-confident – very self-confident, feels inferior – feels superior, does not stand up well under pressure – stands up well under pressure. The scale demonstrated very good internal reliability, α = 0.82.

Importantly, the PAQ measures only desirable instrumental traits (Frable, 1989; Spence, 1991). The valence of traits is important to control for in our study as research shows desirable agentic traits (e.g., independent) are found to relate to perceptions of STEM suitability (Ramsey, 2017), not undesirable agentic traits (e.g., hostile).

Trait Cognitive Rigidity. Cognitive rigidity was assessed using Oreg's (2003) 4-item Likert scale, ranging from 1 (*Strongly disagree*) to 6 (*Strongly agree)*. Participants responded to statements such as, "I often change my mind" and "Once I've come to a conclusion, I'm not likely to change my mind". The scale demonstrated good internal consistency, $\alpha = 0.77$.

Trait Empathic Concern. Empathy was measured using the 7-item empathic concern subscale from the interpersonal reactivity index (IRI) (Davis, 1980; Davis, 1983). The 5-point scale, ranged from 1 (*Does not describe me well*) to 5 (*Describes me very well*) and example statements include "I often have tender, concerned feelings for people less fortunate than me" and "When I see someone being taken advantage of, I feel kind of protective towards them". The scale demonstrated very good consistency $\alpha = 0.83$.

Trait Perspective Taking. Perspective taking was measured using the 7-item perspective taking subscale from the interpersonal reactivity index (IRI) (Davis, 1980; Davis, 1983). The 5-point scale, ranged from 1 (*Does not describe me well*) to 5 (*Describes me very well*) and example statements include "I sometimes find it difficult to see things from the "other person's" point of view" and "I sometimes try to understand my friends better by imagining how things look from their perspective". The scale demonstrated very good consistency $\alpha = 0.81$.

Analyses

As with the previous chapter, differences in outcome measures were analysed between intervention conditions. Following Field's (2017) advice, the robust Welch's t-tests were interpreted when data were normally distributed. When data were skewed, non-parametric tests Mann-Whitney U and Wilcoxon Signed-Rank tests were reported for between-group and within-group comparisons, respectively. As ANOVAs are fairly robust against violations of normality, non-parametric tests are not reported when the data are skewed. However, as a robustness check, non-parametric Brown-Forsythe and Kruskal-Wallis tests were and no meaningful differences were found between the three tests. Thus,

Welch's ANOVAs are reported throughout. For MANOVA, the Wilks' lambda test statistic is reported as it is most appropriate for heterogeneous variance. Pillai's trace was also run and no differences were found between Wilks' lambda and Pillai's trace.

When tests of main effects are broken out at item-level, Bonferonni corrections for multiple comparisons are applied. For post-hoc procedures, the Tukey method is reported as it is considered more powerful than the Bonferroni method when there are a larger number of comparisons made (Field, 2017). However, as the Bonferroni method is considered slightly more conservative, it was examined as a robustness check. No differences were found between Tukey and Bonferroni for any post-hoc analysis. To further increase confidence in findings, the Games-Howell procedure was run, due to the uncertainty of knowing whether the population variances were equal (Field, 2017). Again, the Games-Howell procedures provided the same results. Thus, Tukey is reported throughout.

Lastly, bootstrapping, a common robust estimation method (Efron & Tibshirani, 1993), was applied to all analyses. Bootstrap parameter estimates help to ensure the model parameters are robust and provide 95% confidence intervals of the parameter. All analyses were conducted in SPSS (versions 25, 26, and 27) and JASP (versions 0.12 and 0.16).

Results

To ensure there were no demographic differences between the experimental groups, Welch's *t*-tests and a chi-square test were performed. These tests confirmed that the four experimental conditions did not differ on levels of educational attainment ($F(3, 548) = 0.412$, $p=.745$), age ($F(3, 548) = 0.512$, $p=.674$), political orientation ($F(3,548) = 0.183$, $p = .908$), or gender ($\chi^2(3)=4.411$, $p=.220$). Therefore, in all further analyses, parsimonious models without these variables were estimated.

To ensure there were no important random differences between the debunking and inoculation groups, independent Welch's *t*-tests were run on relevant variables. There were no between-group differences between the debunking and inoculation participants on state perspective taking, state logical thinking, state empathy, perceived increase in knowledge, or perceived credibility.

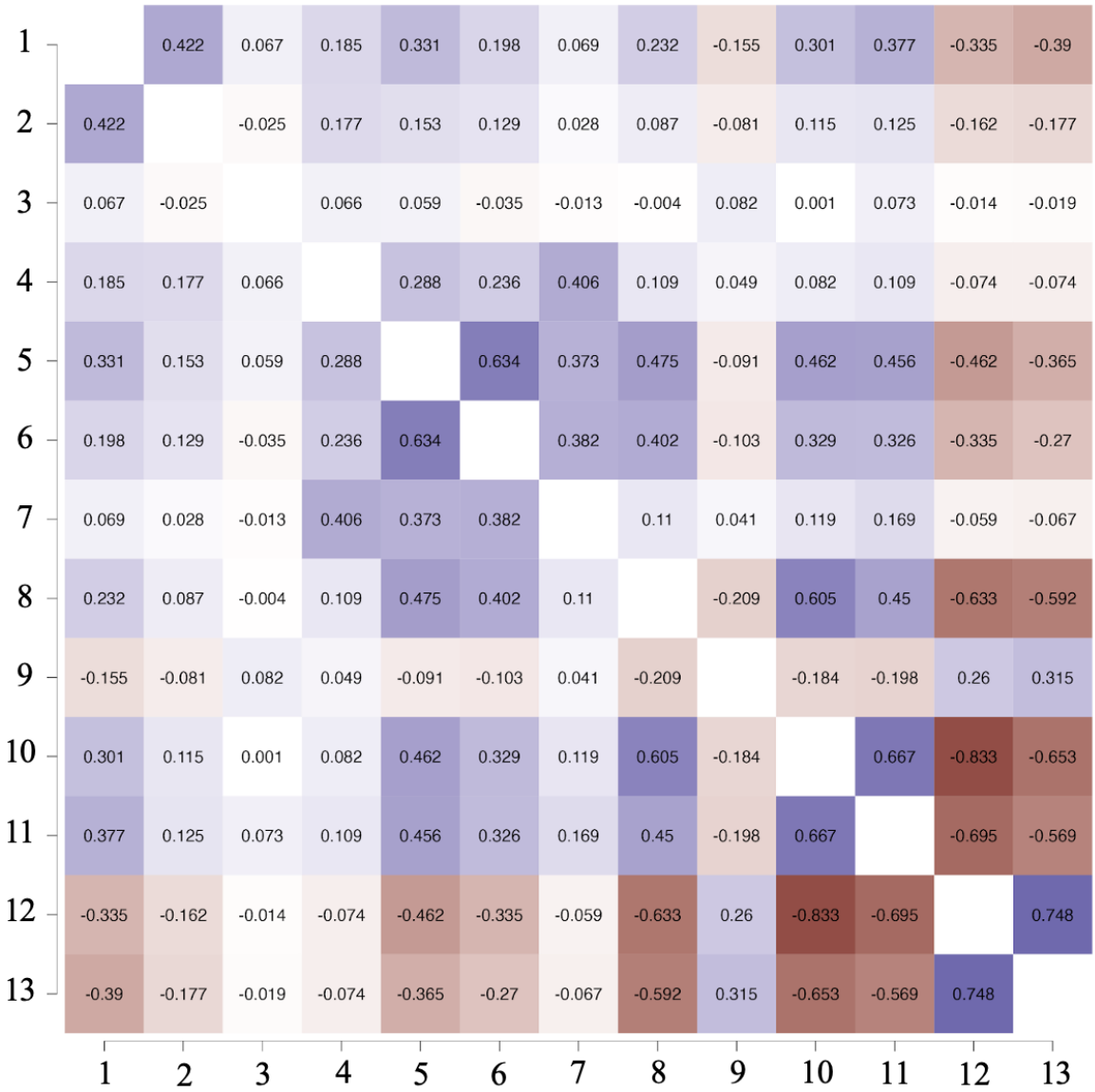
Correlations

Initial bivariate relationships between trait, state, and outcome variables were run and are presented in Table 3.2. As expected, all gender stereotype literacy outcome measures demonstrated significant relationships with one another *(rs* ranging from |.18| to |.75|; *p*s < .001). The *r* range increased to |.45| to |.75| when male-agentic stereotype was not included, suggesting a surprisingly weak association between male-agentic stereotype endorsement and other measures of gender stereotype endorsement or resistance.

Trait empathic concern was significantly associated with trait perspective taking *(r* = .422, *p* <.001), but not so much so that they appear to be measuring the same thing. As such, they were treated independently in analyses. Likewise, state empathic concern and state perspective taking were positively related, and to a higher extent than their trait counterparts *(r* = .634, *p* <.001). However, the strength of the relationship still suggests they are appropriately examined separately.

Table 3.2

Correlation Matrix of Trait, State, and Outcome Variables



Note. 1 = trait empathic concern; 2 = trait perspective taking; 3 = trait cognitive rigidity; 4 = state logical thinking; 5 = state empathic concern; 6 = state perspective taking; 7 = perceived increased in knowledge of gender stereotypes in STEM; 8 = discernment of gender-STEM stereotypes; 9 = male-agentic stereotype; 10 = awareness of gender bias in STEM; 11 = support for affirmative action; 12 = modern sexism; 13 = hostile sexism

Intervention Videos

Before testing main effects of the intervention, the intervention videos were examined to ensure (1) sufficient levels of perceived credibility of the treatment video and misinformation video and perceived increase in knowledge and (2) treatment-elicited perceived increase in knowledge. Then, H1 was tested to examine levels of state empathic concern, perspective taking, and logical thinking across intervention videos. As the debunking (treatment-only) condition and inoculation (treatment + misinformation) condition

were exposed to the same treatment material and measures for the first phase of the study, these two conditions were collapsed for the following analyses.

Credibility

Credibility checks were conducted to test whether the updates to the misinformation video, guided by the findings from the pilot study in the previous chapter, improved its efficacy. The credibility of the misinformation video significantly improved from the pilot (M_{diff} = 1.060, 95% CI [0.395, 1.724], $t(39)$ = 3.226, p <.001).

The control video *(M* = 6.315, *SD*= 1.024), treatment video *(M* = 5.649, *SD*= 1.277), and misinformation video *(M* = 4.087, *SD*= 1.628) all received credibility ratings above 4 (the midpoint on the bipolar semantic scale), suggesting perceived credibility was acceptable for all videos. Although the credibility of the misinformation video significantly improved from the pilot version, it was still rated as significantly less credible than both the treatment video (M_{diff} = -1.560, 95% CI [-1.872, -1.258], $t(295.559)$ = -11.403, p_{tukey} < .001, d = -1.111) and control video (M_{diff} = -2.223, 95% CI [-2.572, -1.907], $t(295.559)$ = -14.085, p_{tukey} < .001, d = -1.635); $F(2, 295.559)$ = 93.907, p < .001, η2 = .282. Thus, the lower credibility rating for the misinformation video may indicate that participants entered the study with pre-existing discernment of gender-STEM stereotypes and evaluated misinformation as less credible. Additionally, it is possible that the differing video lengths between the misinformation video (1:00 minute) and the treatment and control videos (4:07 minutes and 3:39 minutes, respectively) may have influenced the misinformation video's lower credibility rating.

Perceived Increase in Knowledge of Gender Stereotypes

Replicating the pilot study, a one-way ANOVA revealed significant differences in perceived increase in knowledge of gender stereotypes in STEM by video, $F(2, 132.98)$ = 120.361, p < .001, η2 = .248. Post hoc procedures indicated significant differences between all video conditions: the treatment video *(M* = 5.110, *SD*= 1.635) produced the greatest perceived increase in knowledge of gender stereotypes, followed by the misinformation video (M_{diff} = 0.962, 95% CI [0.585, 1.286], $t(132.98)$ = 5.892, p_{tukey} < .001, d = 0.578) and lastly the control video (M_{diff} = 3.148, 95% CI [2.672, 3.498], $t(132.98)$ = 12.131, p_{tukey} < .001, d = 1.985). The misinformation condition *(M* = 4.155, *SD*= 1.685) perceived a significantly greater increase in knowledge of gender stereotypes than the control condition (M_{diff} = 2.184, 95% CI [1.728, 2.609], $t(132.98)$ = 7.978, p_{tukey} < .001, d = 1.378), with the mean value

(4.155) falling slightly above the 'slightly agree' point on the scale. These findings support the increased effectiveness of the updated misinformation video and the continued effectiveness of the treatment video.

Treatment-Elicited Logical Thinking

A one-way ANOVA revealed significant group differences in state logical thinking, $F(2, 113.22) = 9.62$, $p < .001$. The treatment video elicited the highest level of state logical thinking ($M = 5.44$, $SD = 1.08$) followed by the misinformation video ($M_{diff} = 0.481$, 95% CI [0.228, 0.717], $t(113.22) = 4.119$, $p_{tukey} < .001$, $d = 0.413$) and the control video ($M_{diff} = 0.502$, 95% CI [0.174, 0.989], $t(113.22) = 2.765$, $p_{tukey} = .016$, $d = 0.443$). No significant difference was observed between the misinformation video ($M = 4.96$, $SD = 1.28$) and the control video ($M = 4.92$, $SD = 1.28$).

Treatment-Elicited Empathic Concern

A one-way ANOVA revealed significant group differences in state empathic concern, $F(2, 117.58) = 30.90$, $p < .001$. Post hoc procedures indicated significant differences in state empathic concern between all video conditions. The treatment video produced the greatest increase in state empathic concern ($M = 5.18$, $SD = 1.63$), followed by the misinformation video (Mdiff = 0.728, 95% CI [0.408, 1.076], $t(117.58) = 4.349$, ptukey < .001, $d = 0.436$) and the control video (Mdiff = 1.975, 95% CI [1.501, 2.578], $t(117.58) = 7.451$, ptukey < .001, $d = 1.195$). The misinformation video ($M = 4.45$, $SD = 1.75$) produced a greater level of state empathic concern compared to the control video (Mdiff = 1.249, 95% CI [0.778, 1.857], $t(117.58) = 4.469$, ptukey < .001, $d = 0.758$).

Treatment-Elicited Perspective Taking

A one-way ANOVA revealed significant group differences in state perspective taking, $F(2, 121.51) = 64.55$, $p < .001$. Post hoc procedures indicated significant differences in state perspective taking between all video conditions. The treatment video had the greatest increase in state perspective taking ($M = 5.07$, $SD = 1.63$), followed by the misinformation video (Mdiff = 0.810, 95% CI [0.438, 1.119], $t(121.51) = 4.738$, ptukey < .001, $d = 0.475$) and the control video (Mdiff = 2.740, 95% CI [2.238, 3.166], $t(121.51) = 10.088$, ptukey < .001, $d = 1.617$). The misinformation video ($M = 4.26$, $SD = 1.85$) produced a greater level of state perspective taking compared to the control video (Mdiff = 1.917, 95% CI [1.392, 2.470], $t(121.51) = 6.73$, ptukey < .001, $d = 1.142$).

Inoculation Checks

In this section, I test H2. This is to validate the effectiveness of the treatment video as an inoculation intervention. Specifically, inoculation checks were conducted on perceived credibility, counterarguing, and psychological reactance in response to the misinformation video between the inoculation condition and misinformation condition.

Credibility

In support of successful inoculation, the credibility of the misinformation video was rated significantly lower among those who watched it after the treatment video compared to those who watched only the misinformation video (M_{diff} = -0.593, 95% CI [-0.946, -0.202], t(288.858) = -3.060, d = -0.358, p_{tukey} = .002). This is an improvement from the findings in the previous chapter, further indicating the updates to the misinformation video were successful.

Counterarguing

Those who watched the misinformation video after the treatment video engaged in significantly more counterarguing (M = 4.638, SD= 1.950) than those who watched only the misinformation video, (M_{diff} = 0.734, 95% CI [0.257, 1.147], t(288.177) = 3.213, p_{tukey} = .001, d = 0.376). This suggests that the treatment video was successful in eliciting counterarguing to misinformation, a key element of inoculation.

Psychological Reactance

Compared to the misinformation condition (M = 2.919, SD= 2.024), the inoculation condition expressed significantly greater psychological reactance to the misinformation video (M = 3.567, SD= 1.987; $Mdiff$ = 0.652, 95% CI [0.197, 1.109], t(290.052) = 2.765, d = 0.323, p_{tukey} = .006). This further supports successful inoculation as the treatment video elicited psychological resistance to the misinformation messages.

Intervention Main Effects

I will now assess H3 and H4. First, H3 was examined by testing group differences on the main outcome variable, discernment of gender stereotypes in STEM. Then, H4 was examined by testing group differences on more distal measures of gender stereotyping.

Discernment of Gender Stereotypes in STEM

To examine the effect of experimental condition on discernment of gender-STEM stereotypes, an initial omnibus one-way MANOVA was run on the 6 DOGSS items, F (18, 1103.572) = 10.896, Wilk's $\Lambda = 0.630$, partial $\eta^2 = .143$, $p<.001$. Given the significant result, follow-up one-way Welch's ANOVAs were run and demonstrated group differences on all six items.

Table 3.3

Descriptive Statistics and Between-Group Comparisons by Intervention Condition

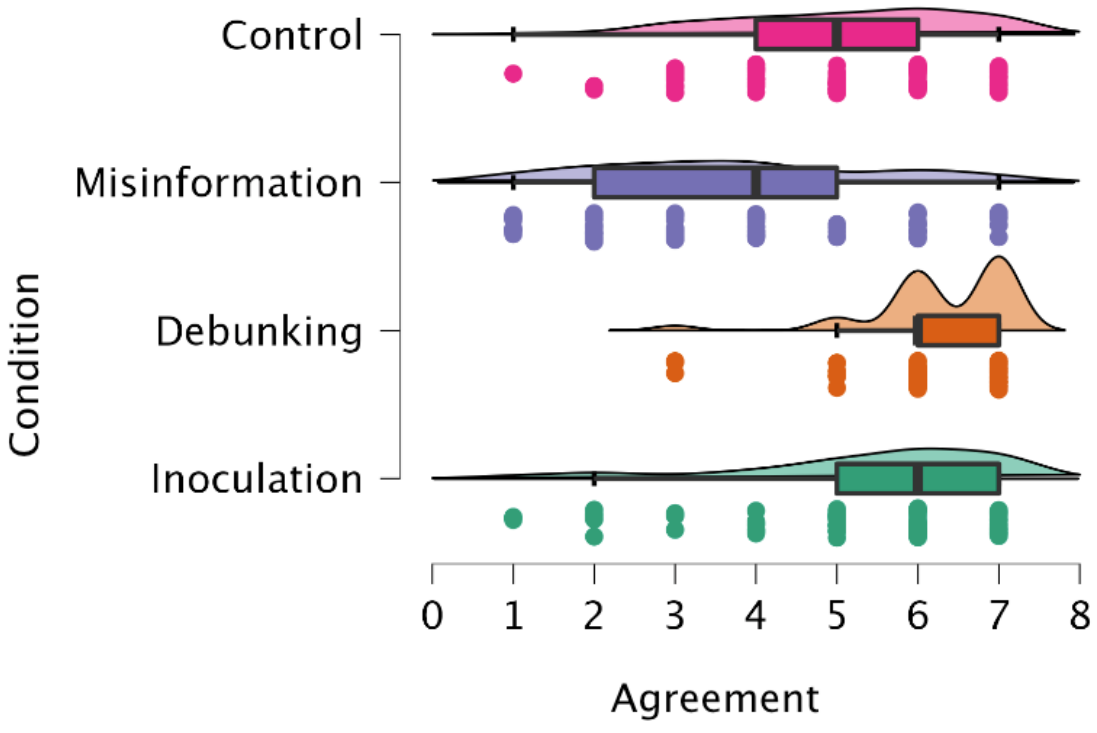
	Inoculation (n=141)		Debunking (n=146)		Misinformation (n=152)		Control (n=153)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
DOGSS								
Scientific consensus	5.37_{b}	1.60	6.31_{a}	0.87	3.77_{c}	1.73	5.09_{b}	1.46
Not naturally interested	3.83_{bc}	1.86	3.08_{a}	1.77	4.30_{c}	1.70	3.39_{ab}	1.82
Brains function differently	3.55_{b}	1.79	2.83_{a}	1.70	4.79_{d}	1.62	4.10_{c}	1.79
Men innately better	2.35_{ab}	1.57	1.93_{a}	1.41	3.60_{c}	1.79	2.80_{b}	1.59
Use misleading tactics	5.85_{a}	1.40	6.17_{a}	1.12	5.09_{b}	1.50	5.43_{b}	1.35
Hiring/promotion effects	5.66_{ab}	1.51	5.79_{a}	1.43	5.23_{b}	1.63	5.40_{ab}	1.54
Composite score	5.19_{ab}	1.16	5.69_{a}	1.05	4.25_{c}	1.20	4.92_{b}	1.26
Male-agentic stereotype	0.34_{a}	0.54	0.32_{a}	0.50	0.34_{a}	0.54	0.41_{a}	0.56
Hostile sexism	2.86_{a}	1.5	2.55_{a}	1.40	2.80_{a}	1.49	2.69_{a}	1.53
Modern sexism	2.82_{a}	1.36	2.74_{a}	1.31	2.83_{a}	1.35	2.74_{a}	1.41
Affirmative action	4.56_{a}	1.64	4.49_{a}	1.63	4.62_{a}	1.68	4.76_{a}	1.65
Gender bias awareness	5.15_{a}	1.42	5.34_{a}	1.46	5.09_{a}	1.48	5.23_{a}	1.38
Knowledge increase	4.97_{a}	1.69	5.23_{a}	1.56	4.15_{b}	1.68	2.25_{c}	1.56
State empathy	5.22_{a}	1.51	5.15_{a}	1.59	4.57_{b}	1.61	3.17_{c}	1.75
State perspective taking	5.08_{a}	1.61	5.04_{a}	1.64	4.26_{b}	1.84	2.66_{c}	1.80
State logical thinking	5.48_{a}	1.03	5.39_{a}	1.11	4.95_{b}	1.27	4.94_{b}	1.30
Psychological reactance [a]	3.57_{a}	1.98	--	--	2.75_{b}	1.83	--	--
Counterarguing [a]	4.64_{a}	1.95	--	--	3.91_{b}	1.90	--	--

Note. [a] Responses provided after viewing misinformation video. Inoculation condition = treatment video + misinformation video; Debunking condition = treatment video; Scientific consensus = "The scientific consensus is that there are no innate gender differences in STEM ability"; Naturally interested = "More females are not in STEM because they are naturally interested in other things"; Brains function differently = "On average, men's and women's brains function differently"; men innately better = "Men are innately better at STEM"; Use misleading tactics = "Some people use misleading tactics to falsely claim that men are better at STEM than women"; Hiring/promotion effects = "Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM"; Knowledge increase = perceived increase in knowledge of gender stereotypes in STEM; Gender bias awareness = Awareness of Gender Bias in STEM

"The scientific consensus is that there are no innate gender differences in STEM ability" revealed a significant main effect of condition ($F(3, 193.299) = 56.955$, $p < .001$, $\eta^2 = .282$; see Figure 3.1). Bootstrapped post hoc comparisons indicated that the debunking condition agreed significantly more with the statement than all three other conditions: inoculation ($M_{diff} = 0.938$, 95% CI [0.582, 1.329], $t(193.299) = 4.333$, $p_{tukey} < .001$, $d = 0.719$), control condition ($M_{diff} = 1.174$, 95% CI [0.824, 1.496], $t(193.299) = 5.468$, $p_{tukey} < .001$, $d = 0.973$), and misinformation condition ($M_{diff} = 2.567$, 95% CI [2.133, 2.922], $t(193.299) = 11.871$, $p_{tukey} < .001$, $d = 1.844$), suggesting the treatment video demonstrated a significant persuasive effect (greater agreement with the "correct" belief). Additionally, the inoculation condition had significantly higher agreement than the misinformation condition ($M_{diff} = 1.633$, 95% CI [1.118, 2.067], $t(193.299) = 7.478$, $p_{tukey} < .001$, $d = 0.958$). Lastly, the control condition agreed significantly more with the statement than the misinformation condition ($M_{diff} = 1.391$, 95% CI [0.965, 1.840], $t(193.299) = 6.447$, $p_{tukey} < .001$, $d = 0.860$). Therefore, the only non-significant group comparison was between the inoculation condition and the control condition ($p_{tukey}=0.695$), suggesting the misinformation video 'neutralised' the treatment video's effect.

Figure 3.1

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "The scientific consensus is that there are no innate gender differences in STEM ability"

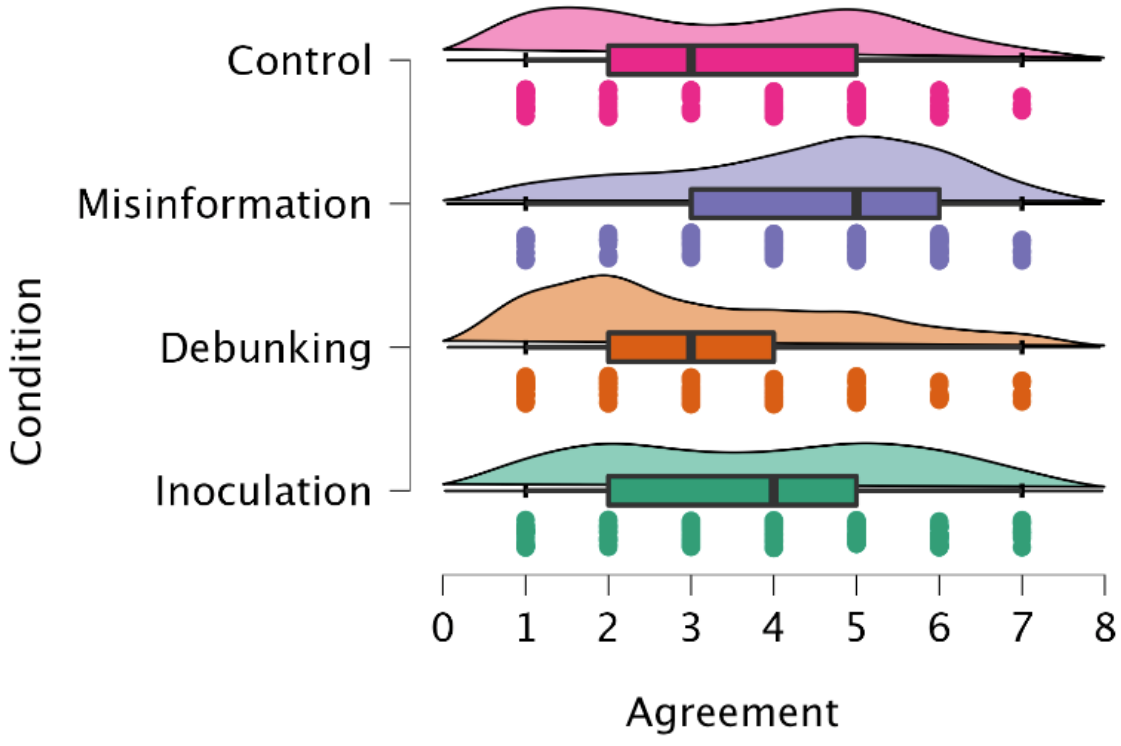


Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges.

"More females are not in STEM because they are naturally interested in other things" revealed a significant main effect of condition ($F(3, 303.741) = 12.368$, $p < .001$, $\eta^2 = .060$; see Figure 3.2). Bootstrapped post hoc comparisons indicated that the debunking condition agreed significantly less with the statement than the inoculation ($M_{diff} = -0.710$, 95% CI [-1.152,-0.292], $t(303.741) = -3.283$, $p_{tukey} = 0.006$, $d = -0.391$) and the misinformation condition ($M_{diff} = -1.204$, 95% CI [-1.580,-0.763], $t(303.741) = -5.564$, $p_{tukey} < .001$, $d = -0.692$), once again suggesting the treatment video demonstrated a significant persuasive effect (decrease in the "right" direction). Additionally, the misinformation condition agreed significantly more with the statement than the control condition ($M_{diff} = 0.895$, 95% CI [0.437,1.294], $t(303.741) = 4.115$, $p_{tukey} < .001$, $d = 0.500$). All other group comparisons were non-significant ($ps > .116$). These findings again indicate that the misinformation attack nullified the treatment video's effects.

Figure 3.2

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "More females are not in STEM because they are naturally interested in other things"

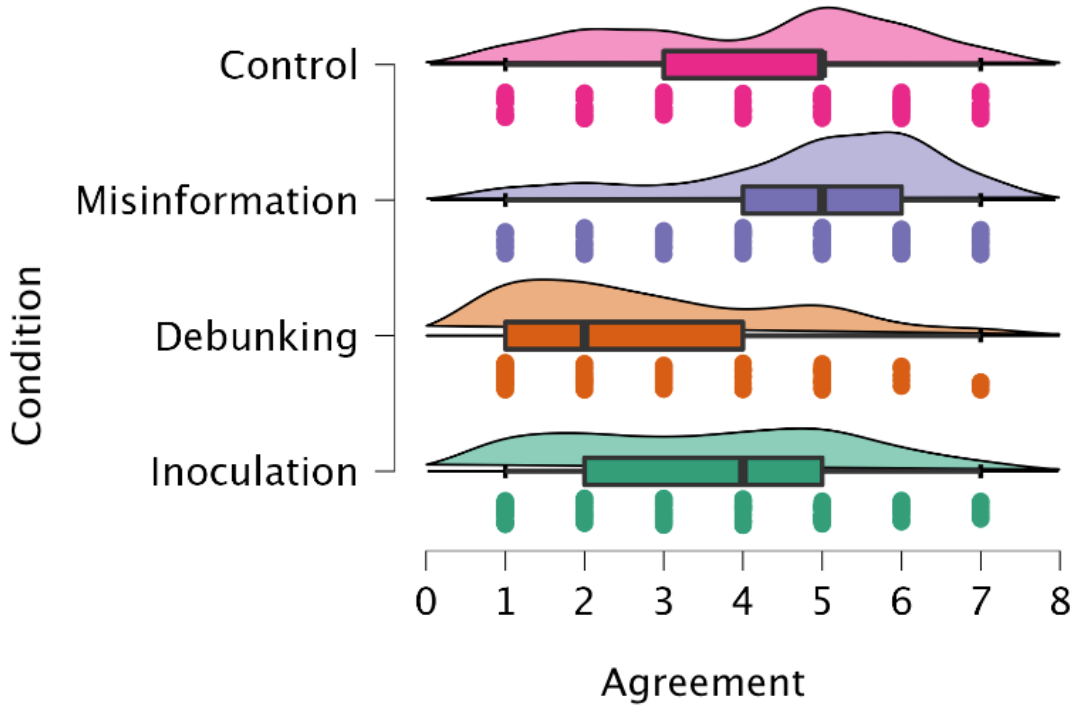


Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges.

"On average, men's and women's brains function differently" revealed a significant main effect of condition ($F(3, 303.847) = 32.956$, $p < .001$, $\eta^2 = .146$; see Figure 3.3). Bootstrapped post hoc comparisons indicated significant differences on all comparisons. The debunking condition agreed significantly less with the statement than all three other conditions: inoculation ($M_{diff} = -0.658$, 95% CI [-1.047,-0.238], $t(303.847) = -3.102$, $p_{tukey} = .011$, $d = -0.371$), control condition ($M_{diff} = -1.217$, 95% CI [-1.598,-0.802], $t(303.847) = -5.486$, $p_{tukey} < .001$, $d = -0.693$) and misinformation condition ($M_{diff} = -1.927$, 95% CI [-2.265,-1.506], $t(303.847) = -9.291$, $p_{tukey} < .001$, $d = -1.149$), suggesting the treatment video demonstrated a significant persuasive effect (increase in the "right" direction). Additionally, the inoculation condition agreed significantly less with the statement than the misinformation condition ($M_{diff} = -1.268$, 95% CI [-1.683,-0.885], $t(303.847) = -6.121$, $p_{tukey} < .001$, $d = -0.749$) and the control condition ($M_{diff} = -0.572$, 95% CI [-0.975,-0.154], $t(303.847) = -2.713$, $p_{tukey} = .035$, $d = -0.318$). Lastly, the misinformation condition agreed more with the statement than the control condition ($M_{diff} = 0.702$, 95% CI [0.321,1.079], $t(303.847) = 3.413$, $p_{tukey} = .004$, $d = 0.413$). Therefore, while the misinformation video attenuated some of the treatment effects in the inoculation condition, it did not fully 'neutralise' it.

Figure 3.3

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "On average, men's and women's brains function differently"



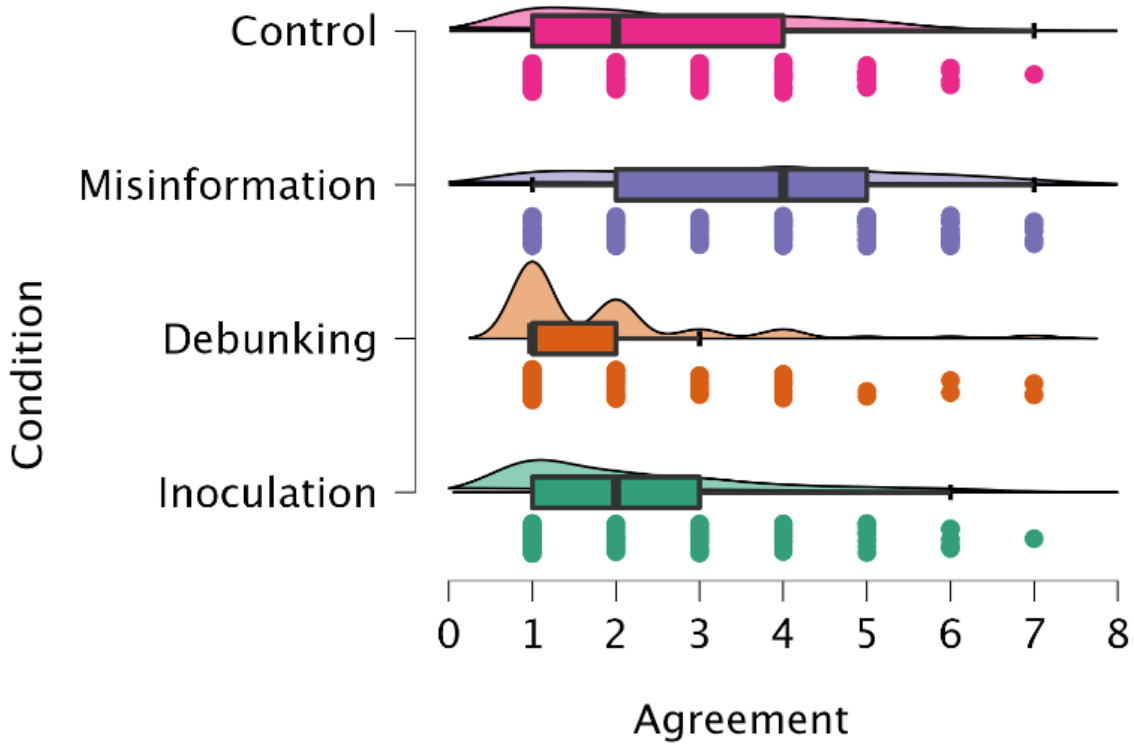
Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges.

"Men are innately better at STEM" revealed a significant main effect of condition ($F(3, 303.203) = 29.045$, $p < .001$, $\eta^2 = .145$; see Figure 3.4). Bootstrapped post hoc comparisons indicated that the debunking condition agreed significantly less with the statement than the control condition ($M_{diff} = -0.860$, 95% CI [-1.203,-0.491], $t(303.203) = -4.551$, $p_{tukey} < .001$, $d = -0.594$) and the misinformation condition ($M_{diff} = -1.732$, 95% CI [-2.096,-1.338], $t(303.203) = -9.238$, $p_{tukey} < .001$, $d = -1.082$). Further, the inoculation condition agreed significantly less with the statement than the misinformation condition ($M_{diff} = -1.281$, 95% CI [-1.728,-0.912], $t(303.203) = -6.877$, $p_{tukey} < .001$, $d = -0.773$). As expected, the misinformation condition agreed significantly more with the statement than the control condition ($M_{diff} = 0.862$, 95% CI [0.481,1.280], $t(303.203) = 4.662$, $p_{tukey} < .001$, $d = 0.517$). All other group comparisons were non-significant ($ps > .132$). In sum, the debunking and control conditions significantly differed from each other, suggesting a persuasive treatment effect (i.e., psychological change in the "right" direction). However, the inoculation condition did not differ from the debunking condition or the control condition, suggesting (a)

the misinformation attack attenuated most, but not all, of the persuasive treatment effect and (b) the psychological distance between the control condition and inoculation condition was narrow, indicating the psychological update towards the "correct" belief was small.

Figure 3.4

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "Men are innately better at STEM"



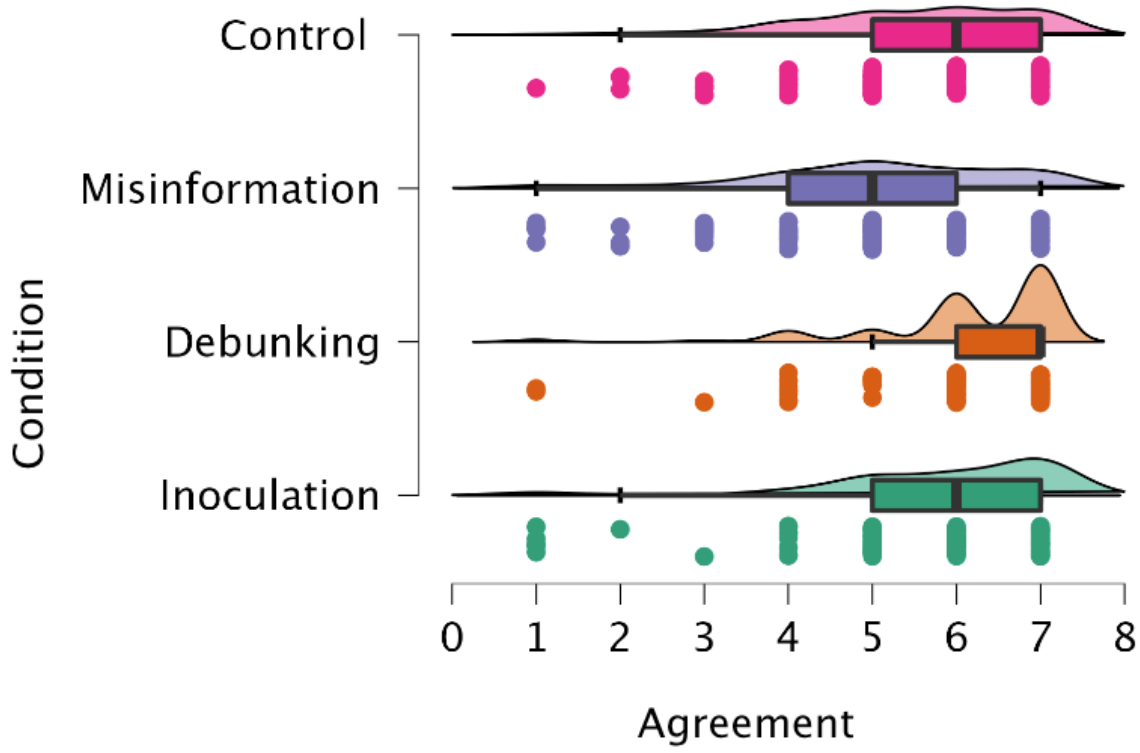
Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges.

"Some people use misleading tactics to falsely claim that men are better at STEM than women" revealed a significant main effect of condition ($F(3, 302.813) = 17.116$, $p < .001$, $\eta^2 = .082$; see Figure 3.5). Bootstrapped post hoc comparisons indicated that the debunking condition agreed significantly more with the statement than the control condition ($M_{diff} = 0.679$, 95% CI [0.400,0.947], $t(302.813) = 4.208$, $p_{tukey} < .001$, $d = 0.557$) and the misinformation condition ($M_{diff} = 1.077$, 95% CI [0.781,1.386], $t(302.813) = 6.676$, $p_{tukey} < .001$, $d = 0.809$). Additionally, the inoculation condition agreed significantly more with the statement than the misinformation condition ($M_{diff} = 0.743$, 95% CI [0.368,1.089], $t(302.813) = 4.560$, $p_{tukey} < .001$, $d = 0.507$). Lastly, in comparison to the control condition, the

misinformation condition demonstrated notable, but non-significant, disagreement with the statement (p_{tukey} = 0.070, d = -0.280). All other group comparisons were non-significant (*ps*>.149). Therefore, although the debunking and control conditions significantly differed from each other, suggesting a persuasive treatment effect (i.e., psychological change in the "right" direction), the inoculation condition did not differ from the debunking condition or the control condition. This again suggests two things: (a) the persuasive effect of the treatment was reduced substantially, but not entirely, by the misinformation attack, and (b) the resulting psychological distance in discernment between the control condition and inoculation condition was narrow.

Figure 3.5

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "Some people use misleading tactics to falsely claim that men are better at STEM than women"



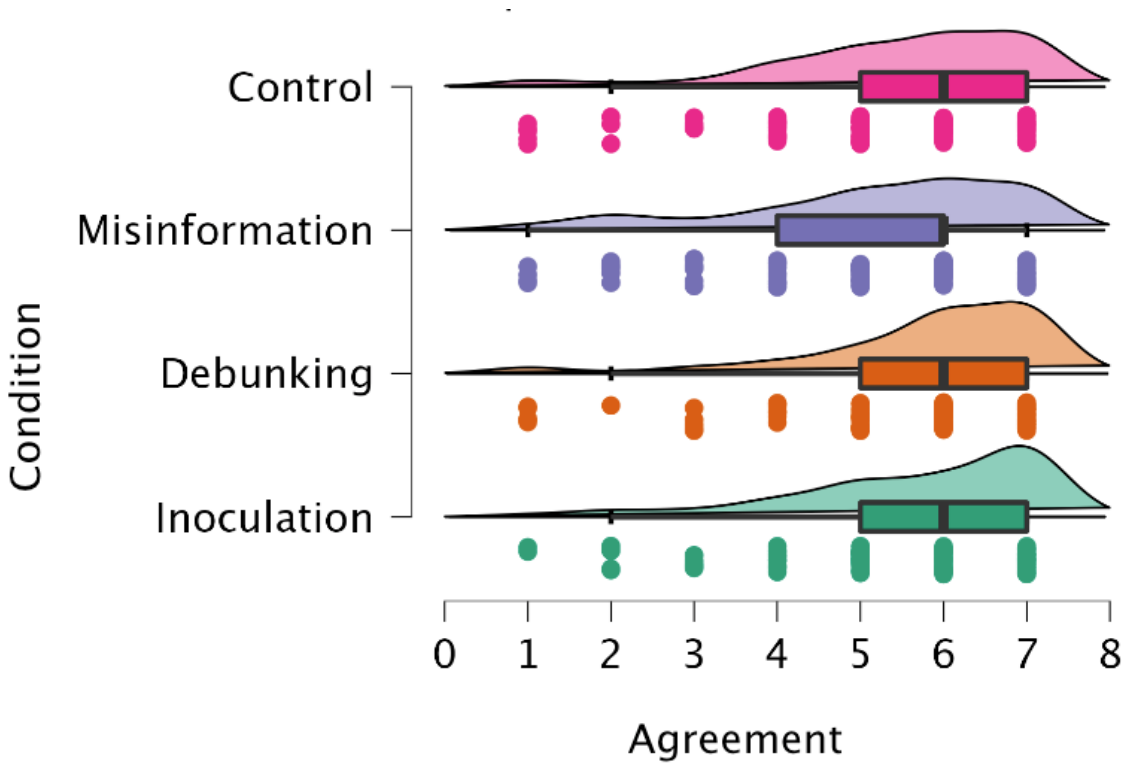
Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges.

"Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM" revealed a significant main effect of condition (F(3, 304.286) = 4.227, p = .006, η^2 = .024; see Figure 3.6. Bootstrapped post hoc comparisons indicated that

the debunking condition agreed significantly more with the statement than misinformation condition (M_{diff} = 0.624, 95% CI [0.247,0.984], t(304.286) = 3.379, p_{tukey} = .004, d = 0.396). The inoculation condition also agreed significantly more with the statement than the misinformation condition (M_{diff} = 0.523, 95% CI [0.183,0.899], t(304.286) = 2.814, p_{tukey} = .026, d = 0.328). All other group comparisons were non-significant (ps>.290). As such, while the inoculation condition and debunking condition did not differ (suggesting inoculative effects), they also did not differ from the control condition (suggesting no treatment effects). As suggested by the means and standard deviations of the control condition (Table 3.3), this may be due to the sample's high endorsement of the "correct" belief prior to the study (i.e., ceiling effects).

Figure 3.6

Raincloud Plot showing Between-Condition Comparison of Agreement with the Statement "Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM"

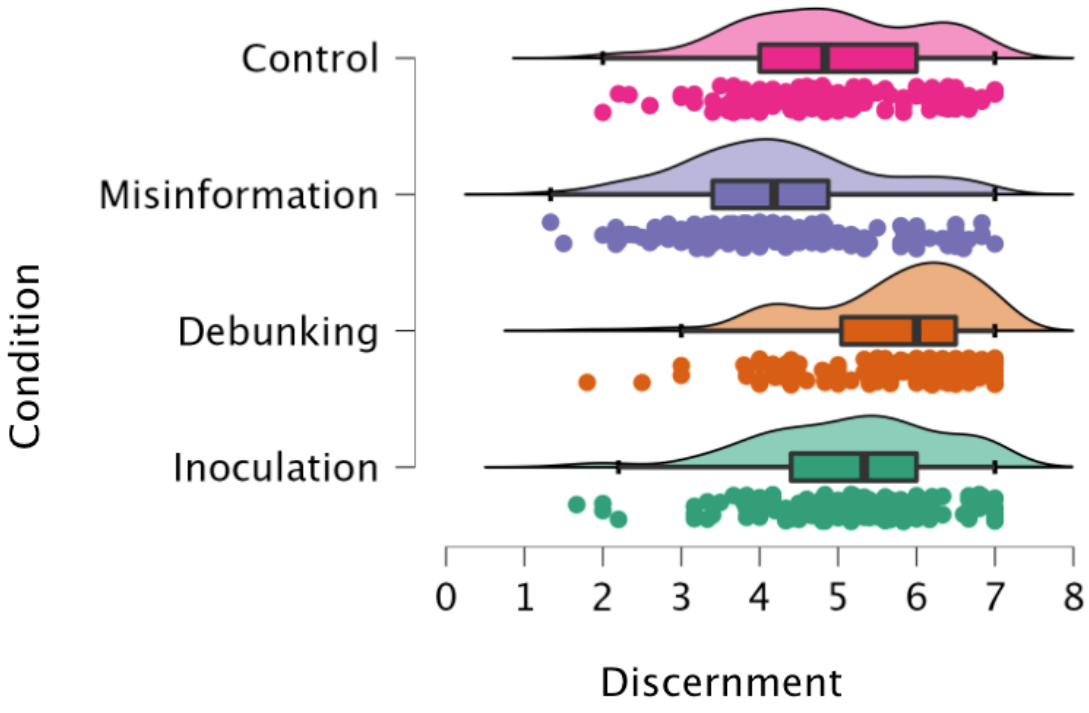


Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges.

Finally, on the total composite score of discernment, a Welch's ANOVA demonstrated significance, $F(3, 325.97) = 41.09$, $p < .001$, $\eta^2 = .172$ (see Figure 3.7). Bootstrapped post hoc comparisons indicated that the debunking condition demonstrated greater discernment than the inoculation condition ($M_{diff} = .50$, 95% CI [0.153, 0.851], $t = 3.704$, $p_{tukey} = .001$, $d = 0.437$), control condition ($M_{diff} = .76$, 95% CI [0.423, 1.107], $t = 5.76$, $p_{tukey} < .001$, $d = 0.667$), and misinformation condition ($M_{diff} = 1.43$, 95% CI [1.096, 1.781], $t = 5.762$, $p_{tukey} < .001$, $d = 1.253$. Additionally, the inoculation condition did not differ from the control condition ($p_{tukey} = .203$) but demonstrated greater discernment than the misinformation condition ($M_{diff} = .93$, 95% CI [0.500, 1.132], $t = 6.979$, $p_{tukey} < .001$, $d = 0.816$). The control condition also demonstrated greater discernment than the misinformation condition ($M_{diff} = .67$, 95% CI [0.280, 0.893], $t = 5.123$, $p_{tukey} < .001$, $d = 0.587$. Therefore, replicating Chapter 2 findings, the treatment provided seemed to provide a successful persuasive effect (significantly higher discernment in the debunking condition), but not an inoculative effect on overall discernment of gender-STEM stereotypes. It appears the misinformation video 'neutralised' the treatment video's effect.

Figure 3.7

Raincloud Plot showing Between-Condition Comparison of Discernment of Gender Stereotypes in STEM



Note. The raincloud plot visualises data density with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges.

As post-hoc comparisons and corrections were performed across all conditions (not only those of interest), and not all post-hoc comparisons were relevant or hypothesized, this impacted the reported statistical performance of the treatment conditions. This statistically conservative approach was taken due to not pre-registering specific contrasts of interest. When tests were run with only hypothesized post-hoc comparisons, the debunking condition significantly differed from the control condition on the composite score and all items, and the inoculation condition significantly differed from the control condition on the composite score and all but two items: "Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM" and "The scientific consensus is that there are no innate gender differences in STEM ability".

Generalised Effect

As expected, there were no main effects of intervention condition on more distal measures of gender stereotype endorsement and attitudes. A one-way Welch's ANOVA on male-agentic stereotype endorsement found no differences between conditions ($F($3,325.600) = 0.730, p = .535, η^2 = .004). Additionally, awareness of gender bias in STEM did not demonstrate any between-group differences ($F($3,325.865) = 0.833, p = .477, η^2 = .004). Similarly, there were no group differences on levels of modern sexism ($F($3,325.066) = 0.170, p = .916), hostile sexism ($F($3, 326.097) = 1.292, p = 0.277), or affirmative action ($F($3,324.333) = 0.715, p = .544). Thus, 'umbrella protection' or a 'blanket of protection' (Papageorgis & McGuire, 1961; Parker et al., 2012) was not established. However, across conditions, group means indicate low levels of sexist attitudes and agentic-trait stereotype endorsement, so perhaps these attitudes were, on average, already "correct" and protected.

Psychological Mechanisms

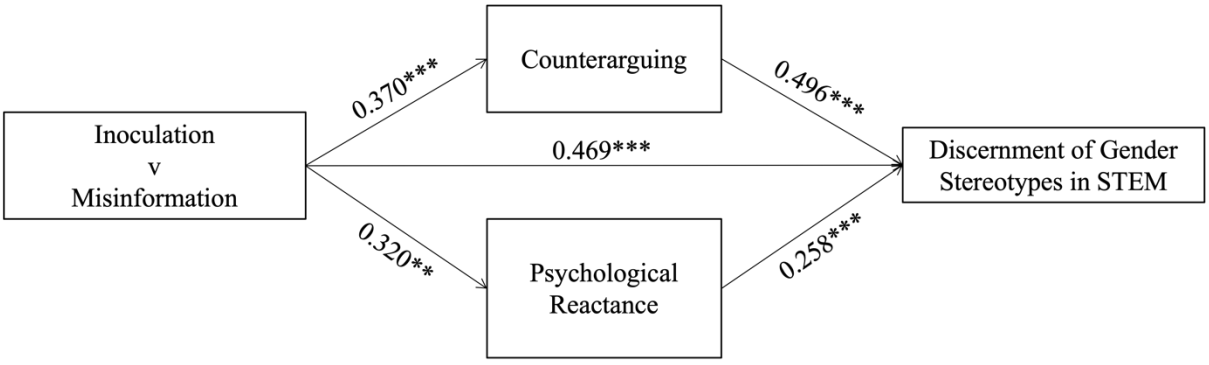
Next, the indirect effects of the intervention on discernment of gender-STEM stereotypes through psychological mechanisms were tested using mediation models. H5 was tested first, to examine the role of the cognitive and affective mechanisms posited in inoculation theory interventions (counterarguing and threat, respectively). Then, H6 was tested by examining treatment-elicited logical thinking, perspective taking, and empathic concern – the psychological mechanisms often associated with more traditional stereotype reduction research.

Inoculation Theory

To test the indirect effect of the inoculation condition on discernment of gender-STEM stereotypes through counterarguing and psychological reactance, a mediation model was tested (see Figure 3.8). The contrast comparing the inoculation condition to the misinformation condition was significant for both counterarguing (β = 0.37, 95% CI [0.145, 0.595], $p < .001$) and psychological reactance (β = 0.320, 95% CI [0.094, 0.545], $p = 0.006$) in response to the misinformation video. As hypothesised, participants who watched the inoculation treatment video experienced greater psychological reactance and engaged in more counterarguing when subsequently viewing the misinformation video than participants who watched only the misinformation video. Counterarguing was positively associated with discernment of gender stereotypes in STEM (β = 0.496, 95% CI [0.385, 0.607], $p < .001$) and, to a slightly lesser extent, so was psychological reactance (β = 0.258, 95% CI [0.148, 0.369], $p < .001$). The inoculation condition versus misinformation condition contrast revealed significant direct and total indirect effects, suggesting partial mediation (see Table 3.4). To ensure robustness of the model, it was re-run controlling for gender, and results did not meaningfully change (see Appendix B; Supplementary Figures and Supplementary Tables). Overall, these cognitive and affective tenants of inoculation theory are supported by the model as the inoculation intervention had a significant indirect effect on discernment of gender stereotypes in STEM through misinformation-elicited counterarguing and psychological reactance.

Figure 3.8

Model Testing the Indirect Effect of the Inoculation Intervention on Gender-STEM Stereotype Discernment Through Counterarguing and Psychological Reactance



Note. Coefficients are standardized. ** $p < .01$; *** $p < .001$.

Table 3.4

Model Estimates for Indirect Effect of the Inoculation Intervention on Gender-STEM Stereotype Discernment through Counterarguing and Psychological Reactance

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Inoculation v Misinformation → DOGSS	0.469	0.072	6.481	<.001	0.327	0.611
Indirect effects						
Inoculation v Misinformation → Counterarguing → DOGSS	0.183	0.061	3.024	0.002	0.065	0.302
Inoculation v Misinformation → Psychological Reactance → DOGSS	0.083	0.035	2.372	0.018	0.014	0.151
Total effects						
Inoculation v Misinformation → DOGSS	0.735	0.109	6.771	<.001	0.522	0.948
Total indirect effects						
Inoculation v Misinformation → DOGSS	0.266	0.083	3.199	0.001	0.103	0.429

State Empathic Concern, Perspective Taking, and Logical Thinking

To explore H6, the relative roles of treatment-elicited empathic concern, perspective taking, and logical thinking in mediating the relationship between the intervention conditions and discernment of gender-STEM stereotypes, a mediation model was tested on the inoculation condition (see Figure 3.9) and debunking condition (Figure 3.10).

The inoculation versus control contrast demonstrated a significant effect on all three state variables, however, only state empathic concern held a significant positive relationship to discernment of gender-STEM stereotypes (see Table 3.5). Similarly, the debunking versus control contrast demonstrated a significant effect on all three state variables, but only state empathic concern held a significant positive relationship to discernment of gender-STEM stereotypes (see Table 3.6)

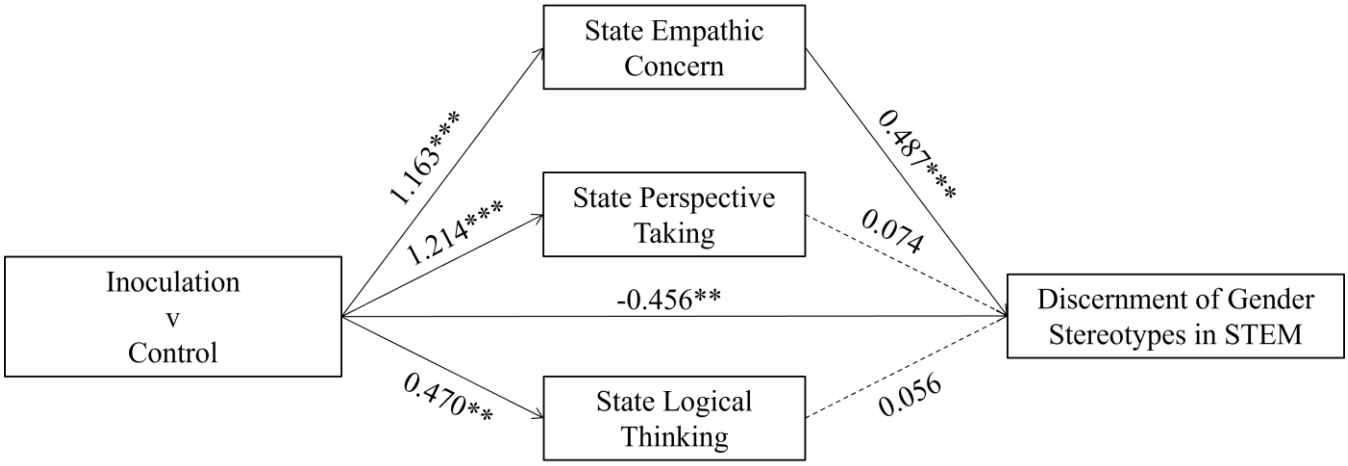
Thus, both models revealed a significant indirect effect of intervention on discernment in gender-STEM stereotypes through state empathic concern. However, significant indirect effects were not found through state perspective taking or state logical thinking. Additionally, while the total effect of the model estimates was significant for the debunking condition, the inoculation condition did not quite reach significance *(p* = .051), suggesting the role of state empathic concern was less explanatory in the inoculation intervention than the debunking intervention. This is consistent with typical stereotype- and prejudice-intervention literature that finds a significant effect of empathy on outcome

measures when they are measured directly post-intervention. The inoculation condition, however, examines the more real-world effect of the intervention-elicited psychological states by measuring outcome variables after exposure to misinformation.

Further, when the models were re-run controlling for gender and the related trait variables (trait empathic concern, trait perspective taking, and trait cognitive rigidity), the indirect effects became non-significant for both inoculation participants and debunking participants (Appendix C; Supplementary Tables and Supplementary Figures). This indicates that the participants' gender and trait-level empathic concern, perspective taking, and logical thinking accounted for the variance initially explained by state empathic concern. Therefore, these three state-level cognitive and affective mechanisms did not mediate the relationships between either of the intervention conditions and discernment of gender stereotypes in STEM.

Figure 3.9

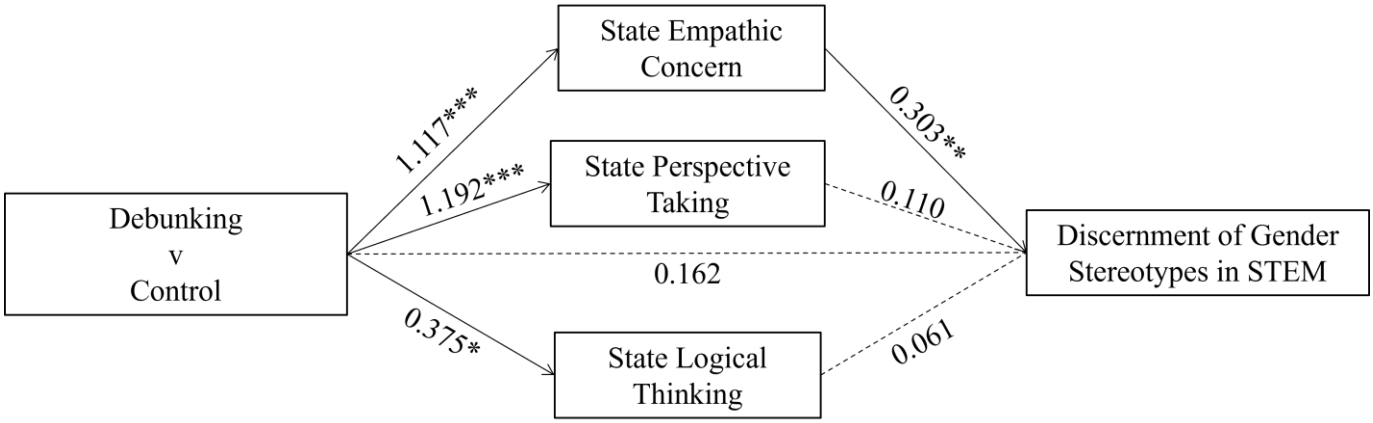
Model Testing the Indirect Effect of the Inoculation Intervention on Gender-STEM Stereotype Discernment through Empathic Concern, Perspective Taking, and Logical Thinking



Note. Coefficients are standardized. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 3.10

Model Testing the Indirect Effect of the Debunking Intervention on Gender-STEM Stereotype Discernment through State Empathic Concern, State Perspective Taking, and State Logical Thinking



Note. Coefficients are standardized. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 3.5

Model Estimates for Indirect Effect of the Inoculation Intervention on Gender-STEM Stereotype Discernment through Empathic Concern, Perspective Taking, and Logical Thinking

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Inoculation v Control → DOGSS	-0.456	0.145	-3.138	0.002	-0.741	-0.171
Indirect effects						
Inoculation v Control → State Empathic Concern → DOGSS	0.566	0.128	4.440	<.001	0.316	0.816
Inoculation v Control → State Perspective Taking → DOGSS	0.090	0.122	0.733	0.463	-0.150	0.329
Inoculation v Control → State Logical Thinking → DOGSS	0.026	0.033	0.803	0.422	-0.038	0.091
Total effects						
Inoculation v Control → DOGSS	0.226	0.116	1.951	0.051	-0.001	0.453
Total indirect effects						
Inoculation v control → DOGSS	0.682	0.118	5.793	<.001	0.451	0.913

Table 3.6

Model Estimates for Indirect Effect of the Debunking Intervention on Gender-STEM Stereotype Discernment Through State Empathic Concern, State Perspective taking, and State Logical Thinking

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Debunking v control → DOGSS	0.162	0.139	1.163	0.245	-0.111	0.434
Indirect effects						
Debunking v control → State Empathic Concern → DOGSS	0.338	0.131	2.575	0.010	0.081	0.596
Debunking v control → State Perspective Taking → DOGSS	0.131	0.137	0.956	0.339	-0.137	0.399
Debunking v control → State Logical Thinking → DOGSS	0.023	0.028	0.822	0.411	-0.032	0.077
Total effects						
Debunking v control → DOGSS	0.654	0.109	5.989	<.001	0.440	0.868
Total indirect effects						
Debunking v control → DOGSS	0.492	0.105	4.690	<.001	0.286	0.698

Application of Gender Stereotypes

To test the influence of the intervention on decision-making, between-groups analyses were conducted on the evaluations of the male/female STEM applicant. Analysis for H7 is presented first, looking at the hireability evaluations by intervention condition and applicant gender. Next, H8 is examined by looking at agency ratings by intervention condition and applicant gender.

Hireability

Contrary to our hypothesis, a 2 (applicant gender; male, female) x 4 (experimental condition; debunking, inoculation, misinformation, control) between-subjects MANOVA revealed no significant interaction between experimental condition and applicant gender on measures of applicant hireability, likeability, competence, leadership potential, and suggested salary, F (12, 1431.643) = 0.932, Wilk's $\Lambda = 0.980$, partial $\eta^2 = 0.007$, $p = .514$.

There were no main effects of experimental condition. However, there was a main effect of applicant gender; surprisingly, across conditions, the male CV received lower ratings than the female CV on competence (M_{diff} = -0.217, 95% CI [-0.353, -0.067],

t(548.652) = -2.362, p_{tukey} = .004, *d* = -0.240), likeability (M_{diff} = -0.667, 95% CI [-0.841, -0.493], *t*(567.478) = -56.53, p_{tukey} < .001, *d* = -0.630), leadership potential (M_{diff} = -0.522, 95% CI [-0.729, -0.351], *t*(561.764) = -5.723, p_{tukey} < .001, *d* = -0.484), and suggested salary (M_{diff} = -916.857, 95% CI [-1237.143, -617.903], *t*(567.957) = -5.792, p_{tukey} < .001, *d* = -0.485). Yet, despite these differences, there was no effect of applicant gender on ratings of hireability, *t*(543.493) = -1.544, *p* = .123, *d* = -0.129. Therefore, the female applicant was rated as significantly more competent, more likeable, with greater leadership potential, and recommended a higher starting salary than the male applicant, but she was not perceived as more hireable.

To ensure this effect of applicant gender was not driven by possible group differences between participants randomly assigned to the female applicant in relation to those randomly assigned to the male applicant, pairwise comparisons were run on the gender stereotype literacy variables. There were no significant differences; however, participants randomly assigned to the female applicant tended to be more aware of gender bias in STEM *(p* =.053, *d* = .163) and tended to have more understanding that gender stereotypes can negatively impact women's chances of being hired *(p* =.074, *d* = .147) than participants assigned to the male applicant; all other *ps*>.109. Therefore, these random group differences may have contributed to the findings.

Agency

In line with our hypothesis, a 2 (applicant gender; male, female) x 4 (experimental condition; debunking, inoculation, misinformation, control) ANOVA revealed no significant interaction between experimental condition and applicant gender on agentic ratings of the applicant, *F* (3, 544) = 1.13, partial η^2 = .006, *p* = .337. However, as expected and in line with past research (Eagly & Steffen, 1984; Rosette & Tost, 2010), there was a main effect of applicant gender, such that the female applicant was rated as more agentic *(M* = 4.02, *SD*= .57) than the male applicant *(M* = 3.87, *SD*= .47), *F* (1, 549.81) = 11.33, η^2 = .047, *p* <.001) across conditions.

Gender x Condition Interaction Effects and Gender Differences

In line with $H_0$9 and findings from the previous chapter, gender did not interact with condition to predict discernment of gender stereotypes in STEM, *p* = .93, but across

conditions women demonstrated greater discernment of gender stereotypes in STEM than men (M_{diff} = 0.57, *SD*= .09, *t(*598.22) = 6.188, p_{tukey} <.001, *d* = .522), as predicted by H10.

Similarly, there was no interaction between gender and condition on male-agentic stereotype endorsement, p_{tukey} = .629. Contrary to the pilot study, but in line with other research (e.g., Hentschel et al., 2019), men rated women as less agentic *(M* = .501, *SD*= 0.683) than women rated women (M_{diff} = -.294, 95% CI [-0.400, -0.188], *t(*596.985) = -5.742, *d* = -.471, p_{tukey} <.001). This may be attributed to the improved measure of agency from the pilot study. These findings suggest men, on average, hold greater male-agentic stereotypes than women.

Likewise, gender did not interact with intervention condition to predict gender bias awareness in STEM, p_{tukey} = .522. Also replicating findings from the pilot study and previous research (Pietri et al., 2017; Luzzo & McWhirter, 2001), women *(M* = 5.758, *SD*= 1.265) were more aware of gender bias in STEM than men irrespective of experimental condition (M_{diff} = 1.038, 95% CI [0.825, 1.250], *t(*591.077) = 9.607, *d* = 0.565, p_{tukey} <.001). aware

Additionally, gender did not interact with condition to predict attitudes toward affirmative action *(p* = .819), modern sexism *(p* = .898), or hostile sexism *(p* =.817), but gender did have a main effect on all three forms of sexism, such that women reported lower levels of modern sexism (M_{diff} = -1.020, 95% CI [-1.227, -.813], *t(*588.513) = -9.917, *d* = -.815, p_{tukey} <.001), hostile sexism (M_{diff} = -1.238, 95% CI [-1.460, -1.015], *t(*596.514) = -11.229, *d* = -.920, p_{tukey} <.001; *U* = 22343, *r* = -.50, *p* <.001), and neosexism (i.e., more positive attitudes toward affirmative action; *Mdiff* = -1.006, 95% CI [-1.265, -.747], *t(*587.056) = -7.600, *d* = -.641, p_{tukey} <.001), in line with past research (e.g., Luzzo & McWhirter, 2001).

In sum, $H_0$9 was supported, as gender did not interact with condition to predict any of the intervention outcomes. Alternatively, as predicted by H10, across all gender stereotype outcome measures, women demonstrated an advantage over men: greater discernment of gender stereotypes in STEM, greater awareness of gender bias in STEM, lower sexist attitudes, and lower male-agentic stereotype endorsement.

To investigate the moderating role of participant gender on applicant ratings, a 2 (participant gender; male, female) x 2 (applicant gender; male, female) x 4 (experimental condition; debunking, inoculation, misinformation, control) ANOVAs were run. There were no three-way interaction effects of participant gender x applicant gender x condition on any applicant evaluations (hireability, leadership potential, competence, suggested salary; all *ps*>.480).

Discussion

Overall, there was an intervention effect on discernment of gender stereotypes in STEM. Additionally, inoculation theory's dual-process model of threat and counterarguing was supported. There were no significant differences between the debunking and inoculation conditions, but the extent to which the intervention demonstrated inoculative effects was less than its persuasive effects. While inoculative effects were demonstrated on some items of discernment, there were no differences between the control condition and inoculation condition on the composite score of discernment. As post-hoc comparisons and corrections were performed across all conditions (not only those of interest), and not all post-hoc comparisons were relevant or hypothesized, this impacted the reported statistical performance of the treatment conditions. This statistically conservative approach was taken due to not pre-registering specific contrasts of interest.

The greater efficacy of the debunking intervention is consistent with typical stereotype- and prejudice-intervention literature that finds a significant effect of treatment immediately post-intervention. The inoculation condition, on the other hand, may demonstrate a more real-world effect of the intervention by measuring outcome variables after subsequent exposure to misinformation. Similarly, this supports intervention research that tends to find less of an intervention effect at follow-ups after people have likely been exposed to counter information and attitudes in their social worlds. To the best of my knowledge, this is the first research paradigm to examine an optimised debunking or inoculation theory intervention in the context of stereotype discernment.

While this research examined cognitive and affective mechanisms, social mechanisms are undoubtedly powerful in experiencing and maintaining psychological change. Extensive research shows perceived normative appropriateness of prejudice influences the extent to which people express prejudiced attitudes and behaviours (Duguid & Thomas-Hunt, 2015; Crandall & Stangor, 2005; Stangor et al., 2001; Crandall, Eshleman, & O'Brien, 2002; Brief, Dietz, Cohen, Pugh, & Vaslow, 2000). Additionally, communicating a social norm to support the intervention message can help to facilitate persuasion (Paynter et al., 2019; Hornsey & Fielding, 2017). Moreover, post-inoculation talk (word-of-mouth communication) has been put forth as a promising 'top-up' technique in inoculation research (Compton & Pfau, 2009; Ivanov et al., 2015). Conversely, when individuals are faced with a persuasive message that is incongruent with their current beliefs or attitudes, they will actively recall others in their social groups (family, friends, coworkers) who share their existing belief or attitude

(Zuwerink, 2003). This social validation process helps them protect their belief or attitude. However, little research has explored the role of social mechanisms in sustaining or deteriorating stereotype-reduction efforts.

Thus, it may be fruitful to investigate the role of counterarguing (cognitive), threat (affective), psychological reactance (cognitive-affective), and post-talk inoculation (social) in the longevity of inoculation effects on stereotype discernment. As with most psychological interventions, if not all, inoculation is found to have decay effects. While 'booster shots' (follow-up treatments) are recommended to buffer decay effects, it would be useful to investigate the importance of these psychological mechanisms both in the main intervention and subsequent boosters for sustained psychological change. Thus, it would be worthwhile for future research to conduct follow-ups to examine the long-term influence of cognitive, affective, and social mechanisms on stereotype change.

The mediation model examining the relative influence of empathic concern, perspective taking, and logical thinking was not well supported. The findings suggest empathic concern may be an important psychological mechanism in psychological interventions – irrespective of perspective-taking. This advances stereotype- and prejudice-intervention research and, specifically, the assumption that empathy is elicited *through* perspective taking. However, as empathic concern was associated with discernment scores immediately following the intervention, and was largely explained by trait empathic concern, the influence of treatment-elicited empathic concern relationship may be state-dependent and fleeting.

Additionally, we did not instruct participants to take the view of the target person's (woman in STEM) perspective, as traditionally done to evoke perspective-taking in intervention research (e.g. Dovidio et al., 2004) based on Batson et al. (1997) seminal study. It is likely that, if our design involved a more explicit manipulation of perspective taking, we would have seen a greater indirect effect through perspective taking. However, the aim of this research was to examine the natural state processes involved in inoculation and debunking interventions.

Some scholars suggest there are two forms of perspective taking: imagining how another feels and imagining how you would feel (Batson et al., 1997). Imagining how the other feels is linked to empathy, which can evoke altruistic motivation, whereas imagining how you would feel can evoke empathy *and* personal distress, which can evoke egoistic motivation. This distinction may be important for future intervention research. Our study

examined how the individual would feel, which may help to explain the non-significant results.

Overall, the relative importance of empathic concern, perspective taking, and logical thinking in psychological change may be better explained by trait-level than state-level. For example, someone who has a propensity for experiencing empathy may be more sensitive to affective mechanisms; likewise, someone who has a propensity for logical reasoning may be more influenced by cognitive components of interventions. Therefore, while the intervention significantly increased all three, individual differences may then account for their relative importance in outcome measures. Thus, designing successful interventions may not be about maximising one psychological mechanism or another, but rather ensuring *both* affective and cognitive components are involved.

Another possible explanation for the reduced discernment effects in the inoculation condition compared to the debunking condition may be related to the time between the treatment and misinformation attack in the inoculation condition. It may have been optimal to include more of a delay between the treatment video and the misinformation video. Following the medical inoculation metaphor, McGuire (1961) suggested that sufficient delay was required between the treatment and attack to allow time for antibodies to build. Although Banas and Rains' (2010) meta-analysis did not find support for the notion of optimal delay, this finding may not uphold in the novel context of stereotyping.

As expected, hostile sexism, modern sexism, and neosexism (measured as support for affirmative action) did not differ between conditions. This suggests that even though *discernment* of gender-STEM stereotypes improved – away from the myth belief and towards the "correct" belief – it did not translate to *attitude* change. Likewise, the intervention did not translate to differences in *decision-making* (applicant hireability). Therefore, the intervention did not demonstrate a blanket of protection, in the language of inoculation scholars. These findings are consistent with the continued influence effect, such that belief change doesn't necessarily translate to attitude or behaviour change. The findings also uphold Vinacke (1957)'s position that stereotypes and prejudice should be conceptualised as separate constructs.

The lack of differences in the hireability evaluations of the identical male and female applicants initially appears as an unbiased reflection of their identical CVs. However, considering the perceived superiority of the female applicant in agency, competence, likeability, and leadership potential, this finding may be interpreted differently. That is, even when a woman in STEM applicant is perceived as more competent, more likeable, and

possessing greater leadership potential than her male counterpart, she is still not perceived as more hireable.

Further, based on the distributions of data on the applicant rating measures, it appears there may have been an influence of ceiling effects. Given the details in the job application (see Appendix B; Job Description) it may have presented an overqualified candidate. These findings may suggest that overqualified women do not experience the effects of gender bias in STEM to the same extent as other women in STEM. As the job description and job application were novel materials, a pilot test should have been conducted on them before this study. Indeed, Dovidio and Gaertner (2000) found that when Black and White candidates' qualifications were very strong, Black candidates do not face hiring discrimination. However, when the candidates' qualifications were moderate or weak, White candidates were recommended significantly more often than Black candidates with the exact same credentials (76% vs 45% and 13% vs 6%, respectively). Thus, replication of this research using an "average" applicant is needed.

The present research did not find a moderating effect of gender on intervention for any outcome variable (discernment of gender stereotypes in STEM, awareness of gender bias in STEM, male-agentic stereotype, hostile sexism, modern sexism, neosexism, or applicant ratings). These findings support our hypotheses and corroborate previous research on gender bias interventions that target *explicit* beliefs, attitudes, and awareness (e.g., Jackson et al., 2014; Pietri et al., 2017; Zawadzki et al., 2012; Moss-Racusin et al., 2018; Chapter 2).

However, as expected, there was a main effect of gender, favouring women, on all outcome variables. This is also consistent with previous research findings that men demonstrate greater gender-trait stereotype endorsement (Hentschel et al., 2019), less knowledge about sexism than women (Luzzo & McWhirter, 2001) and less awareness of gender bias than women (Pietri et al., 2017). It is worth highlighting that, contrary to the pilot study which found no gender difference, men rated women as less agentic than women rated women. This finding may be attributed to the improved measure of agency from the pilot study, as it is in line with other research (e.g., Hentschel et al., 2019). Taken together, the findings suggest that, compared to women, men hold greater gender stereotypes and sexist attitudes, as well as less discernment and less awareness of the effects of gender stereotypes in STEM. This gender difference may be because men demonstrate more scepticism about gender bias evidence than women (Moss-Racusin et al., 2015; Moss-Racusin et al., 2018).

Limitations

The overall sample demonstrated high levels of education and awareness of gender bias, resistance to sexism and gender stereotypes, and a liberal political orientation. Therefore, the sample may not be representative, and findings may not be generalisable.

Further, participants were drawn from the general population and were not necessarily in positions of recruitment or STEM. Nonetheless, the application (and reduction) of gender stereotypes in the evaluation of women in STEM among the general population is relevant to investigate because people who are not in STEM, or STEM recruitment, can still impact gender diversity and what gender stereotypes look like in STEM through social influence (as parents, teachers, peers, and so on). Indeed, substantial research shows the influential role of parents (e.g., Jungen, 2008; Lloyd et al., 2018), teachers (e.g., Lavy & Sand, 2015; Roberts & Hill, 2003), peers (e.g., Blakemore et al., 2009; Fabes et al., 2014), and the media (e.g., Penhall, 2018; Marquez & Porras, 2020) on shaping gender stereotypes and career interests.

We cannot assume participants actively paid attention to the entire video; it is not possible to know if they may have been distracted, multi-tasking, or walked away from their device. This can impact the validity and reliability of the findings. However, it may also be argued that we cannot guarantee all individuals pay close attention to intervention material in laboratory or field settings either.

The perceived credibility of the intervention videos was not matched. Although all three videos (treatment, misinformation, and control) achieved acceptable credibility (group means above the midpoint on the semantic scale), the misinformation video still received a credibility rating significantly lower than the treatment and control videos. This may have impacted the persuasiveness of the misinformation video. The misinformation video's lower credibility rating may be explained in part by the unmatched video lengths (4:07-minute treatment video, 3:39-minute control video, and 1:00-minute misinformation video).

However, as the sample appeared to enter the experiment with an overall resistance to gender-STEM stereotypes (as indicated by grand means as well as the control group's means), a longer misinformation video may not have greatly impacted its credibility or effectiveness. Future research may want to consider matching, or manipulating, the lengths of the treatment message and attack message. To the best of my knowledge, this has not been previously studied or addressed in inoculation research.

Additionally, the lower credibility of the misinformation video may be related to the quality of the video – the YouTuber communicates in a highly-opinionated way, and the points argued may not be as convincing as other arguments for gender differences in STEM

aptitude. Finally, it is possible the sample had strong pre-existing discernment of gender stereotypes in STEM and, therefore, resistance to gender-STEM misinformation. The sample was highly educated and liberal in political orientation, and endorsement of sexism and gender stereotypes was low across all conditions. However, pre-intervention discernment was not directly measured and therefore cannot be inferred.

Due to limited variance and floor effects across conditions, the male-agentic stereotype endorsement may no longer be an accurate measure of gender stereotypic traits, as measured by the PAQ agentic traits. Although there was a significant gender difference, the averages for both women *(M* = 0.2, *SD*= .048) and men *(M* = 0.4, *SD*= 0.57) fall closest to the 0 point on the scale, representing *equally true of men and women*. The PAQ items have not been updated since initial scale development 5 decades ago. Considering the substantial changes in gender roles since the 1970s, it is likely that what is now considered masculine or feminine in society has also changed. Thus, theories and measures of agency related to gender stereotypes may need to be updated to better reflect the social changes of the past 50 years.

It is also important to note that the participants' decision-making process likely did not reflect the weight and consideration given to a typical hiring process, as participants knew there were no real effects or consequences of their decision. The hypothetical nature of the study lacked mundane realism and impacted ecological validity, as participants may have responded differently in a real-world scenario.

Additionally, while sexism was treated a dependent variable, it could also be considered a moderating variable. An international workplace gender-stereotype intervention found the intervention appeared to increase women's tendency to engage in inclusive behaviours toward other women among US employees, but the positive effect on behaviours was concentrated among those who already demonstrated supportive attitudes towards women (Chang et al., 2019). Thus, the success of the intervention was moderated by the individuals' pre-existing levels of sexist attitudes.

Importantly, this research did not assess prior beliefs or attitudes related to gender-STEM stereotypes. Therefore, this study was not able to examine psychological *change* in relation to the intervention. This means we were not able to differentiate between therapeutic versus prophylactic inoculation effects.

CHAPTER FOUR – INVESTIGATING A GENDER-STEM STEREOTYPE TRAINING AND AN UNCONSCIOUS BIAS TRAINING AMONG ENGINEERING STUDENTS: A FIELD STUDY

In Chapters 2 and 3, I demonstrated the effects of the novel debunking and inoculation theory intervention. However, as pre-intervention beliefs were not assessed, I was not able to assess within-participant psychological change. Additionally, the intervention was tested using online samples from the general population. Thus, in this chapter, I sought to (1) examine the effects of the intervention on psychological change and (2) test the efficacy of the intervention in the field using a sample of university engineering students.

Introduction

One of the most popular efforts within male-dominated fields to address the gender imbalance is diversity trainings (Dobbin & Kalev, 2016). Annually, billions of dollars are invested in diversity programmes to reduce prejudice and discrimination (Hansen, 2003; Jayne & Dipboye, 2004). However, the effectiveness of these programmes and, more specifically, which mechanisms work well and which don't, has not been clearly identified (Paluck, 2006; Paluck & Green, 2009; Anand & Winters, 2008; Chang et al., 2019). Paluck and Green (2009) and Paluck et al. (2021) summarize their reviews on prejudice and discrimination reduction procedures with a call for both theory development and evidence from the field. As prejudice and discrimination largely result from negative stereotyped beliefs, diversity training efforts may be well spent targeting stereotype endorsement and application.

Diversity Trainings

Diversity trainings are born out of the assumption, and some research support (Fiske 1998; Nelson et al., 1996), that educating people about biases - and condoning them - reduces their expression of them. However, diversity training can also cause backlash by activating bias (Kidder et al., 2004; Sidanius, Devereux, & Pratto, 2001; Paluck & Green, 2009) or by 'normalising' that everyone has bias (Duguid & Thomas-Hunt, 2015; Stangor et al., 2015), and, thus, negatively impact the promotion of minorities (Naff & Kellough, 2003). Unconscious bias trainings, in particular, are often treated as a silver bullet for gender equity in the workplace, but the effects are likely to be negligible unless more sustained

interventions are also implemented (Williamson & Foley, 2019). Overall, unconscious bias trainings are often a tick-box exercise and do not typically affect change.

In a systematic analysis of federal and private sector data, Kalev et al., (2006) found that managerial diversity training was the least effective programme for increasing representation of women, demonstrating either no effect or a negative effect – suggesting a backfire effect. They note that, although the data indicates managerial bias and social isolation of women drive gender inequality in the workplace, the best approach to addressing underrepresentation is not to target the root causes directly (by adopting diversity training or social-isolation reduction programmes, respectively), but to instead establish organisational responsibility for diversity. For example, organisational structures such as affirmative action plans, task forces, and full-time diversity staff appear to be most effective in improving representation of women. This distinction speaks to the issue of inaction when diversity is a responsibility for everyone, but not a primary responsibility for anyone.

However, affirmative action plans have also been found to backfire (Bond & Pyle, 1988; Linnehan & Konrad, 1999). Although they may rectify some systematic gender inequality by increasing the number of women hired, colleagues and executives may view women who are hired as 'tokens' of diversity (Hogg & Vaughan, 2008) and viewed as less competent (Heilman et al., 1996, 1998; Major et al., 1994). This is the idea that females benefit from positive discrimination and receive positions not because of their qualifications or capability, but because their employer wants to appear inclusive. This can create resentment and a toxic work culture. Further, the hired women themselves may also question their competence, decreasing their self-efficacy and sense of belonging and increasing stereotype threat.

To mitigate the negative effects of some diversity initiatives, research finds that the more diversity initiatives in the organisation the better (Kalev et al., 2006; see Bezrukova et al., 2016 for meta-analysis) – a bit of a 'kitchen sink' approach. This may be because 'diversity initiatives' is a fairly broad umbrella term that encapsulates several different modes and methods of addressing the effects of bias in the workplace, some effective and others inflammatory. So what actually works, and why? Little research has reconciled the conflicting effects of diversity interventions (Bezrukova et al., 2016).

Effects of Gender-STEM Stereotypes within STEM

Research shows that perceived gender discrimination is the most predictive factor in women's selection of a college major. College-bound women report that they do not enter STEM fields, not because STEM is math or science oriented, or because there is an unequal gender distribution, but because of the degree of perceived discrimination (Ganley et al., 2017). This finding remains even when males and females are matched on prior backgrounds, achievements, and interests. This perception appears to be a valid concern, with several studies showing evidence of bias against female students in STEM subjects (Moss-Racusin et al., 2012; Milkman et al., 2015) as well as women in science faculties (Bilimoria & Liang, 2013). In a study of aspiring STEM high school girls, undergraduate women in STEM courses, and female doctoral students in STEM, 61% reported experiencing gender bias, largely from male peers (Robnett, 2016).

Not surprisingly, then, women who do enter agentic fields such as STEM are also less likely than men to remain in them (Turner, 2002). And when they do remain, they are less likely to receive promotions (Ong et al., 2011). This has been described as the "leaky pipeline effect": the progressive loss of women in the STEM fields (Ysseldyk et al., 2019). This loss is described as progressive because the reduction in the number of women becomes more pronounced in more senior positions (Martinez et al., 2007). For example, while the percentage of women obtaining life science doctoral degrees increased from 15% to 52% between 1969 and 2009 (National Science Foundation, 2008; National Science Foundation, 2012) the percentage of women in senior faculty positions is still low, with women representing only 18% of biology professors in 2007 (Nelson, 2007).

Gender-STEM Stereotype Threat

Indeed, one of the insidious effects of gender stereotypes, which may play a role in the slow-reducing gender gap in male-dominated fields, is stereotype threat. Stereotype threat refers to the feelings of anxiety or fear of falling into or confirming the negative stereotype of one's social group (Steele & Aronson, 1995). Extensive research has found that activating group stereotypes elicits behaviour that confirms the stereotypes (see Wheeler & Petty, 2001).

In the context of STEM, the descriptive stereotype that men are better suited for STEM subjects and the prescriptive stereotype that men - not women - should be interested in STEM (Koenig, 2018; Martin, 1995), creates a sense of pressure and inadequacy among

women, which is potentiated by the fear of falling into or confirming the stereotype. Indeed, when women are primed with the stereotype that 'women perform worse than men in math', they then perform worse on a math test, compared to when the stereotype is not mentioned (Spencer et al., 1999). Further, when women are told that men and women perform equally on a math test, they then perform equal to men on the math test (Spencer et al., 1999).

Since the seminal study by Spencer and colleagues in 1999, further work has also found evidence of stereotype threat in STEM. For example, research has found that negative stereotypes about women's competence in maths stimulates a threatened sense of competence in women, which increases anxiety and interferes with available working memory capacity, which in turn results in worse performance (Schmader & Johns, 2003). Further, a meta-analysis carried out in 2015, which aimed to investigate the effects of stereotype threat on girl's performance in math, science and spatial skills, found evidence that their performance was negatively affected by stereotype threat *(d* = -0.22; Flore & Wicherts, 2015). It should be noted, however, that the authors acknowledged the issue of publication bias, which could affect the estimated effect size of stereotype threat.

In a more recent, general (not specific to STEM) large-scale meta-analysis, which included 212 studies and a total of over 10,000 participants, authors examined stereotype threat in the context of high-stakes testing settings, such as employment or college admission tests (Shewach et al., 2019). The authors concluded the effects of stereotype threat were small to negligible. However, when taking a closer look at the findings, *gender*-based stereotype threat demonstrated the greatest significant effects *(d* = |.19| to |.45|). Noteworthy, the authors separately analysed studies that used blatant versus subtle stereotype threat manipulations. In blatant-cue conditions, participants were explicitly told that women perform worse than men on the test. While this is an effective methodology for improving theoretical understanding of stereotype threat, it would be unlikely in a true professional setting due to ethical reasons. Conversely, subtle cues occur when the stereotype, or the message of gender differences in ability, is not communicated; therefore, this may offer a more realistic account of what occurs. Shewach et al. (2019) found that blatant cues produced a stronger effect in women *(d* = |.45|) than subtle *(d* = |.19|). However, Nguyen and Ryan's (2008) meta-analysis found that subtle activation of stereotype threat produced a larger effect in women *(d* = |.24|) than blatant *(d* = |.18|). Thus, findings regarding the differential power of blatant versus subtle cues are mixed, but the effect of stereotype threat on women appears consistent.

Relatedly, benevolent sexism has been found to impact women's performance, even on tasks and in domains that are not traditionally stigmatising against women's competence,

such as customer care, good social abilities, and cooperative skills (Dardenne et al., 2007). Therefore, subtly implying that women lack competence may lead women to consequently perform badly when they wouldn't have otherwise. In contrast, hostile sexism does not seem to impact women's performance, potentially because it is more direct and therefore more easily externalised (Dardenne et al., 2007). Similarly, Major et al. (2003) found that covert sexism was more detrimental than overt sexism on women's self-esteem.

Taken together, these findings on gender identification may help to explain the meta-analytic findings reported above by Nguyen and Ryan (2008), that women were more susceptible to the effects of stereotype threat in subtle versus blatant manipulation conditions; however, this does not support the contradicting findings by Shewach et al. (2019), which demonstrate that blatant conditions are more activating than subtle conditions. In the context of gender stereotype threat in STEM, this differentiation may be less relevant given that women often experience *both* subtle and blatant cues.

Gender Identity Centrality

Self-stereotyping is "the process by which people belonging to a stigmatised social group tend to describe themselves more with stereotypical traits as compared with traits irrelevant to the ingroup stereotype" (p. 92, Latrofa et al., 2012). According to self-categorization theory, self-stereotyping occurs with social identities that are salient. People adopt the beliefs, interests, and behaviours to match the norms, or stereotypes, of their group identity. Indeed, in Hentschel et al. (2019), women applied the stereotype of communality (concern for others, sociability, and emotional sensitivity) to themselves and tended to see themselves as more gender-stereotypical than other women.

Conversely, when stigmatised individuals are in outgroup-dominated domains, self-group distancing can occur as a way to cognitively protect themselves from social identity threat (Veldman et al., 2021; Keane et al., 2011). For example, women in STEM may psychologically distance themselves from their social identity of gender in order to buffer the effects of experiencing gender-STEM stereotype threat. Theoretically, by reframing a social identity as less important, one would be less vulnerable to the consequences of it being threatened.

However, gender identity centrality can also have protective effects in targeted domains. Women higher in gender identification are more likely to detect subtle discrimination than lower-identified women (Major et al., 2003). Their positive image of their gender group is protective as they are better able to identify and disregard negative

attitudes or prejudice toward their gender group. However, other research has found that, although strongly identified women may be better able to *recognise* subtle discrimination, they are still susceptible to the insidious *effects* on cognition and performance (Dardenne et al., 2007; Major et al., 2003). Thus, they suggest that gender identification can moderate the effects of hostile - but not benevolent - sexism on performance (Dardenne et al., 2007). As this research assesses discernment, we would expect that more strongly identified women would experience higher discernment. Indeed, when women are told there are no gender differences in a gender-stereotyped domain, they are more likely to indicate interest in it (Davies et al., 2005). Davies et al. (2005) suggest that, by specifying there is no gender difference, "identity safety" is achieved and, therefore, perceived threat and its consequential anxiety are reduced. Identity-safe environments appear to improve not only performance but aspiration as well (Davies et al., 2005).

Thus, to see considerable improvements in the gender gap in STEM, combined efforts of stereotype and stereotype-threat interventions may be necessary. Inoculation theory/debunking interventions may be able to serve as both. Through explicit messaging that there are no innate gender differences, the intervention may help to establish identity safety for women and treat stereotype threat, as well as debunk, or inoculate, both men's and women's discernment of gender-STEM stereotypes.

Present Study

Researchers caution against laboratory- and online-based findings in intervention research, suggesting there is a need for more field experiments to test the rigor of stereotype reduction interventions (see Paluck & Green, 2009; Paluck et al., 2021; Chang et al., 2019). The primary aim of this study was to investigate the effectiveness of a gender-STEM stereotype-reduction training among engineering university students. The stereotype-reduction training content and techniques were drawn from optimised debunking and inoculation theory. It followed the same intervention messages as the treatment video in Chapters 2 and 3, but was delivered as a presentation in person. This study builds on the previous chapters by (1) investigating endorsement of gender stereotypes in STEM *within* STEM, rather than the general population (2) evaluating psychological *change* in discernment pre-post intervention, including the backfire effect and mediating psychological mechanisms, and (3) comparing post-intervention outcomes against an unconscious bias training.

Gender-specific interventions appear to be more effective than general interventions (Chang et al., 2019). Likewise, previous research has indicated that, although both are effective, more specific inoculation messages are more effective than general ones in conferring psychological protection and change (van der Linden et al., 2017). While the current study did not aim to replicate the differential inoculative effects of specific- versus general-message inoculation interventions, we examined the differential effects of specific (stereotype reduction training) versus general (unconscious bias training) interventions on gender-STEM stereotype discernment. We expected the more specific gender-stereotype reduction training would improve discernment of gender-STEM stereotypes more than the general unconscious bias training. Additionally, we examined the generalised effect, also known as 'umbrella' protection or 'blanket' protection, of the two trainings on more general measures: gender-trait stereotype endorsement and sexist attitudes. As the interventions were both one-hour, one-off trainings, we didn't expect to see differences in gender-trait stereotype endorsement or gender-related attitudes. However, as the unconscious bias training was a more general treatment with more generalised messages, it is possible it could have greater direct effects on more general gender beliefs and attitudes.

Likewise, we expected psychological change in discernment only – not male-agentic stereotype endorsement. As in Chapter 3, the influence of state empathic concern, perspective taking, and logical thinking were examined for their relative effects on intervention effectiveness. Although treatment-elicited perspective taking, empathic concern, and logical thinking did not mediate the treatment effect in Chapter 3 (once gender and trait variables were controlled), the analyses conducted were between-group and on post-intervention scores only. As such, in this chapter, we examine the relative influence of these psychological mechanisms within-group, on experiencing psychological *change* in stereotype discernment, among the stereotype-reduction training participants.

Finally, within the stereotype-reduction training, gender differences were explored, as well as the interactive role of gender identity centrality and experience of gender discrimination in STEM on stereotype discernment.

To note, although the term 'stereotype reduction' is used, it is important to clarify that the goal was not to change participants' mental models or schemas. Such a goal would be quite lofty and amiss. Rather, the aim was to reduce the *use* of stereotypes by increasing participants' discernment of stereotypes.

Method

Participants

Participants (N = 167) were engineering students at a Canadian university, recruited through their engineering course (for the stereotype-reduction training) or their engineering teaching assistant role (for the unconscious bias training). Eligibility was restricted to individuals over the age of 18 and fluent in English. Thirteen participants did not complete the surveys or withdrew from the study, resulting in N = 154 for analyses. Data cleaning procedures checked for straight-lining, and no participants were removed. Participants were not paid for participating in this study. Demographic information is provided in Table 4.1. The demographic information for the unconscious bias training group was completed by only 30% of the group.

Table 4.1

Sample Demographics

	Stereotype-Reduction Training (n = 85)		Unconscious Bias Training (n = 60)	
	n	%	*n*	%
Ethnicity				
White	43	51.8	7	33.3
Black	1	1.2	1	4.8
Asian	13	15.7	7	33.3
Chinese	9	10.8	0	0
Latino/Hispanic	2	2.4	1	4.8
Mixed	4	4.8	0	0
Aboriginal/First Nations	4	4.8	0	0
Other	7	8.4	5	23.8
Gender				
Female	17	20.7	6	28.6
Male	64	78	13	61.9
Transgender female	0	0	0	0
Transgender male	0	0	1	4.8
Non-conforming/queer	0	0	1	4.8
Prefer not to say	1	1.2	0	0
Age[a]				
18-25	66	85.8	7	33.4
26-30	9	11.6	11	52.3
31-35	1	1.3	3	14.3
36-40	1	1.3	0	0
Religion				
No religion	48	57.1	8	38.1

Religious	36	42.9	13	61.9

Note. [a] Age was collected as a continuous variable, but is presented as categorised data here for ease of interpretation. Percentage values represent valid percentages.

Procedure

Participants were assigned to either the stereotype discernment training or the unconscious bias training. In both conditions, the study consisted of an in-person presentation with PowerPoint slides (the training) and a series of questionnaires. The total duration of the study was 60 minutes. I will now detail the procedure for each training condition.

The stereotype discernment training was conducted during class time in a university lecture theatre (n = 85). Upon entering the classroom, participants were given a consent form to read through. After informed consent was provided, consent forms were collected and the first set of questionnaires were completed (Male-Agentic Stereotype, Discernment of Gender Stereotypes in STEM). Next, participants received the stereotype-reduction training. At the end of the training, participants completed another set of questionnaires (Training Credibility, Psychological Reactance, State Perspective Taking, State Empathic Concern, State Logical Thinking, Perceived Increase in Knowledge, Affirmative Action, Hostile Sexism, Male-Agentic Stereotype, Male-Logical Stereotype, Modern Sexism, Trait Perspective Taking, Trait Empathic Concern, Trait Logical Thinking, Gender Identity Centrality, Experiences in STEM, and Discernment of Gender Stereotypes in STEM). To conclude, all participants were thanked for their time, given a debriefing sheet, and invited to contact the researcher with any questions or concerns.

The unconscious bias training was conducted as part of the University's Equity and Inclusion Office training for teaching assistants in the engineering department (n = 69). Upon entering the classroom, participants were given a consent form to read. After informed consent was provided, consent forms were collected and participants received the unconscious bias training. At the end of the training, participants completed a set of questionnaires (Training Credibility, Psychological Reactance, State Perspective Taking, State Empathic Concern, State Logical Thinking, Perceived Increase in Knowledge, Affirmative Action, Hostile Sexism, Male-Agentic Stereotype, Male-Logical Stereotype, Modern Sexism, Trait Perspective Taking, Trait Empathic Concern, Trait Logical Thinking, and Discernment of Gender Stereotypes in STEM). To conclude, all participants were thanked for their time, given a debriefing sheet, and invited to contact the researcher with any questions or concerns.

It was hypothesised that:

$H_0$1a-d: The two intervention conditions would not differ in treatment-elicited empathic concern (a), perspective taking (b), logical thinking (c), or perceived increase in knowledge (d).

H2: The stereotype reduction training condition would report higher discernment of gender-STEM stereotypes than the unconscious bias training condition

$H_0$3a: The unconscious bias training and stereotype reduction training would not differ on levels of male-agentic stereotype endorsement

$H_0$3b: The unconscious bias training and stereotype reduction training would not differ on levels of sexism

H4a: Within the stereotype reduction condition, there would be a training effect on discernment of gender-STEM stereotypes, but

$H_0$4b: no training effect on male-agentic stereotype endorsement

$H_0$5: Within the stereotype reduction condition, participants with "incorrect" prior beliefs (stereotyped beliefs) would not experience a greater decrease in discernment scores than those with "correct" pre-intervention beliefs

H6: Treatment effects on increase in discernment would not be explained by treatment-elicited empathic concern, perspective taking, or logical thinking

H7: There would be no gender differences in intervention effectiveness on stereotype discernment

H8: There would be a main effect of gender, such that women would report lower levels of sexism and male-agentic stereotype endorsement than men

H9: Within the stereotype reduction condition, women would report greater experience of discrimination in STEM

H10: Within the stereotype reduction condition, women would report less gender identity centrality than men

$H_0$11: Within the stereotype reduction condition, women would not demonstrate self-stereotyping, i.e., lower levels of self-perceived agency than men

H12: Within the stereotype reduction condition, participant gender and gender identity centrality would interact to predict change discernment, such that women higher in gender centrality would report greater gender-STEM stereotype discernment

H13: Within the stereotype reduction condition, participant gender and previous experience of discrimination in STEM would interact to predict greater gender-STEM

stereotype discernment, such that women with greater experience of discrimination would report higher gender-STEM stereotype discernment

Materials and Measures

All materials and measures are provided in Appendix C. Measures were administered on paper and completed by pen or pencil.

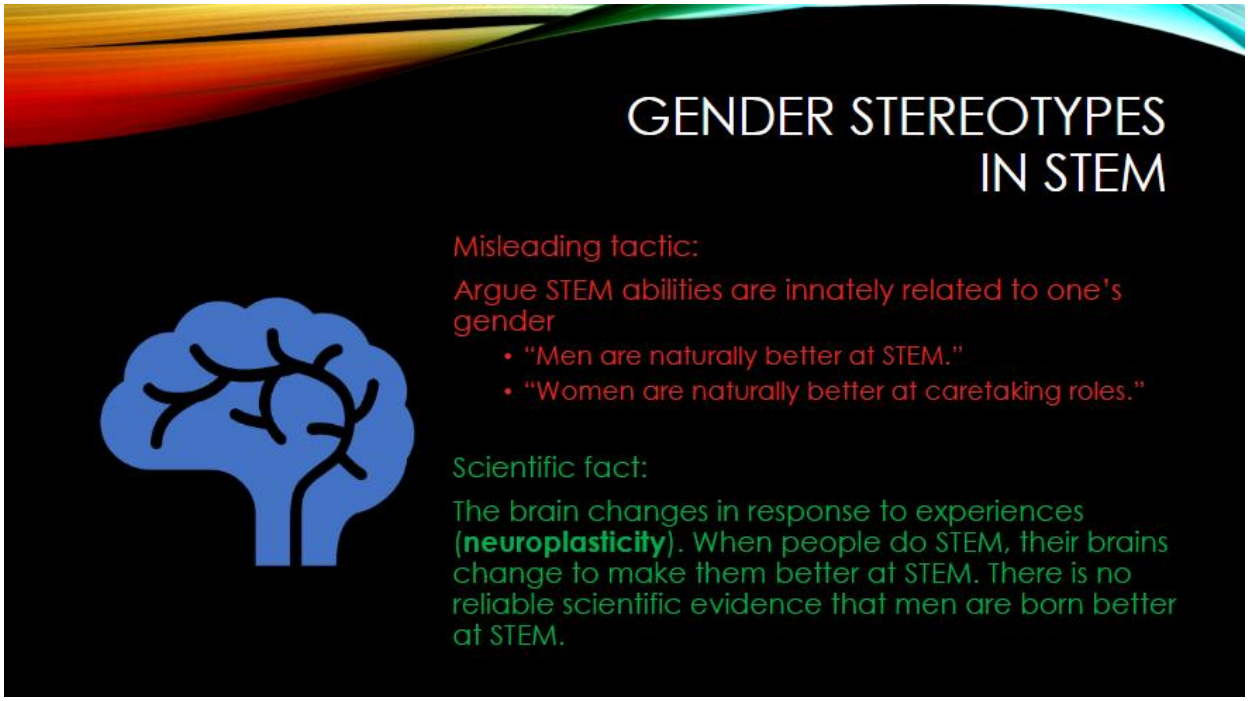
Intervention Presentation

PowerPoint Presentations were used for the purposes of this research. Slides for the stereotype-reduction training included the key messages from the treatment video and were presented according to optimised debunking and inoculation theory (see Figure 4.1). For the unconscious bias training, the University's Equity and Inclusion Office standardised materials were delivered (see Figure 4.1)

Figure 4.1

Screengrabs from the (a) Stereotype-Reduction Training presentation and the (b) Unconscious Bias Training presentation

(a)



(b)

Photo of screengrab removed for copyright reasons.

Measures

The following measures were repeated from Chapter 3: Credibility (of the training; α = .83), Psychological Reactance (α = 0.88), State Perspective Taking (α = 0.80), State Empathic Concern (α = .81), State Logical Thinking (α = .92), Perceived Increase in Knowledge (α = .93), Affirmative Action (α = 0.83), Hostile Sexism (α = 0.86), Male-Agentic Stereotype (α = .55), Modern Sexism (α = 0.81), Trait Perspective Taking (α = 0.75), Trait Empathic Concern, (α = 0.77) and Discernment of Gender Stereotypes in STEM (α = 0.60). Internal consistencies reported are from the current sample.

Discernment of Gender Stereotypes in STEM (DOGSS). The six 7-point Likert-items used in Chapter 2 and Chapter 3 were assessed. The index demonstrated poor internal consistency (α = 0.60), which did not improve by dropping any items. To test the internal validity, a CFA was run. Findings revealed a good model fit, χ^2 (9) = 14.89, p = .09 (CFI = .95, TLI = .91, RMSEA = .07, SRMR = 0.10). However, a principal factor analysis on the 6 items with oblique rotation (direct oblimin) demonstrated two factors with eigenvalues over Kaiser's criterion of 1 – the three stereotype endorsement items loaded onto the first factor, two of the knowledge items on the second factor, and the knowledge-related item assessing perceived scientific consensus did not load onto a factor at all. All KMO values were between .54 and .69, which are above the .5 cut off, but within the .5 - .7 range of barely acceptable to mediocre (Kaiser & Rice, 1974). The factors combined explained 56% of the variance. Therefore, the items were analysed individually for main effects and then combined into a composite score for further analyses. When analyses are performed at item-level, corrections for multiple comparisons are applied. The endorsement-worded items were reverse coded, such that a higher total score indicates greater discernment of gender stereotypes in STEM.

Male-Agentic Stereotype. As in Chapter 3, participants completed the masculine (agentic) subscale of the Personality Attributes Questionnaire (PAQ; Spence et al., 1973) on a scale from –3 (more true of women) to +3 (more true of men), with higher scores indicating more gender-stereotypic beliefs. The masculine (agentic) subscale consists of eight agentic traits: independent, active, competitive, can make decisions easily, never gives up easily, self-confident, feels superior, stands up well under pressure. The scale demonstrated poor internal consistency, α = 0.50 (65). In line with the findings from the online study, the scale's internal reliability improved by dropping "feels superior", though only marginally, to α = .55 (71).

This theoretically makes sense as "feels superior" may not be perceived as a *desirable* agentic trait, even though it is categorised as one by the scale creators. This likely reflects a societal-level shift in socially desirable traits in the last 50 years since the scale was created. Therefore, for both empirical and theoretical reasons, the item was removed from the composite score and subsequent analyses.

Gender Identity Centrality. The importance, or centrality, of one's gender identity to their self-concept was assessed using Luhtanen and Crocker's (1992) 'Importance to Identity' subscale. It is a face-valid 4-item Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree)*. Participants responded to statements such as, "My gender is an important reflection of who I am" and "My gender has very little to do with how I feel about myself". The scale demonstrated good internal consistency, $\alpha = 0.82$.

Analyses

The robust Welch's *t* was interpreted when data was normally distributed. When Shapiro-Wilk tests of normality indicated the sample violated the assumption of normality, Wilcoxon signed-rank tests were performed instead of paired-samples t-tests. For Wilcoxon signed-rank tests, the effect size calculated is rank biserial correlation. The interpretation of the effect size, rank biserial correlations, follows standard practice: 0.10 - 0.3 represents a small effect, 0.30 - 0.5 suggests a moderate effect, and >= 0.5 is considered a large effect. Bootstrapping, a common robust estimation method (Efron & Tibshirani, 1993), was applied to all analyses. Bootstrap parameter estimates help to ensure the model parameters are robust and provide 95% confidence intervals of the parameter.

Change in discernment of Gender Stereotypes in STEM was computed by subtracting the pre-intervention values from the post-intervention values. This was performed at both item-level and composite-level. When analyses are performed at item-level, corrections for multiple comparisons are applied. For the analyses on direction of psychological change, the outcome variable was dummy-coded to include participants who either reported a positive psychological change or a backfire effect. Six participants were excluded in these analyses as they reported zero psychological change. As the proportion of women (n = 17) to men (n = 64) in the stereotype-reduction training was representative of the population ratio (around 20% of the engineering population in North America is women), the variable was not weighted.

All analyses were conducted in SPSS (version 25, 26, and 27) and JASP (versions 0.12 and 0.16).

Results

First, I examine the interventions' perceived credibility and elicited psychological reactance to ensure there were no differences between the two interventions. Following this, I examine H1 by testing the comparative psychological mechanisms elicited in the interventions. Next, I present between-groups analyses to compare the main effects of the unconscious bias training and the stereotype-reduction training on the main outcome variable of gender-STEM stereotype discernment (H2), as well as the differential generalised effect on more general measures of gender stereotyping ($H_0$3a and $H_0$3b). Then, within-group analyses are presented on the stereotype reduction training to examine the intervention's main effects on psychological change in gender-STEM stereotype discernment (H4a) and male-agentic stereotype endorsement ($H_0$4b). Additionally, to gain a deeper understanding, the backfire effect (H5), the mediating role of psychological mechanisms (H6), and the moderating role of participant gender (H7) were examined. Next, gender differences in gender stereotype literacy are presented (H8); and, finally, analyses on gender differences in experience of discrimination in STEM (H9), gender identity centrality (H10), and self-perceived agency ($H_0$11) are presented, as well as their moderating effects on the intervention (H12 – H13).

Training Checks

Both trainings elicited low levels of psychological reactance and high ratings of credibility (see Table 4.2). Next, I examined the psychological processes elicited in the two interventions. Contrary to H1, the unconscious bias training elicited significantly higher levels of empathic concern, perspective taking, and logical thinking than the stereotype-reduction training (all *p*s < .001). Additionally, the unconscious bias training group reported a significantly higher perceived increase in knowledge than the stereotype-reduction group ($p < .001$). This could be interpreted as the more traditional prejudice intervention (unconscious bias training) elicited higher levels of the traditional stereotype-reduction psychological mechanisms.

Between-Groups Analyses

Comparisons between the two interventions on the main outcome variable, discernment of gender stereotypes in STEM, were run first (H2). Next, given the unexpected findings, exploratory mediational analyses were run examining the relative roles of the elicited psychological mechanisms in explaining discernment scores. Lastly, the interventions' differential generalised effects (or 'umbrella effect' or 'blanket of protection') were examined by testing group differences on male-agentic stereotype endorsement and sexism.

Discernment of Gender Stereotypes in STEM

Contrary to H2, discernment of gender stereotypes in STEM reached significance with the unconscious bias training group *(M* = 4.99, *SD*= 0.89) demonstrating higher overall discernment of gender-STEM stereotypes than the stereotype-reduction group *(M* = 4.67, *SD*= 0.99), $t(138.52) = 2.02$, $p = .045$, $d = .336$. Specifically, the unconscious bias training group demonstrated greater discernment on the three negatively-worded (endorsement) items: "On average, men's and women's brains function differently"; "More females are not in STEM because they are naturally interested in other things"; and "Men are innately better at STEM", but when applying a more stringent alpha level for multiple comparisons, only "More females are not in STEM because they are naturally interested in other things" remained significant (see Table 4.2 for means).

Alternatively, the stereotype-reduction training group demonstrated greater discernment on "Some people use misleading tactics to claim there are differences between men's and women's STEM ability" than the unconscious bias training group. Lastly, there were no group differences on ability to discern "The scientific consensus is that there are no inborn gender differences in STEM ability".

Table 4.2

Descriptive Statistics for Measures by Intervention Condition

	Unconscious bias training		Stereotype reduction training	
	M	*SD*	*M*	*SD*
Training credibility	5.72	1.02	5.53	1.24
Psychological reactance to training	1.96	1.49	1.77	1.20
Perceived knowledge increase ***	5.53	1.17	4.43	1.51
Male-agentic stereotype	0.29	0.51	0.30	0.49
Male-logical stereotype	0.15	0.68	0.34	0.85
DOGSS items				
Brains function differently	3.45	1.91	4.32	1.67
Scientific consensus	4.87	1.84	4.92	1.63
Men innately better	2.10	1.31	2.68	1.47
Not naturally interested*	2.93	1.48	3.98	1.78
Misleading tactics *	4.54	1.51	5.35	1.81
Hiring/promoting effects	4.80	1.81	4.82	1.80
Affirmative action *	4.87	1.28	4.29	1.38
Modern sexism	3.20	1.16	2.97	1.10
Hostile sexism	3.14	1.05	3.20	1.24
State logical thinking ***	5.64	1.23	4.79	1.33
State empathic concern ***	5.41	1.34	4.63	1.45
State perspective taking ***	5.00	1.56	3.91	1.67
Trait empathic concern	3.84	0.59	3.67	0.74
Trait perspective taking	3.81	0.67	3.89	0.65
Trait logical thinking	1.95	1.07	1.94	1.01

Note. * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$. Adjusted significance level of $\alpha = .008$ used for DOGSS items to account for multiple comparisons. Mann-Whitney U tests were performed and reported for all but "trait empathic concern" due to non-normally distributed data. Welch's t-tests are reported for normally distributed data.

Exploratory Analyses. To examine if the differential levels of psychological mechanisms elicited may have influenced the intervention differences in discernment, mediation analyses were run. As the following analyses were not part of the hypotheses or analytical plan, these are exploratory and post-hoc; all related tables and figures are presented in Appendix C.

Treatment-Elicited States. The unconscious bias training group experienced significantly greater levels of state empathic concern, state perspective taking, and state logical thinking (see Table 4.2). Thus, to examine the relative roles of these variables in

explaining the significant relationship between intervention condition and discernment of gender-STEM stereotypes, indirect effects of intervention condition on discernment of gender-STEM stereotypes through state empathic concern, state perspective taking, and state logical thinking were tested. The results trended towards a significant indirect effect through empathic concern. While the unconscious bias training evoked higher levels of all three cognitive and affective states, only state empathic concern was associated with scores on DOGSS. However, when the model was re-run controlling for the related trait variables, state empathic concern no longer held significance *(p* = .054). Both models replicate findings from the previous chapter, such that trait-level variables had more explanatory power than treatment-elicited states on discernment.

Perceived Increase in Knowledge. Perceived increase in knowledge, on the other hand, completely mediated the relationship between intervention and discernment of gender-STEM stereotypes *(p* < .001). This is supported by other persuasion and gender bias research, which shows that people are more easily convinced of a message when they perceive an increase in knowledge on the topic (Tormala & Petty, 2007; Pietri et al., 2017). This finding suggests perceived increase in knowledge may be a more relevant process in stereotype discernment than state empathic concern, perspective taking, or logical thinking and, thus, may be particularly relevant for future intervention research design.

Generalised Effects

With the exception of support for affirmative action (*t(*133.67) = 2.61, *p* = .010, *d* = .442), the two intervention groups did not differ on any other outcome measure (hostile sexism, modern sexism, male-agentic stereotype; all *ps*> .130), supporting H3a and partially supporting H3b. The unconscious bias training group demonstrated greater support for affirmative action. As we did not measure pre-intervention attitudes, and the groups were demographically different, we cannot be confident this is an effect of the intervention versus a random group difference in pre-existing attitudes. However, it is plausible it is an intervention effect as the unconscious bias training addressed the effects of unconscious bias on women at work; specifically, the scenario of assuming women have, or will have, children and will choose to be the primary caregiver, which will negatively impact their job performance.

Within-Groups Analyses

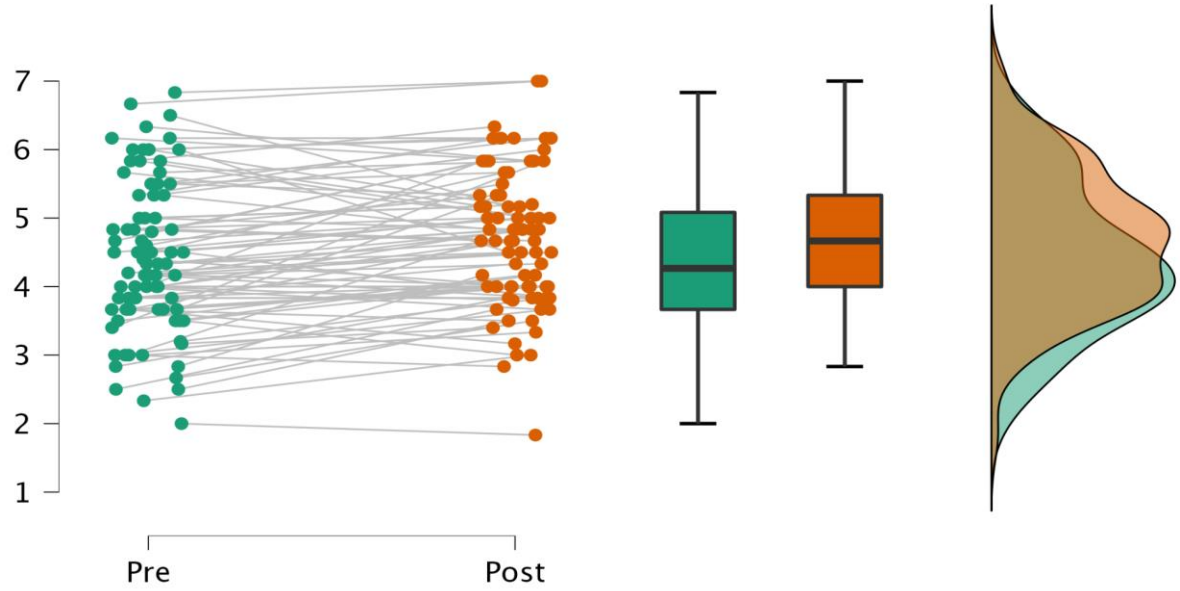
The following pre-post intervention analyses were conducted on the stereotype-reduction training participants only, as the unconscious bias training group did not complete pre-intervention outcome measures.

Change in Discernment of Gender Stereotypes in STEM

As expected (H4a), there was an increase in discernment of gender stereotypes in STEM after the intervention, t (82) = 3.935, $p < .001$, $d = .432$ (see Figure 4.2). Further pre-post comparisons demonstrated a significant change in discernment on three items (see Figure 4.3): "Some people use misleading tactics to falsely claim that men are better at STEM than women" significantly increased *(*$W = 610.5$, $z = 1.99$, $p = .038$, $r = .352$; see Figure 4.3); "The scientific consensus is that there are no innate gender differences in STEM ability" significantly increased *(*$W = 697$, $z = 3.45$, $p < .001$, $r = .619$; see Figure 4.3); meanwhile, "On average, men's and women's brains function differently" significantly decreased *(*$W = 230$, $z = -2.60$, $p = .008$, $r = -.466$; see Figure 4.2). The rank biserial correlations suggest that changes in stereotype literacy and endorsement, measured by the three items above, were moderate in size. However, pre-post differences on the other three items, "Men are innately better at STEM", "More females are not in STEM because they are naturally interested in other things", and "Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM", were not significant, *ps*>.123.

Figure 4.2

Raincloud plot showing individual pre-post comparisons on discernment of Gender Stereotypes in STEM



Note. The raincloud difference plot visualises data density for pre- and post- scores with a box plot and a one-sided violin plot. Within the box plots, the bold black line shows the sample median, the hinges indicate the 25th and 75th quantile, and the whiskers point to 1.5 interquartile ranges beyond the hinges. Range on y-axis represents range on response scale.

Table 4.3

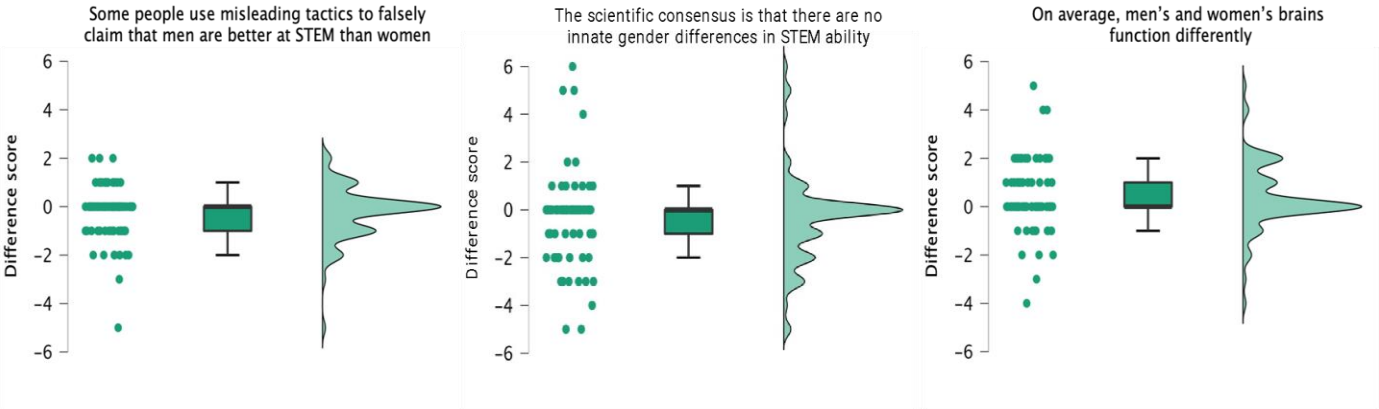
Pre- and Post-Intervention Descriptive Statistics for Discernment of Gender-STEM Stereotypes (N = 84)

	Pre		Post	
	M	*SD*	*M*	*SD*
Men are naturally better at STEM [R]	2.89	1.57	2.68	1.47
Some people use misleading tactics to falsely claim that men are better at STEM than women *	5.04	1.41	5.35	1.27
On average, men's and women's brains function differently [R] **	4.76	1.68	4.32	1.66
The scientific consensus is that there are no innate gende differences in STEM ability ***	4.52	1.63	5.15	1.48
More females are not in STEM because they are naturally interested in other things [R]	4.08	1.84	3.98	1.78
Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM	4.64	1.87	4.82	1.80
Composite score * **	**4.39**	**1.10**	**4.67**	**0.99**

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Wilcoxon signed-rank comparisons. Responses collected on a Likert-type scale from 1 to 7. [R] represents reverse-scored items in the composite score.

Figure 4.3

Raincloud Difference Plots showing Significant Within-group Comparisons for Discernment Difference Scores



Note. Difference scores are calculated post - pre.

Backfire Effect. Consistent with $H_0$5, a chi-square test revealed that a decrease in stereotyped beliefs did not occur more among those who held stereotyped pre-intervention beliefs than those who held the correct beliefs ($\chi^2(1) = 1.95$, $p = .163$). Thus, the findings support previous literature which finds that backlash is not a concern in interventions, even when the topic touches on a salient stereotype. Conversely, this finding is inconsistent with diversity training initiatives which tend to find a backfire effect. I believe this may be explained by differences in outcome measures. While we measure discernment in this intervention, diversity initiatives tend to measure *attitudinal* or *behavioural* outcomes (support for affirmative action, hiring decisions, etc.). Further, it is worth noting that, in total, 35% of the sample reported a decrease in discernment. However, these pre-post differences are derived from average scores, so this difference may mean selecting just one point lower on just one of the six Likert-type items at post-test, which may not be a very meaningful decrease to interpret.

Because participants were categorised into backlash versus positive increase for the chi-square test, the *degree* to which participants demonstrated an increase or reduction in stereotype endorsement was not taken into account in these analyses. For example, a participant who reported a change of -.17 and a participant who reported a change of -1.33 were categorically the same. By categorising continuous data, the loss of statistical variation can mean a loss of a more nuanced understanding. However, distinguishing participants

above versus below the meaningful threshold of zero was the aim of the analysis, in order to identify if there were differences in direction of change.

Following this, we then tested the *degree* of change among those who held correct pre-intervention beliefs and those who held stereotyped beliefs. There was a greater psychological shift in the "right" direction among those who held incorrect beliefs pre-intervention *(M* = 0.51, *SD*= 0.57) than those who entered the training with correct beliefs *(M* = 0.11, *SD*= 0.57), *t(*56.39) = 2.94, *p*= .004, *d* = .700 (see Figure 4.3). This does not appear to be explained by ceiling effects in the correct-belief group, as the pre-intervention average among those with the correct beliefs was *M* = 5.13 (*SD*= 0.77) on the 7-point scale. However, the degree of positive increase in discernment in either group was relatively small (half of a scale point and one tenth of a scale point, respectively).

Further, this does not tell us whether the increase in discernment among those with the incorrect belief remained in the myth zone. Indeed, of those who held the stereotype belief pre-intervention and experienced an increased ability to discern, 66% remained in the incorrect-belief camp (see Table 4.4). Again, these values are split by participants who fall above and below 4 (the midpoint on the scale), not a median split, as the meaning of the mid-point on the scale is more relevant for these research questions than dividing the sample by low and high. As a robustness check, these analyses were re-run using a median split (Median = 4.26) for pre-intervention discernment, and results did not meaningfully change.

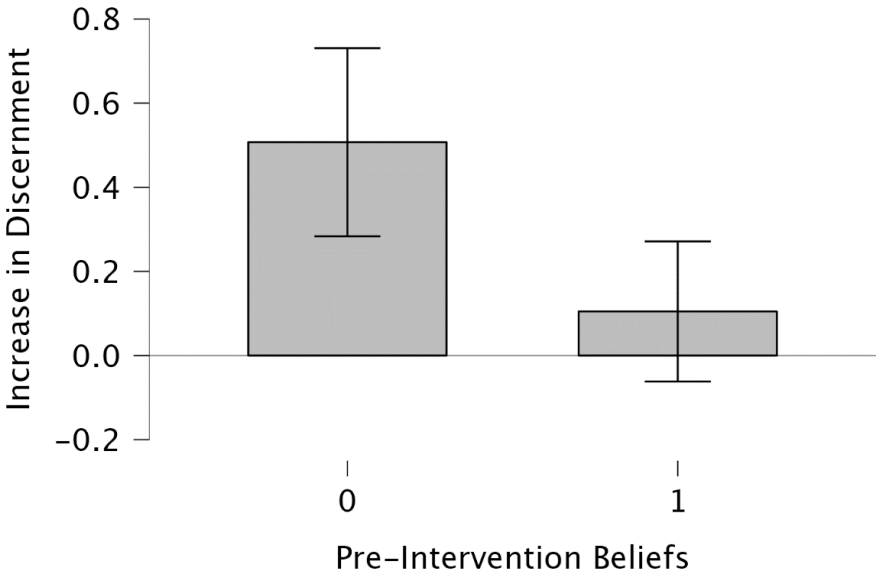
Table 4.4

Contingency Table for Pre-Post Intervention Discernment

	Pre-intervention Discernment		
Post-intervention Discernment	Incorrect belief	Correct belief	Total
Incorrect belief	16	2	18
Correct belief	8	44	52
Total	24	46	70

Figure 4.3

Average Increase in Discernment by "Correct" versus "Incorrect" Pre-Intervention Beliefs



Note. Pre-intervention discernment is dummy coded 0 = "incorrect" (less than 4; *disagree* side of the scale) and 1 = "correct" (greater than 4; *agree* side of the scale). Error bars represent standard error. Increase in discernment is calculated as a post – pre difference score.

Taken together, while true therapeutic versus prophylactic inoculation comparisons cannot be examined due to no counter message attack following the treatment, the findings indicate a persuasive effect of the intervention, particularly among those who held contrary prior beliefs. Although there wasn't as much of a psychological increase among those who held correct beliefs pre-intervention, the intervention may have had a *protective,* i.e., inoculative, effect on discernment among those who already held the correct beliefs. Again, this could not be tested due to no exposure to a counterattack message. Therefore, while the intervention had a successful debunking effect, it is unclear whether it had a successful inoculative effect.

Psychological Mechanisms. To test $H_0$6, the relative influences of state empathic concern, state perspective taking, and state logical thinking in predicting psychological change in gender-STEM stereotype discernment were tested with a linear regression. There were no significant effects for any psychological mechanism in the model *(p* = .294). Replicating findings from Chapter 3, the model controlling for gender and trait variables demonstrated a significant effect of trait empathy *(p* = .012). Thus, those who experienced

greater psychological shifts in discernment were not differentiated from those who didn't by these treatment-elicited affective and cognitive mechanisms, but rather trait empathic concern.

Gender Differences. As predicted by H7, there were no gender differences in direction of psychological change in discernment, meaning there were no gender differences in participants who experienced a backfire effect or increased discernment of gender stereotypes in STEM, $\chi 2$ (1, N = 74) = .02, p=.88. Likewise, there were no gender differences in overall magnitude of psychological change, $t(24.72) = -0.22$, $p = .83$, $d = -.06$). As a robustness check, items were examined independently (see Table 4.6). Levene's tests for equality of variances demonstrated homogeneity of variance (all ps > .12); however, Shapiro-Wilk tests indicated non-normal distributions across all items (all ps < .05). Therefore, Mann-Whitney U tests, or Wilcoxon rank-sum tests, were conducted. One of the six items demonstrated a gender difference in psychological change: women reported a greater increase in agreement with "Some people use misleading tactics to falsely claim that men are better at STEM than women" (U = 339, $p = .01$, $r = .36$; adjusted $p = .06$.

Table 4.6

Descriptive Statistics for Psychological Change in Discernment of Gender-STEM Stereotypes by Gender

Items	Women n = 17		Men n = 64	
	M	*SD*	*M*	*SD*
Men are naturally better at STEM [R]	0.06	0.90	0.20	1.31
Some people use misleading tactics to falsely claim that men are better at STEM than women	0.88	1.36	0.08	0.95
On average, men's and women's brains function differently [R]	0.35	1.06	0.40	2.17
The scientific consensus is that there are no innate gender differences in STEM ability	0.00	2.17	0.47	1.86
More females are not in STEM because they are naturally interested in other things [R]	0.29	1.21	0.13	1.21
Inaccurate beliefs about women's ability in STEM can have effects on who is hired and promoted in STEM	0.06	1.06	0.13	1.28

Note. Means (M) and standard deviations (SD) for women's and men's difference score (post – pre). [R] represents reverse-scored items. Positive values represent a positive psychological change, i.e., an increase in stereotype literacy and stereotype reduction.

Change in Male-Agentic Stereotype Endorsement

As expected (H4b), change in endorsement of the male-agentic stereotype pre-post intervention was not significant *(p* = .378), suggesting no change in stereotypic gender-trait beliefs (see Table 4.5). Replicating the other chapters, the sample's overall low endorsement pre-intervention *(M* = 0.29, *SD*= 0.52; *M* = 0 represents no perceived gender differences) likely created floor effects. These corroborated findings across chapters suggest that the male-agentic stereotype may not be directing the underrepresentation of women in STEM - at least *explicitly* – and/or is perhaps no longer a gender stereotype at all.

Table 4.5

Pre- and Post-Intervention Descriptive Statistics for Male-Agentic Stereotype (N = 83)

Items	Pre		Post	
	M	*SD*	*M*	*SD*
PAQ				
Independent	0.04	1.09	0.08	0.91
Active	0.31	0.82	0.27	0.73
Competitive	0.58	1.15	0.52	0.93
Can make decisions easily	0.35	1.05	0.22	0.76
Never gives up easily	-0.01	0.57	-0.01	0.47
Self-confident	0.29	0.92	0.22	0.80
Stands up well under pressure	0.28	1.05	0.22	0.84
Composite score	0.25	0.52	0.29	0.52

Note. All Wilcoxon signed-rank comparisons were non-significant, all *p*s > .28. Responses measured on a scale from -3 'more true of women' to +3 'more true of men'

Main Effects and Interactive Effects of Gender

Due to the small number of women, the following analyses are preliminary and exploratory in nature.

Unlike previous findings (Chapter 3), and contrary to H8, women did not differ from men on hostile sexism *(p* = .127) or attitudes towards affirmative action *(p* = .133), only modern sexism (*p* < .001). This suggests female engineers are more sexist than females in the general population and/or male engineers are less sexist than general-population males. Similarly, contrary to Chapter 3 findings, men and women did not differ on male-agentic stereotype endorsement (*p* = .399). Descriptive statistics by gender are presented in Table 4.7.

As expected (H9), women reported greater experience of discrimination in STEM *(p* < .001) and, unexpectedly (H10), women reported greater gender identity centrality *(p* = .002; see Table 4.7). Therefore, the female engineers did not demonstrate self-group distancing as a protective mechanism against negative stereotypes targeted at their gender group (i.e., gender identity centrality was not significantly lower among women than men). Instead, this may be explained by the increased saliency of gender for women in STEM and a subconscious effort to protect against gender stereotypes in STEM. For example, if a woman's gender identity is important and strong, she may be more resilient to the negative effects than a woman whose gender identity is weaker and, therefore, more easily targeted. Conversely, as expected ($H_0$11), findings demonstrate that the female engineers didn't engage in self-stereotyping *(p* = .556; no gender differences in self-reported agentic traits).

Contrary to H12, gender identity centrality did not interact with gender to predict pre-intervention discernment of gender-STEM stereotypes *(p* = .246) or change in discernment *(p* =.10). Additionally, contrary to H13, previous experience of discrimination in STEM and gender did not interact to predict pre-intervention discernment *(p* = .385) or change in discernment *(p* = .296). Therefore, women's greater identity centrality and experience of discrimination in STEM did not influence their discernment of gender stereotypes.

Table 4.7

Descriptive Statistics by Gender

	Women n = 17		Men n = 64	
	M	*SD*	*M*	*SD*
Credibility of training	5.95	0.96	5.51	1.18
Trait logical thinking	3.77	1.07	3.79	1.16
Trait perspective taking	3.90	0.67	3.90	0.66
Trait empathic concern	3.96	0.71	3.66	0.69
Gender identity centrality ** *(d* = .96)	5.04	1.33	3.72	1.40
Agentic traits [a]	0.68	1.16	0.69	0.82
Increase in gender-STEM stereotype discernment	0.29	0.64	0.25	0.62
Male-agentic stereotype endorsement	0.13	0.44	0.24	0.43
Modern sexism *** *(d* = 1.15)	2.13	0.65	3.15	1.07
Attitude towards affirmative action	4.82	1.21	4.21	1.36
Hostile sexism	2.76	1.04	3.25	1.19
Have experienced discrimination in STEM*** *(d* = 2.00)	5.13	1.36	2.23	1.52
University's engineering department is diverse ** *(d* = .71)	2.88	1.36	4.00	1.67
Felt sense of belonging in STEM * *(d* = .64)	4.82	1.13	5.53	1.06

Note. * $p < .05$, ** $p < .01$, p *** $< .001$. d = Cohen's d effect size. Welch's t-tests conducted. All responses measured on a Likert-type scale from 1 to 7 unless otherwise noted.
[a] Responses on a Likert-type scale from -3 (not at all like me) to +3 (very much like me)

Discussion

This study compared two psychological interventions, stereotype reduction and unconscious bias, on discernment of gender stereotypes in STEM among engineering students. Further, it examined the effects of the stereotype-reduction training on psychological *change*. This study builds on previous chapters by examining gender-STEM stereotype discernment interventions *within* STEM and by measuring within-subject stereotype change.

Overall, findings suggest the superiority of the unconscious bias training to the stereotype-reduction training. Specifically, the unconscious bias training participants demonstrated higher levels of discernment overall, but the stereotype reduction intervention reported higher discernment of misleading tactics used to perpetuate gender-STEM stereotypes. There was a small but significant increase in discernment among the stereotype-reduction participants, with greater shifts among those who entered the intervention with "incorrect" (stereotyped) beliefs than those who held "correct" ones. These findings are

consistent with an international workplace gender-stereotype intervention, which found the positive effect on gender attitudes was driven by the subgroups who demonstrated less supportive attitudes towards women pre-intervention (Chang et al., 2019). However, the increase in discernment among those with incorrect pre-intervention beliefs still remained within the myth zone for most. This is perhaps not surprising given that beliefs and attitudes intrinsically tied to one's self-concept, like gender, are associated with reluctance to change (Johnson, 1989; Eagly & Chaiken, 1993; Sherman, 2002).

The unconscious bias training group reported significantly higher levels of perceived increase in knowledge. Persuasion research suggests that people are more easily convinced of a message when they perceive an increase in knowledge on the topic (Tormala & Petty, 2007). Indeed, perceived increase in knowledge fully mediated the relationship between intervention and stereotype discernment scores. Further, the greater amount of group-based discussions present in the unconscious bias training - and not in the stereotype-reduction training – may have influenced this perceived increase in knowledge, given that elaboration is positively associated with the process of psychological change (Petty & Cacioppo, 1986) as well as processes associated with resistance to counter information (Clear et al., 2021). Future intervention research may want to consider prioritising a discussion component particularly given the additional research which suggests the positive impact of post-inoculation talk (e.g., Compton & Pfau, 2009; Ivanov et al., 2015)

Treatment-elicited empathic concern, perspective taking, and logical thinking did not predict psychological change; however, trait empathic concern did. Therefore, it is possible that the intervention material may elicit either a more mentalising component of empathy or a more affective component of empathy, depending on individual differences in propensity for each of these states. *Which* of these components is evoked may not be as important as *whether* they are evoked. While some may have a tendency to engage in affective empathy, others may have a tendency to engage in cognitive empathy, and yet others may naturally engage in both. Thus, the specific pathway taken to achieve an empathic experience is perhaps less important and more indicative of dispositional empathy than intervention components or effectiveness.

As previously discussed, social mechanisms are powerful in shaping stereotype endorsement (Duguid & Thomas-Hunt, 2015; Crandall & Stangor, 2005; Stangor et al., 2001; Crandall, Eshleman, & O'Brien, 2002; Brief, Dietz, Cohen, Pugh, & Vaslow, 2000; Paynter et al., 2019; Hornsey & Fielding, 2017) and are considered important in inoculation research too (post-inoculation talk; Compton & Pfau, 2009; Ivanov et al., 2015). When individuals are

faced with a persuasive message that is incongruent with their current beliefs or attitudes, they will actively recall others in their social groups (family, friends, coworkers) who share their existing belief or attitude (Zuwerink, 2003). This social validation process helps them protect their belief or attitude. However, little research has explored the role of social mechanisms in sustaining or deteriorating intervention effects.

Dasgupta's (2011) stereotype inoculation model presents a theoretical framework for inoculating underrepresented students against stereotype threat. In this application of the inoculation metaphor, ingroup peers and experts act as social vaccines by protecting one's self concept. In a social-belonging intervention, the minority achievement gap among college students halved over 3 years (Walton & Cohen, 2011); and there is considerable support for the model and the benefit of same-gender experts and peers on women's sense of belonging and self-efficacy (Dennehy et al., 2018; Dasgupta, 2011; Stout et al., 2011; Sekaquaptewa, 2011; Hoyt & Murphy, 2016; Dasgupta et al., 2015). Therefore, social factors may be critical for reducing both stereotyping and experiencing stereotype threat.

Thus, it may be fruitful to investigate the role of counterarguing (cognitive), threat (affective), psychological reactance (cognitive-affective), and post-talk inoculation (social) in the longevity of inoculation effects on stereotype discernment. As with most psychological interventions, if not all, inoculation is found to have decay effects. While 'booster shots' (follow-up treatments) are recommended to buffer decay effects, it would be useful to investigate the importance of these psychological mechanisms both in the main intervention and subsequent boosters for sustained psychological change. Thus, it would be worthwhile for future research to conduct follow-ups to examine the long-term influence of cognitive, affective, and social mechanisms on stereotype change.

The training was not voluntary, which may help to explain some of the findings. It is recommended that diversity trainings are voluntary to improve intervention effectiveness (Dobin et al., 2015). This is not surprising, as it is likely that individuals who voluntarily participate are also individuals who are open to and/or interested in the topic, which is fertile ground for psychological change. However, it also poses the issue of self-selection bias. Are we reaching enough of the population, and particularly those who are most influenced by stereotypes, if these psychological interventions are voluntary? Conversely, simply including the presence of participants who have the most psychological room to shift does not inherently necessitate that they will engage in that psychological change. Indeed, a concern is that if trainings are made mandatory, we run the risk of eliciting psychological reactance and the backfire effect. However, findings from the present study and previous research (see

Wood & Porter, 2018 for meta-analysis) suggests the backfire effect, even on polarised issues, is negligible.

However, the backfire effect has been found in real-world managerial diversity training (Kalev et al., 2006) and affirmative action programmes (Bond & Pyle, 1988; Linnehan & Konrad, 1999), suggesting more research is needed to elucidate these mixed findings. Perhaps, in research contexts, participants experience less backfire effect because they are aware they are participants in a contained study that will likely not have real-life consequences for them; alternatively, employees may experience greater psychological threat to real-life interventions with real-life consequences. Again, this may point to the value of investigating and prioritising psychological safety.

Limitations

Due to the nature of field research, some of the research plans did not go as planned. Firstly, the unconscious bias training group did not provide pre-training measures and, therefore, we were not able to examine within-group psychological change or between-group differences in psychological change – only differences in post-intervention measures. Secondly, the study did not include a control group. Without a control group, we cannot robustly test the manipulations and associated psychological mechanisms. The study would be improved by the inclusion of a control condition, in which participants receive a training on a topic unrelated to gender-STEM stereotypes, in the same form of an in-person presentation and supporting PowerPoint slides, to match the treatment conditions, and for the same length of time as the treatment trainings.

Additionally, demographic data was only completed by less than a third of the unconscious bias training participants. Based on these responses, the unconscious bias training group was significantly older, more liberal, gender diverse, and ethnically diverse. Further, they were hired as undergraduate teaching assistants, and the training was delivered as part of the training for this position. Thus, the sample may suffer from selection bias based on the type of individual who would both (1) apply to be a teaching assistant and (2) also be *selected* and hired for the role. Alternatively, the treatment group received the training during class time and was more representative of a general student engineering population. Altogether, we cannot be confident the between-groups findings are explained by intervention effects.

Further, although not captured by the quantitative data, participants in the stereotype discernment training, but not in the unconscious bias training, explicitly expressed gender stereotypes attitudes during the training. This, together with the demographic differences between the groups reported above, suggests the two groups may not be sampled from the same population. Therefore, as we did not measure pre-intervention beliefs in the unconscious bias training, and the groups are arguably qualitatively different, we cannot be confident these differences are an effect of the intervention.

As well as the inclusion of a control group and pre-intervention measures across conditions, the study design would have improved with the inclusion of post-intervention misinformation exposure as well as follow-up tests one week and one month after the intervention to test the robustness long-term and decay effects. Indeed, stereotype-reduction research shows that the positive effects typically wear off after a day (Paluck & Green, 2009). Moreover, it would improve study design to give more time between pre-test measures and the intervention to reduce testing effects in the repeated-measures design.

In summary, the two intervention groups appear to be sampled from different student populations and there was no control group, misinformation 'attack', or post-intervention follow-ups. Additionally, the responses for trait variables (treated as control variables) were collected post-intervention, which may have resulted in post-treatment conditioning. In other words, it's possible that these variables were affected by the intervention and, therefore, not accurate measures of confounding variables. Considering these methodological limitations, the findings are considered largely exploratory and observational.

Because the trainings were delivered by a woman, it may have decreased the effect of the intervention. Female professors are rated less positively and perceived as more sexist than male professors when reading identical information on gender gaps, especially among those with more gender stereotypical attitudes (Abel & Meltzer, 2007). In other words, a male researcher delivering the training may have been more effective in eliciting psychological change, as men are rated more positively when discussing gender bias issues (Abel & Meltzer, 2007). This would be an interesting manipulation to test in future intervention research.

Although there was a small 'generalised' effect of the unconscious bias training (as measured by support for affirmative action), it could also be considered a more effect of the intervention, given that an example discussed was on the topic of female-family unconscious bias. Indeed, there were a number of small differences between the interventions that could

be confounding variables in the between-groups analyses. As such, these are strongly recommended to be interpreted as associative in nature.

Practically, it would be advantageous if general interventions were just as effective as topic-specific ones, as the same intervention could be systematically employed across different targeted domains - and potentially improve untargeted beliefs/attitudes as well through 'blanket of protection'. Future research would benefit from testing this generalised effect of specific versus general interventions across multiple categories of stereotypes. Theoretically, general messages provide a generalised cognitive framework that is not context-dependent and could therefore be applied to any stereotyped domain; thus, it is likely that this approach requires greater cognitive effort and elaboration, which appear to be mechanisms that improve the success of inoculation (Clear et al., 2021; Cárdaba et al., 2014; Haugtvedt & Petty, 1992). However, this would also require greater motivation.

Inoculation interventions with specific messages and refutations may be more useful for individuals who do not engage in elaborative, higher-thinking processes. Further, as research shows individuals may use both high-level thinking (e.g., elaboration) and low-level thinking (e.g., heuristics) in successful inoculation interventions (Clear et al., 2021), future research examining interventions with both high-level and low-level cognitive components - and analysing their relative and/or interactive effects - is advised.

The lack of interaction between gender centrality and gender on discernment scores may be due to the self-report nature of them. Some research shows that women's gender identification is associated with their behavioural gender bias, but not self-reported gender bias (Kaiser & Spalding, 2015). This may be explained by implicit stereotypes operating at a subconscious level (Banaji & Greenwald, 2013). Alternatively, it may be explained by social desirability bias, such that women may be aware of their preference, but may be reluctant to acknowledge it in a self-report measure. Methodologically, engineering centrality was asked at the beginning of the questionnaire, and this may have primed individuals. However, given the explicit content of the training, it is likely that gender identities would have been primed by the training itself, which could have subsequently influenced post-intervention ratings on importance to identity; hence why these identity-centrality questions were asked prior to the training material.

Finally, although the training was informed by inoculation theory, this field experiment did not empirically test inoculation theory there was no follow-up counterattack message after the training. Overall, the study requires replication using larger samples, a more matched and representative comparison treatment group, inclusion of a control group,

inclusion of a post-treatment misinformation ‘attack’, and multiple post-intervention follow-ups. Thus, future research on the effects of inoculation interventions as gender stereotype *and* stereotype threat interventions in STEM is recommended.

CHAPTER FIVE – IMPROVING UNDERSTANDING AND EVALUATION OF GENDER-TRAIT STEREOTYPES

In this final empirical chapter, gender-trait stereotypes were explored. Given the poor internal consistency and poor performance of the male-agentic stereotype across all previous studies (Chapters 2, 3, and 4), conceptualisation and measurement of the male-agentic stereotype was investigated. The female-communal stereotype was also examined.

Introduction

A methodological concern related to studying agency is the diversity in conceptualisation and operationalisation of the term. Firstly, there is a lack of consistent measurement in regards to the items/scales. The majority of researchers do not use established or validated scales, and different items are selected by different researchers to assess agency across studies (see Figure 5.1). Therefore, researchers may be measuring different constructs but calling them all agency. This may help to explain the mixed findings in the literature. Secondly, the few well-established measures either have not been re-validated or do not stand up well to re-validation, likely because they have not been updated since creation 5 decades ago. Social roles and associated stereotypes have arguably undergone significant change in that time. Thirdly, the labelling of agentic dimensions varies across studies, making it challenging once again to make valid comparisons. Further, these dimensions are not well-identified or differentiated by the items included. Indeed, different researchers assign the same items (e.g., assertive) to different factors (e.g., competency, agency, assertiveness) and quite different items (e.g., boastful, calm in emergencies) to the same factor label (e.g., agency), see Figure 5.1. Lastly, the relationship/distinction between agency and competency is not well understood and is defined and measured inconsistently across studies (see Figure 5.1). To the best of my knowledge, these methodological issues have not yet been addressed or discussed in past literature.

Agency and Competency

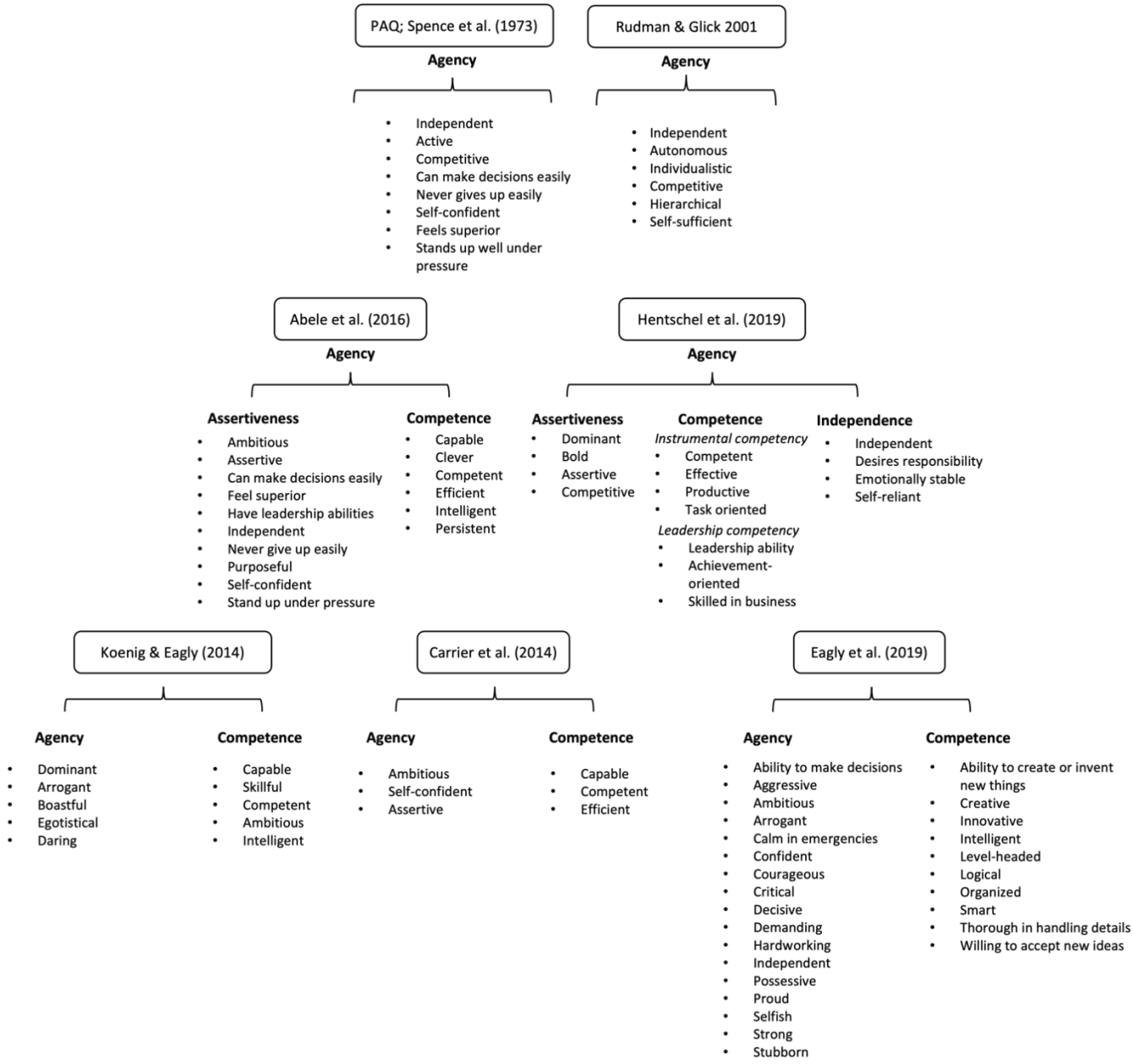
Constructs

Some researchers (e.g., Hentschel et al., 2019; Abele et al., 2016; Rudman & Glick, 2001; Rosette et al., 2016) include competency as a dimension of agency. Competent-agency

involves performance related to skills, tasks, and functioning (Abele & Wojciszke, 2014); whereas dominant-agency involves domineering, competitive, and controlling behaviour (Carrier et al., 2014), Rudman et al., 2012). In other words, competence is perceived as instrumental and represents utility and accomplishment, whereas dominance is perceived in relation to others and represents control and advancement over others.

Figure 5.1

Schemata Comparing Agency Measures



Note. The top row of indices represents measures that do not distinguish between agency and competence; the middle row of indices measure competence as a dimension of agency; the bottom row of indices measure agency and competence as distinct constructs.

Other researchers distinguish competency and agency as completely distinct constructs (e.g., Carrier et al., 2014; Eagly et al., 2019; Schaumberg & Flynn, 2017). Meanwhile, other researchers do not acknowledge a distinction at all, such that they collapse the two by including competency-related items under an umbrella factor of 'agency' (e.g.,

Spence et al., 1973; Bem, 1981a), or refer to agency and competency interchangeably (e.g., Fiske et al., 2002; Cuddy et al., 2008; Oldmeadow & Fiske, 2007).

Given the interchangeable use of the terms agency and competence in stereotype research, there appears to be general agreement that these pairs of labels share meaning. However, Abele et al. (2008) proposed that competence may not be synonymous with agency. In particular, Abele et al. (2008) argue that agency encompasses self-interest motivation and assertiveness, whereas competency only refers to the ability to carry out intents. Therefore, qualities of competency and communion can be seen as compatible, while self-focused agentic traits are considered incompatible with other-focused communal traits (Schwartz, 1992).

Indeed, people respond more favourably to dominant men than dominant women (Rudman et al., 2012), but respond similarly to competent men and competent women (Rudman & Glick, 2001). In support of this, over time, women have been perceived as more competent – but not dominant – by both others (Diekman & Eagly, 2000) and themselves (Spence & Buckner, 2000). Further support from Eagly et al.'s (2019) meta-analysis found that perceived gender differences in levels of competency have significantly decreased, to the point that men and women are now seen as equally competent.

Within the camp of researchers that operationalise competency as a dimension of agency, some include 'dominance' as a second, separate dimension of agency (e.g., Rudman & Glick, 2001; Rosette et al., 2016), while others include 'assertiveness' (e.g., Abele et al., 2016; Hentschel et al., 2019), and 'independence' (e.g., Hentschel et al., 2019). Additionally, recent research has further dissected competence into instrumental competence and leadership competence (Hentschel et al., 2019). Among researchers who measure competency as a separate construct to agency, varying other dimensions of agency are specified, including *self-reliance* and *dominance* (Schaumberg & Flynn, 2017), and *ambition* and *administration* (Livingston, 2013), while most others simply measure agency as a single factor (e.g., Eagly et a., 2019; Carrier et al., 2014; Koenig & Eagly, 2014).

Carrier et al. (2014) aimed to disentangle the relationship between agency and competence and found that they indeed operate independently. Carrer et al. (2014) conducted a factor analysis on their 9 items relating to warmth (warm, likeable, friendly), competence (capable, competent, efficient), and agency (ambitious, self-confident, assertive). The three factor model, in which competence and agency were separate factors, performed much better than the two factor model, in which agency and competency items were combined into one factor, suggesting agency and competence are distinct from each other. Likewise, Koenig and

Eagly (2014) conducted factor analyses on a pool of 18 agentic and communal traits and, when evaluated as gender-trait stereotypes, a three-factor solution emerged: communality, agency, and competency, suggesting agency and competency are separate constructs in relation to gender stereotypes. Conversely, when these same items were evaluated across other social groups, a two-factor solution of communality and agency/competency emerged (e.g., Fiske et al., 2002; Koenig & Eagly, 2014), suggesting that the differentiation of agency and communality is unique to gender stereotypes. This makes theoretical sense as competency can be related to both communal and agentic roles (Koenig & Eagly, 2014). However, Carrier et al.'s (2014) findings were in relation to general social judgement, not gender specifically, indicating mixed support for distinguishing between agency and competency beyond gender stereotypes. Despite the more consistent empirical and theoretical support for distinguishing competency from agency when investigating gender stereotypes, many researchers still test and discuss the two interchangeably as one construct.

Even when competency and agency are dissociated from each other, measuring agency as a single factor may be too high-level in conceptualisation to be theoretically or empirically useful. For example, in one study, Koenig & Eagly (2014) measured agency and competency separately, based on a factor analysis, and agency was not further delineated into different dimensions. However, upon examination of the specific agentic-trait items included in the factor solution (dominant, arrogant, boastful, egotistical, daring; α = .86), it appears they may have been tapping into a dominance-specific dimension of agency. Indeed, the item 'assertive' was originally included, but it cross-loaded onto the competence factor and was therefore excluded. While these agentic items certainly measure a facet of agency, they may be measuring just that - one facet of agency. It is unlikely that they adequately capture the whole concept of agency; thus, the composite score may have been more accurately and usefully labelled with a lower-level abstraction of 'dominance'.

Likewise, Eagly et al.'s (2019) meta-analysis found that men still maintained an advantage over women in perceived agency between 1946 to 2018. However, when taking a closer look at Eagly et al.'s (2019) meta-analysis, which also operationalised agency as a single factor, a variety of agentic items were included relating to dominance (e.g. aggressive, possessive) as well as assertiveness (e.g. ability to make decisions, confident, independent). Unfortunately, the factor structure of the scale was not tested. Therefore, it is possible that Eagly et al.'s (2019) agentic items were tapping into two (or more) separate constructs and that the dominance-related items may have carried the perceived gender differences in agency.

Taken together, there may be an important distinction between the agentic dimensions of assertiveness and dominance. Indeed, Abele et al. (2016) did not find a gender difference in assertiveness (or competence); however, Hentschel et al. (2019) did. These conflicting findings may be a result of the semantic difference in items used (see Figure 5.1). For example, Hentschel et al. (2019) did not find gender differences in self-ratings or ratings of men/women on *independence* (or instrumental competence), which appears to be more similar to Abele et al.'s (2016) measure of *assertiveness* (see Figure 5.1).

Overall, while some researchers (e.g., Carrier et al., 2014; Livingston, 2013) measure a motivational component of agency, others (e.g., Koenig & Eagly, 2014) measure a dominance domain of agency, but they are both being assigned the same high-level label of agency. This is problematic as these seem to be conceptually and psychometrically different constructs.

Items

As well as the inconsistencies in operationalisation of agency and competency, as described above, there are inconsistencies in the items specified to each factor. In Carrier et al. (2014)'s study, the agentic items were described by the authors as targeting a *motivational* or ambitious component of agency. In contrast, 'ambitious' was categorised under competence by Koenig and Eagly (2014). Similarly, Livingston (2013) describes a distinction between *ambitious* agency (seeking power, self-promotion, and ambition) and *administrative* agency (assertiveness and task accomplishment), the latter appearing quite similar to competence.

Similarly, Koenig & Eagly (2014) assigned *dominant* under 'agency', while Hentschel et al. (2019) measured it as 'assertiveness'. In Carrier et al. (2014)'s research, assertiveness was measured as part of motivational/ambitious agency, while in Livingston (2013)'s paper it was not assigned to their motivational/ambitious factor, but their *administrative* factor. Thus, as well as the issue of measuring different psychological constructs (e.g., dominance and competence) and labelling them the same construct (e.g., agency), there is the issue of measuring the same construct and labelling it as different constructs.

Such differences in measurement and meaning may in part explain the mixed findings in both gender stereotypes and self-perceived gender-trait differences. For example, according to Twenge's meta-analyses (1997; 2001), self-reported agency has increased in women over time to the point of mostly no reported gender differences by the l990s. Importantly, agency was measured as masculinity with the BSRI and PAQ (Twenge, 1997)

and as dominance and assertiveness – the terms were used interchangeably (Twenge, 2001). Upon examination of the scales included, it appears the ones measuring 'dominance' were measuring something closer to social/leadership competence or, as referred to in the paper, assertiveness. For example, 'speaking in public never bothers me' and 'enjoy starting a conversation with a stranger' are example items from the dominance scale of the Bernreuter Personality Inventory; and the Edwards Personal Preference Schedule (EPPS) dominance scale measures 'a need to be a leader and influence others'. Therefore, one might conclude that women's levels of assertiveness, but perhaps not dominance, are no longer different than men's.

Communality

Communality, on the other hand, persists as a female-trait stereotype among both men and women (Diekman & Eagly, 2000; Hentschel et al., 2019); and women's self-ratings are significantly higher than men's (Spence & Buckner, 2000; Hentschel et al., 2019; Donnelly & Twenge, 2017). However, the gender difference in self-rated communality also appears to be narrowing over time. Women's self-rated communality decreased between 1993 and 2012 *(d* = -.26; Donnelly & Twenge, 2017), suggesting a devaluation of women identifying as feminine or communal since the 1990s. But when the time period was expanded to include studies between 1974 - 2012, no differences were found (Donnelly & Twenge, 2017). This may be explained by the second-wave feminist movement in America in the 1970s and early 1980s. Conversely, the meta-analytic data show that men's levels of self-rated communion have remained low and not changed over time (Twenge, 1997; Donnelly & Twenge, 2017). However, men rate tend to rate themselves higher in communality (concern for others, emotional sensitivity, and sociability) than they rate others in their gender group (Hentschel et al., 2019), further evidencing the stereotype of female-communal.

Communion and agency were traditionally seen as in opposition to one another, such that being perceived as communal or feminine means being perceived as not agentic or masculine and vice versa (Foushee et al., 1979). Indeed, early researchers discussed and measured these two psychological constructs as a bipolar construct (Spence et al., 1975; Bem, 1994), with the understanding that what is seen as desirable for a man is undesirable for a woman and vice versa. However, current gender-trait differences and stereotypes appear to be positioning communality and agency more independently of each other: associations

between female-communal have remained high, whereas associations between male-agentic is more muddled. Thus, while gender stereotypes related to agency are more complex, gender stereotypes related to communality largely persist.

Present Study

The overarching aim of this study was to improve current understanding of gender-trait stereotypes through examining potential measurement issues with gender-trait stereotype questionnaires.

As the variance in male-agentic stereotype endorsement was relatively small in all previous chapters, and the mean was consistently very close to the centre value of 0 (*equally true of men and women),*

(1) the first aim was to investigate the issue of anchoring on the centre response choice. This was investigated in two ways:

First, the response options were edited to include labels for all seven response choices for each trait (-3 = *more true of women,* -2 *moderately more true of women,* -1 = *slightly more true of women,* 0 = *equally true of men and women,* 1 = *slightly more true of men,* 2 = *moderately more true of men,* 3 = *more true of men*). The versions used in previous chapters included labels only on the extreme end points (-3 = *more true of women*; +3 = *more true of men*) and centre point (0 = *equally true of men and women);* the other four options on the 7-point scale (-2, -1, +1, +2) were left blank. Krosnik (1991) suggests fully labelled scales are better for assessing attitudes. Indeed, previous methodological research on attitudinal scales has found that labeling only some response options evokes more responses on those values and that full labeling helps to reduce this response bias (Moors et al., 2014).

Some research suggests that the extremes of Likert-type scales tend to be selected less than central choices (Guilford, 1954). This may make reasonable sense given that we may expect averages to generally sit around the middle and assume normal distributions of data. However, this can lead to issues of kurtosis and lack of variability; importantly, if this scoring tendency is a response bias, it may not accurately reflect true values. Likewise, when respondents are asked to directly compare men and women on a trait, and they see *equally true of both men and women* as a response option, this may prime social desirability bias and influence them to respond in a way that they would not have otherwise.

Thus, second, we constructed and tested separate scales to evaluate each gender separately – one for the assessment of men and one for the assessment of women - and calculated a difference score by subtracting the ratings of women from the ratings of men on each trait. This method may allow for a more accurate representation of gender stereotypes, as respondents are asked to think of each gender group on its own rather than directly comparing on a bipolar scale. Additionally, it may circumvent the potential issue of social desirability bias related to centre-anchoring responses, by eliminating the centre cue of *equally true of men and women*. To help achieve this, the two scales (one for assessing men, one for assessing women) were presented on separate pages, and respondents could not return to their responses once the first page was submitted.

Research that measures gender-trait attributions (or, the trait dimension of gender stereotypes) often use different indexes of items, including their own (e.g., Abele, 2003; Abele & Wojciszke, 2007; Rudman & Glick, 2001; Hentschel et al., 2019; Koenig & Eagly, 2014; Carrier et al., 2014 Cejka & Eagly, 1999; Van Veelen & Derks, 2021; Charlesworth & Banaji, 2021; Eagly et al., 2019; see Figure 5.1) However, more consistently, traits are evaluated on a bipolar scale from *more true of women* or *strongly female* to *more true of men* or *strongly male* (e.g., Rudman & Glick, 2001; Charlesworth & Banaji, 2021; Eagly et al., 2019).

(2) The second aim of this research was to determine a theoretically and empirically derived factor structure of gender-trait stereotypes.

Given the mounting evidence that agency, or certain dimensions of agency, is no longer a gender-stereotyped trait (Chapters 2, 3, and 4) and/or that the PAQ's measurement of agency is insufficient (Chapters 3 and 4), confirmatory factor analyses were run to test the validity of the PAQ.

Confirmatory Factor Analyses

We examined the psychometric validity of the revised PAQ scales (Ward et al., 2006; then Curran & Warber, 2011), including the factor structure and uni/multidimensionality using CFAs. CFAs test theory-based factor structures and allow for the measurement of unidimensionality within each factor. Each indicator (or item) must load onto a single factor in order to be unidimensional. If indicators cross-load, then more than one construct is being estimated by the indicator, making it uninterpretable (Anderson & Gerbing, 1988; Curran &

Warber, 2011). In order to assign meaning to the constructs, each factor must achieve acceptable unidimensionality.

The PAQ is one of the most popular scales used in gender-trait research as well as the measure used throughout this thesis. However, the PAQ items for agentic and communal traits have not been updated since the scale was published in 1974; and there have been substantial social changes related to gender in the last five decades. For example, in the United States in the same year (1974), it only first became illegal to force pregnant women to leave their job on the basis they were incapable of working while pregnant. Measures like the PAQ and BRSI were previously well validated (see Coffé, 2018), but are arguably now outdated and may not be accurately capturing current gender stereotypes given the substantial social changes in the last five decades. (Donnelly & Twenge, 2017).

Therefore, it is important to conduct confirmatory factor analyses (CFAs) on previously validated scales to ensure they are still psychometrically sound (Levine et al., 2006). Psychometric evaluations of the PAQ have demonstrated support for the two-factor structure, but these were 20-40 years ago (Helmreich et al., 1981; Hill et al., 2000).

The Personality Attributes Questionnaire (PAQ; Spence et al., 1973) claims to assess desirable 'masculine' and 'feminine' personality traits. The desirability ratings were determined by asking participants to rate the ideal man and ideal women (Spence et al., 1973). The original scale consists of three subscales, one for masculine traits, or traits more desirable for men than women (M); one for feminine traits, or traits more desirable for women than men (F); and one for traits in which the ideal man and ideal woman fall at opposite ends of the scale (M-F) (e.g., "aggressive" to "cries very easily"). Conversely, average ratings of the ideal man and ideal women, although significantly different, fall on the same side of the midpoint on the *M* and *F* scales.

Due to the nature of the M-F scale development, with the average ratings of the ideal man and ideal woman falling on opposite sides of the midpoint, the M-F subscale tends to differentiate men and women better than the *M* or *F* subscales (see Ward et al., 2006). However, there are a number of reasons why I and other researchers do not include it. First, it is a bipolar measure of agentic and communal traits scored in the masculine direction. This means it treats agency and communion as a unidimensional construct, with agency at one pole and communality at the other. As previously discussed, agency and communality are distinct dimensions that can vary independently of each other. Therefore, it lacks construct validity as well as face validity, and researchers have questioned what exactly the subscale is measuring (e.g., Bromberger & Matthews, 1996). Additionally, the M-F subscale contains

items that are misspecified, which diminishes the internal consistency of the subscale (Hill et al., 2000; Helmreich et al., 1981). For the purposes of this research, only the agentic and communal subscales were considered/assessed/included/evaluated.

Ward et al. (2006) and Curran and Warber (2011) are the two most detailed and recent psychometric investigations of the PAQ. Ward et al. (2006) revised and renamed the three subscales as agency (M), communion (F), and emotional vulnerability (M-F). As Ward et al. (2006) retained items with multi-dimensionality, Curran and Warber (2011) aimed to clarify the factor structure and, importantly, unidimensionality. However, this investigation was conducted a decade ago now and included only university student samples.

Further, the PAQ was developed with bipolar descriptive adjectives (e.g., "passive" at one end and "active" at the other). Combining two different traits into single items may result in interactional effects between the two traits and therefore muddle the trait intended to be measured (in this case, "active"). Therefore, in our revised PAQ bipolar measures, each item assessed the target trait as a single trait on a scale from "not at all" to "very" for the separate gender measures.

Exploratory Factor Analyses

Due to the expected poor performance of the CFAs above, exploratory factor analyses (EFAs) were conducted on a range of traits (including agency, communality, and competency) to elucidate current gender-trait associations. While CFAs are theoretically-driven, EFAs are data-driven. Practically, this means CFAs permit factor structures to be specified, whereas EFAs allow all indicators to load onto all factors.

Method

Participants

All participants (N = 127) were recruited through social media posts, email lists, and the Department of Psychology e-bulletin at the University of Cambridge. Eligibility was restricted to those fluent in English and over the age of 18. Six participants did not complete more than 50% of the questionnaires, resulting in a total of 121 participants.

The sample included mostly women (78.7%; 12 participants did not report gender; 1 indicated prefer not to say). Age ranged from 18 to 88, with a median of 24 and a mean of 29.5. Nearly 85% of the sample (N = 121) was from the UK, while 11.9% were from

European countries: Denmark, Germany, Hungary, India, Ireland, Norway, Sweden, Switzerland, and less than 5% were from the USA. Only 1 participant left school before GCSE-level and 73% had a university degree. The majority of the sample was students and about 22% of the sample were ethnic minorities (see Table 5.1) for further demographic information). Participants were not paid for their time; instead, they were given the option to enter into a draw for a £20 gift card as a thank-you for their time. Ethical approval and insurance was obtained before participant recruitment began.

Procedure

The study was advertised through the Psychology Department e-bulletin, email, and social media. The study advertisement included the invite link for participants, which directed them to the online survey hosted on Qualtrics Survey Platform. After reading the information sheet and providing informed consent, participants were randomly assigned to one of two conditions for the first phase of the study. Half of the sample received forty traits on a 7-point bipolar scale with response options labelled for *all* seven response choices (-3 = *more true of women*, -2 = *moderately more true of women*, -1 = *slightly more true of women*, 0 = *equally true of men and women*, 1 = *slightly more true of men*, 2 = *moderately more true of men*, 3 = *more true of men*), and the other half of the sample received the forty traits on a 7-point bipolar scale with response labeling on the extreme end points (-3 = *more true of women*; +3 = *more true of men*) and centre point (0 = "equally true of men and women") only. In the second phase of the study, participants evaluated men and women separately on the same forty traits assessed in phase one. This time, participants were prompted with, "In general, men/women are…" and response options were on a 7-point bipolar semantic scale (e.g., -3 = *not at all independent* to +3 = *very independent)* for both gender assessments. Once again, all participants were randomly assigned to one of two conditions: assessing men first and women second, or assessing women first and men second. Therefore, for all forty traits, all participants completed (a) a bipolar measure directly comparing men and women and (b) a measure separately evaluating men and women. Participants then completed measures related to perceived gender differences in toy preferences, the data of which is not included in this thesis. To conclude, participants were debriefed (Appendix D; Debrief Form) and offered the option to enter their email address into a draw for a £20 gift card. Ethics approval was obtained before recruitment began.

Materials and Measures

Responses for each item were optional. The display order of all measures, as well as the items within each measure, were randomised. All materials and measures were administered in Qualtrics Survey Platform and are included in Appendix D.

Table 5.1

Sample Demographics (N =121)

	n	%
Ethnicity		
White	80	77.6
Black	4	4.1
Asian	13	12.7
Mixed	6	5.7
Gender		
Female	85	78.7
Male	22	20.4
Prefer not to say	1	0.9
Age[a]		
18-24	48	44.4
25-34	41	38.0
35-44	5	4.6
45-54	3	2.8
55-64	8	7.4
65-74	0	0.0
75-84	2	2.0
85-94	1	0.9
Education		
No university degree	34	31.8
Undergraduate university degree	46	43
Graduate university degree	27	25.2
Employment		
Full-time employment	24	22.2
Part-time employment	8	7.4
Self-employed	2	1.9
Student	67	62
Unemployed, looking for work	2	1.9
Retired	5	4.6
Religion		
No religion	71	65.7
Religious	37	34.3
Country		
Europe	13	11.9
UK	92	84.4
USA	4	3.7

Note. [a] Age was collected as a continuous variable, but is presented as categorical data here for ease of interpretation. Percentage values reported are valid percentages and therefore estimate missing values as distributed proportionately among response options.

Traits

Participants assessed the typical man and typical women on a list of forty traits (see Appendix D), including widely-measured communal, agentic, and competent traits, as well as STEM-typed and leadership-typed traits (traits more specifically associated with male dominated roles). These traits were derived from previous research and widely held associations. The forty traits were assessed using two different scale types: a 7-point bipolar scale from -3 (*more true of women)* to +3 (*more true of men*) and separately by gender ("In general, men/women are…"), and a 7-point semantic differential scale from -3 (e.g., *not at all independent)* to +3 (*very independent)*. To evaluate the latter measure in analyses, a difference score was calculated by subtracting the ratings for women from the ratings for men on each trait.

Analyses: Part I

For all analyses, corrections for multiple comparisons were applied.

Results

To ensure there were no display order effects on responses, a MANOVA on all traits was conducted. It revealed no differences between participants who viewed the bipolar measure first versus participants who viewed the independent gender measure first, $p = .495$. An additional MANOVA was run on all items to test the effect of evaluating men first versus women first; it also revealed no significant differences in display order effect of men versus women, $p = .321$.

Centre-Anchoring

Response Option Labels

To test the influence of response option labeling, a MANOVA was run on the agentic items. It showed no difference in responses between participants who saw labels on all 7 response options versus those who saw labels only on the extreme values and middle values of the 7-point scale, F (6, 119) = .949, Wilk's $\Lambda = .952$, $p = .463$. Item comparisons supported the omnibus test, with all ps > .100. Therefore, the issue of anchoring on the centre

response choice in the previous studies was likely not influenced by response labeling. A subsequent MANOVA was run on all communal traits, and it also demonstrated non-significance, F (6, 119) = .967, Wilk's Λ = .887, p = .492. Item comparisons supported the omnibus test, with all ps > .100. Descriptive statistics for the item comparisons are in Table 5.2. Due to the lack of between-group differences, these two groups were combined together in subsequent analyses.

Bipolar versus Difference-Score Scales

As an initial investigation, distributions of each scale were analysed. All Shapiro-Wilk tests demonstrated significance (all ps<.001), suggesting neither scale would be preferred for meeting the assumption of normality. The bivariate relationships between the two scale types suggested there was adequate variability between the measures *(rs* between |.58| and |.67, ps < .001). Therefore, comparison analyses were conducted. Data was standardised in order to conduct the following comparison analyses. Scores were not mean-centred, so as to not lose the meaning of the 0 value. Instead, the raw values were divided by their standard deviation.

To examine whether participants responded differently on the bipolar and difference-score scales, paired-Samples t-tests, or Wilcoxon Signed-rank tests in cases of non-normality (significant Shapiro-Wilk tests), were conducted and corrected for multiple pairwise comparisons (see Table 5.2). Among the agentic items, 'feels superior' approached significance, suggesting that when men were directly compared against women, they were rated higher in 'feels superior' than when the gender groups were evaluated separately, W = 3315, z = 1.87, p = .058, r = .214. All other items were non-significant, ps > .179. As a robustness check, a Wilcoxon Signed-rank test was conducted on the agentic composite score. As expected, it revealed no differences between bipolar and difference-score averaged responses, W = 3732, z = 1.465, p = .143. Therefore, the issue of anchoring on '*equally true of men and women'* on the bipolar scale used in previous studies is likely not influenced by social desirability bias to a greater extent than separate evaluations of the gender groups are.

However, among the communal items, 'warm' demonstrated significant differences, such that women were rated more warm on the bipolar scale than on the difference-score scale, t(117) = -2.296, p = .023, d = -.212. This means that women are rated as more warm than men to a greater extent when directly compared to men than when each gender group is evaluated separately. All other communal items demonstrated non-significance, ps > .208. As

a robustness check, a Wilcoxon Signed-rank test was conducted on the communal composite score. As expected, it revealed no differences between bipolar and difference-score averaged responses, $W = 3021$, $z = .071$, $p = .945$.

Table 5.2

Descriptive Statistics, Between-Group Comparisons for Response Labelling, and Within-Group Comparisons for Scale Type

Items	Bipolar Response Labels [a]				Scale Type [d] (n =120)			
	Reduced [b] (n = 59)		All (n = 62) [c]		Bipolar Scale [e]		Difference score [f]	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Agentic								
Independent	0.00	1.25	0.14	1.06	0.06	1.00	0.06	0.97
Active	0.40	0.85	0.14	0.88	0.31	1.00	0.21	0.98
Competitive	0.71	1.09	0.82	1.15	0.68	1.00	0.68	1.00
Never gives up easily	0.08	0.97	-0.03	0.76	0.02	1.00	-0.05	1.01
Self-confident	1.05	1.31	0.98	1.13	0.83	1.00	0.76	1.00
Feels superior	1.33	1.26	1.32	1.14	1.10	1.00	0.95	1.00
Communal								
Able to devote self to others	-1.20	1.18	-1.16	0.92	-1.11	1.00	-1.14	1.00
Helpful to others	-1.01	1.16	-0.82	0.95	-0.86	1.00	-0.85	1.01
Kind	-0.71	1.11	-0.88	0.97	-0.76	1.00	-0.89	1.00
Aware of others' feelings	-1.33	1.10	-1.45	1.09	-1.27	1.00	-1.11	1.00
Understanding of others	-1.30	1.23	-1.37	1.09	-1.15	1.00	-1.18	0.98
Warm in relations with others	-1.30	1.22	-1.21	0.96	-1.15	1.00	-0.94	0.99

Note. [a] Values for between-groups analyses are standardised. Mann-Whitney U were interpreted due to non-normal data distributions. No tests demonstrated significance after correction for multiple comparisons (alpha level of $p = .004$). [b] Half of participants were randomly assigned to complete the bipolar scale with response labels on only the extreme and middle values (-3 = *more true of women*, 0 = *equally true of men and women*, and +3 = *more true of men*) of the 7-point scale. [c] The other half of participants were randomly assigned to complete the bipolar scale with response labels on all 7 response options. [d] Wilcoxon Signed-Rank tests were interpreted for the within-groups analyses due to non-normal data distributions. No tests demonstrated significance after correction for multiple comparisons (alpha level of $p = .004$). [e] Bipolar scale values include data from all participants (data combined from both 'Reduced' and 'All' response-label conditions), as there were no significant differences between these two groups (see [a]). [f] Difference score values are calculated from separate evaluations of men and women by subtracting participants' ratings of women from their ratings of men on each trait.

McNemar tests were conducted to compare the frequency of counts of 0 between the bipolar and difference-score scale. As expected, on all items, respondents reported a value of 0 (*equally true of men and women)* significantly more on the bipolar scale than the difference-score measure (see Appendix D). Thus, to examine if responses on each scale type differentially significantly differed from zero, One Sample Wilcoxon Signed-Rank tests were

performed and tested against a value of zero (see Table 5.3). Scores significantly greater than zero indicate the trait is rated as more true of men, and scores significantly less than zero indicate the trait is rated as more true of women. Findings demonstrated that for both the bipolar and difference-score scales, 'independent' and 'never gives up' were not significantly different from zero, but 'competitive', 'self-confident', and 'feels superior' were significant. After corrections for multiple comparisons, 'active' remained significant only on the bipolar scale. Thus, men were rated as significantly more active than women when compared directly to women, but when the two genders were evaluated independently from each other, the gap between men and women shrunk. This demonstrates a male advantage, such that the direct comparison of men and women appears to benefit men's perceived levels of 'active'. All communal items were significantly different from zero for both scales. Thus, the items on each scale type did not demonstrate differences when compared to zero.

Table 5.3

One Sample Wilcoxon Signed-Rank Tests for Agentic and Communal Traits

Items	Bipolar	Adj *p*	Difference-score	Adj *p*
Agentic				
Independent	$V = 1071.50$, $p = .492$, $r = .097$	ns	$V = 1117.50$, $p = .445$, $r = .097$	ns
Active	$V = 754.50$, $p = <.001$, $r = .595$	*	$V = 1422.50$, $p = .019$, $r = .097$	ns
Competitive	$V = 2271.50$, $p = <.001$, $r = .828$	*	$V = 2929.50$, $p = <.001$, $r = .097$	*
Never gives up easily	$V = 498.00$, $p = .755$, $r = .053$	ns	$V = 910.00$, $p = .634$, $r = .097$	ns
Self-confident	$V = 3867.50$, $p = <.001$, $r = .808$	*	$V = 3982.50$, $p = <.001$, $r = .097$	*
Feels superior	$V = 4374.00$, $p = <.001$, $r = .918$	*	$V = 4341.50$, $p = <.001$, $r = .097$	*
Communal				
Devote	$V = 21.00$, $p = <.001$, $r = -.988$	*	$V = 12.00$, $p = <.001$, $r = .097$	*
Helpful to others	$V = 0.00$, $p = <.001$, $r = -1.000$	*	$V = 100.00$, $p = <.001$, $r = .097$	*
Kind	$V = 39.00$, $p = <.001$, $r = -.960$	*	$V = 40.00$, $p = <.001$, $r = .097$	*
Aware of others	$V = 81.00$, $p = <.001$, $r = -.966$	*	$V = 140.50$, $p = <.001$, $r = .097$	*
Understanding	$V = 49.50$, $p = <.001$, $r = -.975$	*	$V = 14.50$, $p = <.001$, $r = .097$	*
Warm	$V = 20.50$, $p = <.001$, $r = -.989$	*	$V = 129.00$, $p = <.001$, $r = .097$	*

Note. Effect size given is rank-biserial correlation. Standard error given is for rank-biserial correlation. Adj. p = adjusted for multiple comparisons, $\alpha = .008$; * p<.008; ns = non-significant. Devote = Able to devote self completely to others; Aware of others = Aware of others' feelings; Understanding = Understanding of others; Warm = Warm in relations with others

Discussion

Overall, findings suggest the bipolar scale is no more problematic than a difference-score measure of gender-trait stereotypes. These results indicate that the lack of variation in

and endorsement of the male-agentic stereotype in the previous chapters is likely not a result of centre-anchoring due to scale type.

Asking participants to comparatively evaluate men and women on traits, versus considering and evaluating the gender groups separately, may elicit a cognitive process more reflective of the real world. For example, when hiring an applicant for a STEM position, it is likely that the applicant pool is not all female. Therefore, all applicants, including men, women, and gender minorities, are evaluated against each other for suitability for the position. To minimise this comparative bias, some hiring practices have been recommended, which require evaluating each candidate separately and blindly. While these hiring procedures appear to be effective for the first round of the selection process, it is challenging to implement such techniques for the interviewing stage – as well as the evaluation and promotion processes after hiring.

However, it could also be argued that bias against women in STEM exists independently of a direct comparison to men. Applying the descriptive stereotype that women are communal, not agentic, can disadvantage women in STEM, or women interested in STEM, regardless of the explicit presence of male competition.

Analyses: Part II

Based on the findings above, as well as the poor internal consistency of the male-agentic stereotype in the previous chapters, it is likely that the measurement issues with the male-agentic stereotype is related to the items not the response type.

To examine the factor structures of agentic, communal, and competent scales, a series of CFAs were conducted. The aim of the CFAs were to evaluate the psychometrically-revised PAQ (agentic and communal) scales by Ward et al. (2006) and Curran and Warber, (2011), as well as the competence scale put forth and used by Eagly et al. (2019).

The following model fit indices were assessed and reported: chi-square (X^2), to assess overall model fit and discrepancy between the sample and fitted covariance matrices; comparative fit index (CFI), compares the fit of a target model to the fit of the null model; Tucker Lewis index (TLI), also compares the model of interest relative to the null model - and is suitable for smaller samples; root mean square error of approximation (RMSEA), a parsimony-adjusted index; and standardised root mean square residual (SRMR), the standardised square root of the difference between the residuals of the sample covariance

matrix and the hypothesized model (Kline, 2015). In addition, modification indices and residual covariances were examined to identify multidimensionality and misspecificity.

Following Hu and Bentler (1999), Kline (2005), and Browne and Cudeck (1993), the following standards were used to assess model fit: 0.90, or .095 to be conservative, for the CFI and TLI; 0.08 = mediocre, .05 = good, .01 = excellent for the RMSEA; less than .09 for the SRMR; and a non-significant chi-square is desirable, but rare in social psychology. Modification indices were set at a cut-off of 3.84, $p<.05$. Finally, as suggested by Kline (2016), a threshold of less than |.10| was used for standardardised residuals.

Factors and variables were permitted to correlate, as it is unlikely for psychological constructs to be completely uncorrelated (Field, 2017). However, correlations between constructs were not permitted to exceed 0.85 (Hair et al., 2010). Initial bivariate correlations were run on all items. All item pairs demonstrated bivariate correlations of less than 0.8 and therefore all were included in the factor analyses.

Analyses were conducted in SPSS version 26 and 27 as well as JASP using EQS emulation.

Results

Communal and Agentic Traits

Ward et al. (2006) and Curran and Warber (2011) are the two most detailed and recent psychometric investigations of the PAQ. Ward et al. (2006) revised and renamed the three subscales as agency (M), communion (F), and emotional vulnerability (M-F). As Ward et al. (2006) retained items with multi-dimensionality, Curran and Warber (2011) aimed to clarify the factor structure and, importantly, unidimensionality. Initial CFAs were run on Ward et al.'s (2006) model and confirmed unacceptable model fit (see Table 5.4; Appendix D; Supplementary Results and Supplementary Figures). Curran and Warber's (2011) factor model, the one used in Chapters 3 and 4, was then examined and reported below.

Bipolar Scale

First, the Curran and Warber (2011) revised model was tested on the bipolar scale. Curran and Warber (2011) identified the same problematic items as Ward et al., (2006), as well as four additional ones: active (M), never gives up easily (M), feels superior (M), and helpful (F). Therefore, these were removed in their revised model. The CFA suggested an

improved model fit from both the bipolar and difference-score Ward et al. (2006) versions, but still not desirable, X^2 (19, N = 121) = 41.31, p=.002 (CFI = .944, TLI = .917, RMSEA = .099, SRMR = .104). The factor correlation was -.293, p = 0.006. The internal reliability of the communal subscale was excellent (α = 0.90), while the internal reliability of the agentic subscale was unacceptable (α = 0.52). Curran and Warber (2011) noted the internal consistency of their agentic subscale was poor and considered unacceptable (α = 0.57). The authors suggested that the agentic subscale be dropped entirely.

"Independent" did not significantly load onto a factor (see Table 5.4). As the previously problematic items, "active", "never gives up easily", "feels superior", and "helpful", were removed from this model, there were no issues to be found with these items. However, as with the Ward et al. (2006) model, modification indices indicate that "independent" cross-loads onto the communal factor and "able to devote self completely to others" cross-loads onto the agentic factor.

Table 5.4

CFA Standardised Factor Loadings of Agentic and Communal Items

	Bipolar Scale		Difference Score	
Items	**Ward et al. (2006) Model**	**Curran et al. (2011) Model**	**Ward et al. (2006) Model**	**Curran et al. (2011) Model**
Agentic				
Independent	0.111	0.121	0.290**	0.285**
Active	0.409***		0.399***	
Competitive	0.737***	0.737***	0.467***	0.482***
Never gives up easily	0.260*		0.160	
Self-confident	0.729***	0.758***	0.80***	0.760***
Feels superior	0.728***		0.685***	
Communal				
Devote	0.750***	0.750***	0.770***	0.760***
Helpful	0.827***		0.777***	
Kind	0.762***	0.737***	0.619***	0.620***
Aware of others' feelings	0.791***	0.825***	0.791***	0.800***
Understanding of others	0.826***	0.847***	0.881***	0.884***
Warm	0.821***	0.784***	0.619***	0.610***

Note. * p<.05, ** p<.01, *** p<.001

The residual correlations between variables (i.e., the standardised residual covariance matrix) suggests there are considerable issues with some items. Ten residuals were equal to or greater than 0.10 in the standardised residual matrix (Figure 5.2). The RMSEA is based on the residual covariances, so these deviations are driving the undesirable RMSEA value in this model. Although these items demonstrated statistically significant factor loadings, they demonstrated multiple standardised residuals greater than 0.10.

Further, there were several covariances between items on the same latent factor, which means they share additional relationships beyond their loadings on the agentic/communal factors. On the agentic subscale, between "competitive" and "self-confident"; and on the communal subscale between "able to devote self completely to others" and "understanding of others", "able to devote self completely to others" and "aware of feelings of others", and "aware of feelings of others" and "kind". This can be explained if they are more closely correlated with each other than the other items on the factor. With the exception of the covariance between "able to devote self completely to others" and "aware of feelings of others", these are all different covariances than the Ward et al. (2006) model.

Lastly, the modification index for residual covariances indicated extra covariance between items that do not share the same latent factor - "independent" and "kind", and "competitive" and "able to devote self completely to others". This further explains the multidimensional and unacceptable factor structure. These relationships were also different from our findings from the Ward et al. (2006) model. While the modification indices indicate the parameters that would significantly improve the model fit if we included them in the model, it also needs to make theoretical sense to free the parameters. Further, these indices are subject to sampling bias, so including the parameters in the model can result in the issue of overfitting. Taken together, an EFA may be more appropriate.

Difference-score Scale

Next, a CFA on the difference-score scale of the Curran & Warber (2011) model also demonstrated poor fit, X^2 (19, N = 121) = 51.719, p < .001 (CFI = .90, TLI = .852, RMSEA = .123, SRMR = .084). The factor correlation was -.502, p < .001. The internal reliability of the communal subscale was very good (α = 0.85), while the internal reliability of the agentic subscale was unacceptable (α = 0.60).

Unlike in the bipolar scale, "independent" significantly loaded onto the agentic factor, although with less significance than the other indicators. The residual correlations between

variables suggests there are considerable issues with some items. Fourteen residuals were equal to or greater than 0.10 in the standardised residual matrix, four more than the bipolar version (Figure 5.2). Although these items demonstrated statistically significant factor loadings, they also demonstrated multiple standardised residuals greater than 0.10. The RMSEA is based on the residual covariances, so these deviations are driving the undesirable RMSEA value in this model.

As items "active", "never gives up easily", "feels superior", and "helpful" were removed for this model, there were no issues to be found with these items. As with the bipolar Ward et al. (2006) model, "independent" cross-loads onto the communal factor and "able to devote self completely to others" cross-loads onto the agentic factor. Additional cross-loadings were found, with "self-confident" and "never gives up easily" cross-loading on the communal subscale.

Additionally, there were multiple covariances between items on the same latent factor, which means they share additional relationships beyond their loadings on the agentic/communal factors; on the agentic subscale, between "competitive" and "self-confident" and "competitive" and "never gives up easily"; and on the communal subscale between "understanding of others" and "warm", "aware of feelings of others" and "warm", and "able to devote self completely to others" and "aware of feelings of others".This can be explained if they are more closely correlated with each other than the other items on the factor.

Lastly, the modification index for residual covariances indicated extra covariance between items that do not share the same latent factor - "self confident" and "able to devote self completely to others", "independent" and "aware of feelings of others", "competitive" and "kind", and "self confident" and "understanding of others".

According to the modification indices, removing the equality constraint on the parameters above would significantly improve the model fit. However, as these are based on the specific sample, it is not advisable to include them unless there is very good theoretical reasoning. Once again, due to the sample size and to avoid 'overfitting', an EFA may be more appropriate.

Overall, the CFAs above suggest that Curran and Warber's (2011) model does not demonstrate satisfactory fit, neither with the bipolar nor the difference-score scale.

Figure 5.2.

CFA Standardised Residual Covariance Matrix of Agentic and Communal Items for the Curran & Warber (2011) Model

	1	2	3	4	5	6	7	8
1 - Independent		0.01	0.05	0.24	0.34	0.26	0.19	0.22
2 - Competitive	0.04		0.00	0.16	0.02	0.04	0.02	0.03
3 - Self confident	0.03	0.05		0.09	0.07	0.01	0.02	0.08
4 - Devote	0.14	0.04	0.21		0.01	0.05	0.05	0.00
5 - Kind	0.17	0.18	0.03	0.01		0.05	0.04	0.03
6 - Aware of feelings of others	0.29	0.05	0.06	0.06	0.01		0.01	0.02
7 - Understanding of others	0.13	0.05	0.02	0.02	0.01	0.00		0.01
8 - Warm	0.14	0.01	0.09	0.08	0.06	0.08	0.05	

Note. Values on the top right of the diagonal correspond to the bipolar measure and values on the bottom left of the diagonal correspond to the difference-score measure.

Competence Traits

Next, two CFAs were conducted on Eagly et al.'s (2019) competence measure. To the best of my knowledge, the traits were not and have not yet been validated as a psychometrically sound factor.

Bipolar Scale

The CFA of Eagly et al.'s (2019) competence items on the bipolar scale demonstrated extremely poor fit, X^2 (35, N = 121) = 169.860, p < .001 (CFI = .580, TLI = .460, RMSEA = .181, SRMR = .144). All indicators demonstrated problematic residual covariances, and nearly every indicator exhibited standardised residuals above .10. Taken together, these indices suggest an unacceptable model fit.

In terms of factor loadings, "level-headed" and "good with details" did not significantly load onto the factor. "Innovative", "accepts new ideas", and "logical" loaded to a lesser extent than "able to create and invent new things", "creative", intelligent", organised", and "smart" (see Table 5.5).

Table 5.5

CFA Factor Loadings of Competence Items from Eagly et al. (2019)

Items	Bipolar	Difference Score
Able to create and invent new things	0.379***	0.355***
Creative	0.322***	0.554***
Innovative	0.291**	0.249*
Intelligent	0.879***	0.656***
Level-headed	0.185	0.379***
Organised	0.350***	0.526***
Smart	0.921***	0.493***
Good with details	0.150	0.439***
Accepts new ideas	0.250**	0.600***
Logical	0.239*	0.225*

The residual correlations between variables suggests there are considerable issues with most items. Sixteen residuals were equal to or greater than 0.10 in the standardised residual matrix (Figure 5.3). The RMSEA is based on the residual covariances, so these deviations are contributing to the unacceptable RMSEA value in this model. As this scale was treated as one factor, there were no possible issues of cross-loadings. However, there were fifteen residual covariances between the 10 items, which means they share additional relationships beyond their loadings on the competence factor. Specifically, “intelligent and “smart”, “organised” and “accepts new ideas”, “able to create and invent new things” and “innovative”, “organised” and “good with details”, “able to create and invent new things” and “level-headed”, “creative” and “organised”, “smart” and “good with details”, “creative” and “good with details”, “level-headed” and “logical”, “good with details” and “accepts new ideas”, “intelligent” and “level-headed”, “creative” and “accepts new ideas”, “innovative” and “good with details”, “innovative” and “smart”, “creative” and “logical”.

Figure 5.3

CFA Correlation Matrix of Competence Items from Eagly et al. (2019)

	1	2	3	4	5	6	7	8	9	10
1 - Create/invent new things	0.00	0.08	0.43	0.03	0.31	0.10	0.00	0.08	0.00	0.12
2 - Creative	0.04	0.00	0.12	0.02	0.01	0.27	0.01	0.28	0.17	0.17
3 - Innovative	0.37	0.16	0.00	0.02	0.15	0.03	0.04	0.18	0.03	0.04
4 - Intelligent	0.02	0.02	0.00	0.00	0.07	0.05	0.01	0.03	0.04	0.00
5 - Levelheaded	0.04	0.15	0.08	0.16	0.00	0.03	0.00	0.10	0.04	0.27
6 - Organised	0.21	0.14	0.21	0.13	0.05	0.00	0.01	0.36	0.49	0.01
7 - Smart	0.07	0.02	0.03	0.12	0.06	0.07	0.00	0.08	0.03	0.01
8 - Details	0.25	0.11	0.17	0.02	0.05	0.40	0.23	0.00	0.25	0.03
9 - Accept new ideas	0.10	0.03	0.16	0.07	0.02	0.07	0.03	0.02	0.00	0.06
10 - Logical	0.22	0.16	0.21	0.05	0.12	0.12	0.22	0.25	0.07	0.00

Overall, due to the small sample size and to avoid the issue of ‘overfitting’, these findings suggest an EFA may be more appropriate.

Difference Score

The confirmatory factor analysis of Eagly et al.’s (2019) competence items measured on the difference-score scale demonstrated extremely poor fit, X^2 (35, N = 121) = 155.711, p < .001 (CFI = .500, TLI = .357, RMSEA = .176, SRMR = .135). Like the bipolar/difference score scale, all indicators demonstrated problematic residual covariances, and nearly every indicator exhibited standardised residuals above .10. Taken together, these indices suggest an unacceptable model fit.

All items significantly loaded onto the factor structure, including “level-headed” and “good with details”, which did not significantly load on the bipolar scale. However, “innovative” and “logical” loaded to a lesser extent than all other indicators (see Table 5.5)

The residual correlations between variables suggested there are considerable issues with some items. Twenty-two residuals were equal to or greater than 0.10 in the standardised residual matrix (Figure 5.3). The RMSEA is based on the residual covariances, so these deviations are contributing to the unacceptable RMSEA value in this model.

As this scale was treated as one factor, there were no possible issues of cross-loadings. However, there were nineteen residual covariances between the 10 items, which means they share additional relationships beyond their loadings on the competence factor. Specifically, “organised” and “good with details”, “smart” and “good with details”, “able to create and invent new things” and “good with details”, “good with details” and “logical”, “able to create and invent new things” and “organised”, “logical” and “smart”, “intelligent” and “organised”, “innovative” and “organised”, “intelligent” and “level-headed”, “able to create and invent new things” and “logical”, “creative” and “organised”, “intelligent and “smart”, “innovative” and “logical”, “innovative” and “accepts new ideas”, “creative” and “level-headed”, “creative” and “logical”, “creative” and “innovative”, and “innovative” and “good with details”. Based on the above findings, an EFA may be more appropriate.

Overall, the difference score measure demonstrates a slightly better fit than the bipolar measure, but, once again, neither the bipolar scale nor the difference-score scale demonstrated acceptable fit for Eagly et al.’s (2019) competence measure.

Discussion

Taken together, all four CFA models demonstrated unacceptable model fit and indicate EFA is necessary. The difference-score scale demonstrated small to negligible superiority to the bipolar scale in all models. Given the similarity in performance and that the difference-score scale is double the length, the bipolar measure would likely be preferred in research to avoid fatigue effects. Thus, the bipolar measure will be explored in the EFA, though the difference-score solution is presented in Appendix D; Supplemental Tables.

Analyses: Part III

EFAs were run using minimum residual estimation, which produces solutions similar to maximum likelihood, but is more robust. Oblique methods allow factors to correlate, whereas orthogonal methods produce uncorrelated factors. Theoretically and empirically, most arguments suggest agentic and communal are correlated (cf. Koenig & Eagly, 2014). Additionally, if factors are truly uncorrelated, oblique and orthogonal rotations should produce nearly identical results (Costello & Osborne, 2005), which they did not in this dataset. Therefore, oblique rotation should render a more accurate solution and was applied,

using direct oblimin, to allow factors to correlate. As advised by Fabrigar et al. (1999), in cases of non-normally distributed data, the analyses were re-run using principal axis factoring estimation and the factors and item loadings did not change.

Although factor loadings greater than 0.4 are generally used as a threshold, Stevens (2002) notes the loading should be greater for smaller samples (e.g., 0.512 for sample size of 100). Therefore, based on our sample of 121, a cut-off of 0.5 was used for retaining indicators. Likewise, the common method of retaining all factors with eigenvalues larger than 1.0 is not considered an accurate method (Velicer & Jackson, 1990). Parallel analysis is more robust and, therefore, the method used in the following EFAs. Following Kaiser and Rice (1974), the Kaiser–Meyer–Olkin values were considered mediocre between 0.5 - 0.7, good between 0.7 - 0.8, great between 0.8 - 0.9, and superb above 0.9 (Hutcheson & Sofroniou, 1999; Field, 2017).

Results

The Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis, KMO = 0.82, and all KMO values for individual items were greater than 0.67. Table 5.6 shows the factor loadings after rotation. Based on item clusterings, factor 1 represents the traditional female-trait stereotype of communality. Indeed, it includes all the communal traits from the PAQ, as well as some from Eagly et al.'s (2019) competence dimension. Factor 2, however, does not appear to represent the traditional male-trait stereotype. Indeed, it does not include any of the agentic items from the PAQ, with the exception of 'competitive'. It includes two additional items from other competence and/or agentic and/or assertive subscales, namely, 'ambitious' and 'level-headed' (e.g., Abele et al., 2016; Eagly et al. 2019; Koenig & Eagly, 2014, Carrier et al., 2014). These together with the other items, which include both agentic traits (e.g., ambitious, competitive) and competence traits (e.g., mathematical, able to create and invent new things), appear to represent a STEM-aptitude dimension of agency. The third factor included only 2 items (less than the typically recommended three), but was retained for exploratory purposes. It appears to represent a relational dimension of leadership competency. Together, they explained 44% of the variance, with Factor 1 (female-communal stereotype) accounting for nearly half the variance. Table 5.6 shows the factor loadings after rotation.

Descriptive statistics of the items are presented by factor in Table 5.7. These indicate that, indeed, communality remains a solid female-trait stereotype. Meanwhile, the STEM-aptitude factor emerged on the side of male-trait stereotype, but to a much weaker extent.

Lastly, the relational dimension of leadership competency does not appear to be a gender stereotyped domain.

Table 5.6

Factor Loadings

	Factor 1	Factor 2	Factor 3	Uniqueness
Helpful	0.837			0.282
Understanding of others	0.813			0.342
Warm	0.804			0.356
Kind	0.772			0.367
Empathic	0.762			0.386
Aware of others' feelings	0.758			0.421
Devote	0.756			0.412
Collaborative	0.712			0.478
Organised	0.677			0.521
Honest	0.575			0.570
Encouraging of others	0.557			0.659
Accepts new ideas	0.551			0.685
Able to assess others	0.538			0.660
Good communicator	0.521			0.669
Good with details	0.506			0.705
Mathematical		0.814		0.397
Analytical		0.712		0.516
Logical		0.696		0.542
Able to create/invent new things		0.661		0.399
Competitive		0.607		0.409
Level-headed		0.584		0.575
Ambitious		0.565		0.583
Persuade others			0.692	0.531
Able to convince others			0.692	0.531
Independent				0.798
Active				0.718
Never gives up easily				0.710
Self-confident				0.509
Feels superior				0.489
Decisive				0.683
Creative				0.651
Innovative				0.575
Intelligent				0.523
Smart				0.561
Likeable				0.765
Driven				0.821
Charismatic				0.697
Curious				0.624
Able to inspire others				0.549
Strategic				0.641

Note. Applied rotation method is oblique oblimin. Items measured on a scale from -3 (*more true of women)* to +3 (*more true of men)*

Table 5.7

Descriptive Statistics for Items

	M	*SD*	Min.	Max.
Factor 1				
Helpful	-0.92	1.06	-3.00	0.00
Understanding of others	-1.33	1.16	-3.00	2.00
Warm	-1.26	1.09	-3.00	1.00
Kind	-0.80	1.05	-3.00	1.00
Empathic	-1.44	1.20	-3.00	1.00
Aware of others' feelings	-1.40	1.10	-3.00	2.00
Devote	-1.18	1.06	-3.00	1.00
Collaborative	-0.67	1.03	-3.00	1.00
Organised	-0.93	1.09	-3.00	1.00
Honest	-0.40	0.93	-3.00	1.00
Encouraging of others	-0.62	1.04	-3.00	1.00
Accepts new ideas	-0.71	1.07	-3.00	2.00
Able to assess others	-0.59	0.94	-3.00	1.00
Good Communicator	-0.72	1.02	-3.00	1.00
Factor 2				
Mathematical	0.35	0.90	-3.00	3.00
Analytical	0.26	1.11	-3.00	3.00
Logical	0.14	0.99	-3.00	3.00
Able to create/invent new things	0.02	0.82	-3.00	2.00
Competitive	0.77	1.12	-3.00	3.00
Level-headed	-0.13	1.12	-3.00	3.00
Ambitious	0.33	1.06	-3.00	3.00
Factor 3				
Persuade others	0.07	1.20	-3.00	3.00
Able to convince others	-0.02	1.04	-3.00	3.00

Note. Items measured on a scale from -3 (*more true of women)* to +3 (*more true of men)*

Findings from the difference-score scale demonstrated similar findings (Appendix D; Supplementary Tables). Although factor 1 and factor 2 appeared to tap into the same female versus male stereotype constructs, respectively, there were some differences in the items that loaded. Additionally, five items loaded onto factor 3, representing a more general measure of leadership competency, not just the relational aspect. However, the overall MSA was lower (.76) and the cumulative variance explained was also lower (40%). Therefore, the bipolar measurement model is a stronger model.

Discussion

The present findings suggest that the traditionally conceptualised global factor of agency is perhaps no longer measuring a male-trait stereotype. Indeed, nearly none of the commonly assessed agentic traits significantly loaded onto a male-agentic stereotype factor. Instead, it appears the current male-trait stereotype may be a STEM-aptitude dimension of

agency. Further research is needed to better understand these preliminary findings and current male-trait stereotypes. Nonetheless, these findings shed light on the poor internal consistency, variance, and performance of the male-agentic stereotype as it was measured in Chapters 2, 3 and 4.

Agency can involve cognitive, physical, or personality traits, which may help to explain inconsistent findings across studies that use the term 'agency'. As well, this may help to explain the poor internal consistency in measures like the PAQ that include more than one of these dimensions (e.g., 'active' and 'independent'). Thus, it may be useful to update current operationalisations of agency in a way that distinguishing these dimensions.

These findings suggest using generalised measures of agency and competency may not be suitable in gender-trait stereotype research. Further, the distinction between agency and competency did not emerge as in these findings, likely due to including only competence- and assertive-based agentic items (versus dominant-agentic items). Thus, we question whether there is theoretical or empirical value in using the global terms of 'agency' and 'competency' in gender-trait research, or if lower-level abstractions may be more useful.

Alternatively, communality appears to remain a psychometrically sound and consistent gender-trait stereotype. All communal items loaded onto the female-trait stereotype factor as expected.

Considering the consistently strong communal stereotype assigned to women, and therefore not men, it may be understood that women's underrepresentation in STEM and other fields is not only due to a perceived lack of agency, but rather a rejection of communality. Future research would benefit from examining the relationship between female-communal stereotype endorsement and gender-STEM stereotype endorsement. It may serve as a better indicator than the male-agentic stereotype in the perpetuation of underrepresentation of women in STEM. Likewise, the female-communal stereotype may be a more effective target myth than the male-agentic stereotype in gender-STEM intervention research. Specifically, this could be addressed, not from the message that men are as communal as women (this is largely not supported by research), but rather that being successful in STEM requires communal traits and abilities too (e.g., communicating effectively, attention to detail, teamwork, and so on).

Limitations

The small sample size is a limitation, particularly for factor analyses. Thus, findings are considered exploratory and require replication.

Additionally, this research did not include aggressive- or dominant-agentic traits. It is possible that a stronger factor would have emerged with such traits, demonstrating a more robust male-agentic stereotype. These traits were purposefully not included in this research study, as they are not typically associated with gender differences in male-dominant roles or fields, such as STEM. Thus, it may be interesting to investigate what factors emerge when dominance items are included, though we suspect this dimension may not be so useful in gender-STEM intervention research.

Although it is assumed that desirable, not undesirable, agentic traits drive gender bias in STEM, there could be a relationship to explore between the role of undesirable agentic traits (dominance) and backlash. Future research may benefit from administering measures that examine the valence of traits (e.g., EPAQ; Spence et al., 1979). Perhaps the prescription for women's agency related to competence has changed, but agency related to social dominance has not (Spence & Helmreich, 1972; Spence & Buckner, 2000). Therefore, women are allowed to be competent and communal, but cannot be dominant (competitive and directive) and communal (submissive and nice), as social dominance traits are incompatible with communal traits (Rudman & Glick, 2001). The double-bind and backlash effect, then, may be explained by perceived *dominance* – meaning women perceived as competent but not dominant may avoid this (Rudman & Glick, 2001; Eagly et al., 1992; Carli, LaFleur & Loeber, 1995). The need to be perceived as agentic enough (competent) but not too agentic (dominant) would make women's experience in STEM more tenuous than men's.

Lastly, only *explicit* gender-trait stereotypes were assessed. Therefore, social desirability bias may have influenced responses. Additionally, it is very plausible that gender-trait stereotypes function largely at an implicit level. Indeed, given that nearly 75% of men and women hold an implicit bias for men = science, "implicit bias warrants a central focus for research and application" (p. 7228, Charlesworth & Banaji, 2019).

CHAPTER SIX – GENERAL DISCUSSION

The guiding question for this doctoral thesis was, how can we reduce the detrimental effects of gender stereotypes in STEM? To investigate this, I designed a novel psychological intervention integrating debunking and inoculation theory, which has not been previously applied to the context of stereotypes. Additionally, I sought to identify cognitive and affective mechanisms, and their relative importance, involved in improving and protecting gender-STEM stereotype discernment. Using a multi-dimensional approach to gender stereotyping, I examined the effectiveness of the intervention among both the general population and STEM university students. Finally, as interpretations of findings can only be as valid as the methods used to quantify them, I sought to improve scientific understanding and measurement of current gender-trait stereotypes.

Summary of Findings

Overall, the present research suggests that inoculation theory and optimised debunking strategies can be effective interventions (at least in the short term) for improving gender-STEM stereotype discernment, in both the general population and engineering student population. To the best of my knowledge, this is the first research programme to test an inoculation theory intervention in the context of stereotyping.

In Chapter 2, I designed and tested the effectiveness of the novel intervention. I found that the debunking condition demonstrated the greatest discernment of gender stereotypes in STEM. There were no significant differences between the debunking condition and inoculation conditions, but the inoculation condition did not differ from the misinformation and control conditions to the same extent as the debunking condition. As the two conditions received the same treatment, this can be interpreted in line with most stereotype-reduction research, which finds significant change immediately following an intervention treatment (methodologically similar to the debunking condition) and reduced treatment effects at follow-up (methodologically similar to the inoculation condition). In other words, the inoculation condition may have been representative of the treatment's real-world effects because participants were subsequently exposed to convincing counterinformation before completing the outcome measures, and exposure to counterinformation in the real world is likely to happen after completing the intervention. Conversely, Chapter 2 found no treatment effects on male-agentic stereotype endorsement. In fact, there was very low to no male-

agentic stereotype endorsement across conditions. This finding highlights that gender-trait stereotypes may be changing; and together these findings support a multi-dimensional approach to studying gender stereotyping. This chapter contributes preliminary theoretical and empirical findings to gender stereotype-reduction literature.

The findings from Chapter 2 were replicated in Chapter 3. Once again, there were no intervention effects on male-agentic stereotype endorsement, with an overall very low to no endorsement of the male-agentic stereotype across conditions. And, again, compared to the control and misinformation conditions, the intervention conditions demonstrated greater discernment of gender-STEM stereotypes. However, like in Chapter 2, these findings were more consistent and more pronounced in the debunking condition than the inoculation condition. Exposure to the misinformation video seemed to neutralise the treatment effect in the inoculation condition. Importantly, this does not suggest that the debunking intervention was more effective than the inoculation intervention (the treatments for the two conditions were identical). Instead, it suggests that the *inoculative* effects, or protective effects against misinformation, of the intervention need to be improved, as some of the increased ability to discern is diminished after exposure to convincing misinformation about gender differences in STEM. Further, the intervention effects on gender stereotype *discernment* did not generalise to gender stereotyped *attitudes* (i.e., sexism) or *decision-making* (i.e., applicant evaluations). This is consistent with most diversity trainings, which seem to have more influence on cognitive outcomes than attitudinal or behavioural ones (see Bezrukova et al., 2012; Bezrukova et al., 2016; Killinoski et al., 2013 for meta-analyses). Taken together, these findings from Chapter 3 suggest (1) the limited efficacy of the intervention and (2) the multi-dimensional nature of gender stereotypes and their effects. Finally, Chapter 3 findings support inoculation theory by demonstrating that the inoculated participants demonstrated greater counterarguing and psychological reactance to subsequent gender-STEM misinformation than participants who did not receive the treatment; and inoculated participants' greater gender-STEM stereotype discernment was partially mediated through a dual-process model of counterarguing and psychological reactance to the misinformation. While these cognitive and affective mechanisms held significant explanatory power in discernment, treatment-elicited perspective taking, empathic concern, and logical thinking did not. Perspective taking, empathic concern, and logical thinking are psychological mechanisms commonly theorised to play a role in stereotype- and prejudice-reduction research, but are empirically understudied. Thus, these findings contribute new theoretical and empirical insights to stereotype intervention literature.

Paluck and Green (2009) and Paluck et al. (2021) highlight the need for field interventions, and in Chapter 4 I designed and tested a field intervention for university engineering students. The intervention design followed inoculation theory and optimised debunking and built on the previous studies by measuring pre-intervention beliefs. Although significant, the increase in gender-STEM discernment was small. Beliefs that are intrinsically tied to one's self-concept, like gender, are reluctant to change (Johnson, 1989; Eagly & Chaiken, 1993; Sherman, 2002). However, findings did not demonstrate a backfire effect, in fact, the greatest psychological shifts were among those who held contrary beliefs pre-intervention. Corroborating findings from Chapters 2 and 3, changes in gender-STEM stereotype discernment did not translate to changes in male-agentic stereotype endorsement. Once again, the sample demonstrated very low to no endorsement of the male-agentic stereotype pre-intervention. In further support of findings from Chapter 3, treatment-elicited empathic concern, perspective taking, and logical thinking did not mediate the intervention effects on discernment. Finally, compared to an unconscious bias training, the stereotype training did not perform as well on post-intervention discernment. Altogether, Chapter 4 responds to the call for more research on interventions in the field (Paluck & Green, 2009; Paluck et al., 2021). Additionally, it contributes empirical insights to optimised debunking and therapeutic inoculation interventions. Finally, it adds to the literature on the backfire effect that indicates it is a small to negligible effect even on polarised issues (see Wood & Porter, 2018 for a meta-analysis). However, in real-world diversity training (Kalev et al., 2006) and affirmative action programmes (Bond & Pyle, 1988; Linnehan & Konrad, 1999), researchers have found a *behavioural* backfire effect, suggesting further research is needed to elucidate the potential attitudinal and behavioural backfire effects of optimised debunking and inoculation interventions.

Across Chapters 2 through 4, there were no perceived gender differences in agency in any experimental condition. Previous literature shows that both women and men agree on the importance of agentic traits in STEM (Ramsey, 2017) and propose the male-agentic association underpins the underrepresentation of women in STEM (Heilmen et al., 2001; Eagly & Karau, 2002; Ellemers, 2018; Rudman & Phelan, 2008; Eagly & Wood, 2012; Wood & Eagly, 2010). So, why was there so little endorsement of the male-agentic stereotype in Chapters 2, 3, and 4? Given the poor internal consistency and poor performance of the male-agentic stereotype across the previous chapters, gender-trait stereotypes were explored in Chapter 5. Chapter 5 suggests the lack of male=agentic findings in Chapters 2 through 4 may be related to inappropriate measurement of agency based on findings from the

CFAs. Alternatively, findings from the EFAs suggest the gender-trait stereotype related to the gender-STEM gap may not driven by the requirement of *agency*, but more so the rejection of *communality*. Thus, aiming to enhance and promote women's agentic competence in interventions may be misguided or insufficient. Instead, it may be effective to draw attention to the communal aspects necessary to perform well in STEM (e.g., effective communication, team-work, attention to detail, etc.). Indeed, women's agency has steadily increased since the 1970s, plateauing in line with men's agency (Twenge, 1997; Twenge et al., 2001; Egan & Perry, 2001; Ramsey, 2017; Donnelly & Twenge; 2017; Hentschel et al., 2019). Therefore, Chapter 5 advances gender stereotype literature by employing CFAs and EFAs to update current scientific understanding and measurement of gender-trait stereotypes.

In summary, this thesis presents the first investigation of inoculation theory and optimised debunking as an intervention approach for targeting the effects of gender stereotypes in STEM. The key take-away finding is that debunking and inoculating against gender stereotypes in STEM is potentially promising but insufficient. Gender stereotypes are multi-dimensional (Eisand, 2019), and they operate at multiple levels, including internalised, interpersonal, institutional, and structural (Bearman & Amrhein, 2013). All levels co-exist and are intertwined with each other, and this complexity implores a multi-level, multi-disciplinary treatment strategy. Likewise, this research points to the importance of adopting a multi-dimensional approach to studying gender stereotypes and treatment effects. Finally, this research shows that treatment-elicited counterarguing and psychological reactance to gender stereotypes may underpin resistance to gender stereotypes, but it did not find clear support for treatment-elicited empathic concern, perspective taking, or logical thinking in improved discernment of gender stereotypes. More research is needed to elucidate the psychological mechanisms that underpin stereotype change.

I will now discuss the limitations of the present research and suggested future directions.

Limitations and Future Directions

The first limitation of this thesis I would like to address is discussing gender as a binary construct, both in the analysis of participants and intervention materials. Data, and therefore research conclusions, are largely derived from WEIRD – and cisgender - samples. Gender is far more diverse than male and female, including individuals with differences in sex development (DSD), transgender, non-binary, gender-fluid, and two-spirited individuals

(Factor & Rothblum, 2008; Nanda, 2015; Wilson, 1996). By not including those who don't neatly fit into the categories of male or female, research fails to recognise and represent them (Butler, 2004); thus, producing a problematic bias in scientific research and knowledge (Cameron & Stinson, 2019; Balarajan et al., 2011).

Although gender was measured on a spectrum when collecting demographic data (including options for transgender, non-binary, etc.), too few gender-minority individuals participated to be able to conduct meaningful analyses with the data. Future research should aim to specifically recruit gender minorities - not only in gender stereotype research, but across psychological research. Through the inclusion of only male-identifying and female-identifying participants in analyses, this research contributes to the marginalisation and exclusion of gender minorities. A lack of scientific inclusion indirectly perpetuates the myth of a gender binary and tends to focus on differences between men and women. Gender minorities are likely disproportionately impacted by gender stereotypes given the extent to which their gender identity violates prescriptive and descriptive gender stereotypes. Thus, future research should aim to recruit and represent a gender diverse sample when possible.

Likewise, future gender-stereotype research and interventions should aim to reduce the effects of gender stereotypes against gender minorities. Previous research has found transgender men and women are rated significantly less likeable and hireable than identical cis counterparts (Moss-Racusin & Rabasco, 2018). Including gender categories beyond male and female in intervention material may help bolster gender stereotype discernment, as it would dismantle the foundational framework of gender stereotypes – that male and female are an oppositional binary.

As this research offers some novel approaches and findings, replication is required. Replication is especially needed as the sample sizes were relatively small (Chapter 2, 4 and 5) and the composition of the samples were mostly online (Chapter 2, 3, and 5) and mostly WEIRD participants (Hruschka et al., 2018). Additionally, across studies, the effect sizes were generally small to medium. Although this is common and to be expected in social psychological research, and even more so in stereotype- and prejudice-reduction research, it is important to acknowledge the limited explanatory power of the findings. Thus, replication with more representative samples would be valuable, as well as cross-cultural validation samples.

Due to the nature of the intervention content, along with the self-report measurement method, demand characteristics and social desirability bias may have inflated findings in the treatment groups (Chapter 2, 3 and 4). Additionally, due to the explicit assessment of gender

stereotypes and sexism, social desirability bias may have influenced findings across conditions, not just the treatment group, and therefore reduced true group differences (Chapter 2 and 3). Likewise, the non-intervention assessment of current gender-trait stereotypes (Chapter 5) may also suffer from social desirability bias, which could impact the validity of the factor structure solution.

Many of the study materials used were original to this research (e.g., intervention videos, training presentation, STEM job description, and male/female application form) and would benefit from further improvement and validation (Chapters 2, 3, 4 and 5). Specifically, the quality of the intervention materials could be improved with professional editing of the materials, in particular the intervention videos and presentation slides. Likewise, it would be fruitful to test different video lengths and content. Additionally, some measures used were original to this research and should be corroborated with other scales to confirm reliability and validity of measurement and findings. For example, the index used to measure discernment of gender stereotypes in this research is the first of its kind. While it is not uncommon practice for researchers to design and use their own face-valid measures (e.g., Pietri et al., 2017; Rudman & Glick, 2001; Hentschel et al., 2019; Hill & Augoustinos, 2001; Tompkins et al., 2015), further development and validation is required.

Further, scientifically, there needs to be more cohesion on how we define and measure gender stereotypes in research. The lack of parsimony makes it challenging to integrate evidence and may help to explain the field's contradictory findings. Due to the inconsistent, operationalisation of these variables across research, it is challenging to compare findings, including ours, with other stereotype-reduction interventions. In contrast, hostile sexism, modern sexism, and neosexism were measured with well-established measures. Additionally, the PAQ, one of the most widely used gender-trait measures, was selected as the outcome measure for the agentic-trait dimension of gender stereotypes. However, the PAQ agency scale did not seem to be an accurate measure of current gender-trait stereotypes. In general, it appears most gender-stereotype instruments need to be updated. Among researchers that aim to use a more updated measure of agency, the traits assessed and their respective domains of agency are so varied that it is challenging to make meaningful comparisons and interpretations across findings. Thus, the trait dimension of gender stereotypes in particular is fraught with inconsistency. Overall, improvements and agreement on specifying gender stereotype dimensions, and their instruments, is needed.

Moreover, more research investigating the intra- and inter-relatedness of different dimensions of gender stereotyping is needed. For example, gender stereotypes can play out in

intrapersonal processes, such as identity, attitudes, gender roles, social perceptions, and values. Interpersonally, gender stereotypes can play out in different contexts: at home, at work, in public spaces, in health care, and so on; as well as in different relational dynamics: with friends, family, coworkers, and strangers. It is very plausible that individuals may perpetrate or experience gender stereotyping in one or some of these areas and not others. For example, a woman may experience gender discrimination in her workplace but not at home, or vice versa. Current measures and this thesis do not do a good job of capturing such complexities and therefore may not capture important nuances.

Beliefs, attitudes, and psychological change can be difficult to empirically quantify as they are not observable, objective constructs. Further, the self-report method is subject to social desirability bias, particularly with socially-sensitive topics like gender stereotypes and sexism. Future research may benefit from prioritising and measuring psychological safety, so that participants are able to engage as honestly and openly as possible. This would be helpful for two purposes: increasing confidence in validity of scores and, because psychologically safe environments are more conducive to belief change, improving efficacy of the research programme. Additionally, future research may benefit from task-based methods to avoid self-report biases and inflation effects.

Gender-stereotypic beliefs can shape attitudes and behaviours, but behaviours can also shape attitudes and beliefs (e.g. self-fulfilling prophecy; pygmalion effect). Therefore, intervention research aimed at significantly reducing the effects of gender stereotypes would benefit from novel methods to target beliefs, attitudes, *and* behaviours. Our intervention (Chapter 2 and 3) influenced beliefs only, not attitudes or behaviours. But if the belief change doesn't translate to changes in attitudes or behaviours, can it be considered effective? Some behavioural research suggests interactive game-based interventions may be more effective than passive videos for targeting attitudes (Morewedge et al., 2015; Simonovits et al., 2017; Cundiff et al., 2014). Additionally, behavioural interventions, such as cognitive based compassion training (CBCT) and mindfulness meditation, appear to improve empathy (Mascaro et al., 2013; Block-Lerner et al., 2007; Shapiro & Izett, 2008), likely through effects of neuroplasticity. This could be a helpful technique for increasing empathy and increasing stereotype resistance. Thus, behavioural-based interventions that aim to target attitudes and behaviours, such as game-based and CBCT interventions, are recommended.

Data on pre-intervention beliefs were not collected in the online intervention studies (Chapter 2 and 3), meaning within-subject analyses were not possible. As such, between-condition results were analysed and interpreted. Therefore, it was not possible to evaluate

true intervention effects or psychological *change*, as we do not know the participants' psychological starting point. Although groups did not differ from each other on key demographic variables like level of education, age, political orientation, and gender, or on the range of sexist attitudes measures, it is possible that there were other random group differences on important variables that influenced findings. Indeed, in both online intervention studies (Chapter 2 and 3), participants randomly assigned to the inoculation condition and participants randomly assigned to the debunking condition demonstrated differences on variables that may have confounded findings. While a repeated-measures design was used in the field experiment (Chapter 4), the study did not include a control group, so between-group comparisons of psychological change were not possible. Thus, we cannot be as confident that any change in pre-post values were a result of the training (Cook & Campbell, 1979).

Additionally, the pre-post intervention comparisons in the field experiment (Chapter 4) may suffer from priming, testing, and/or fatigue effects. While additional, distracting questions were included in the pre-intervention measures to minimise priming and anchoring effects, this may have increased the likelihood of fatigue effects. Further, the aim of the intervention may have been understood from the consent form and the intervention, which may have exacerbated priming effects and demand characteristics. Finally, as the pre- and post-measures were collected immediately before and after the intervention, they may have suffered from testing effects.

Likewise, while the use of identical items for pre-post comparison is useful for internal validity and statistical analysis and interpretation, participants may update their understanding or definition of the outcome variable during the training. This phenomenon, known as response shift bias, is of particular concern when the aim of the training is to change participants' understanding of the variable being measured (Robinson & Doueck, 1994). Thus, this potential change in understanding could have biased change in discernment, particularly on two of the DOGSS items that used the term *gender stereotypes* in the statement. Future research may benefit from changing the wording, but not the meaning, of items between the pre and post measures. However, with a change in statements, it would be challenging to know whether the paired items have been interpreted in an identical way. Indeed, previous research has found item effects when different items are used at post-test (Roozenbeek et al., 2021). It would be important to use crossover design to counterbalance the items, but this wouldn't ameliorate the issue. Alternatively, the use of post-then (retrospective pre-tests) comparisons may be more reliable (Howard & Daily, 1979; Howard

et al., 2981). However, validation studies of post-then measurement have looked at more "neutral" topics, such as self-perceived assertiveness. This type of retrospective data may not be suitable for assessing more socially-sensitive topics like gender stereotypes.

A longitudinal design was not used and, therefore, long-term intervention effects were not measured. This would have improved robustness of the findings. However, neither significant nor long-term change was expected with these one-off interventions. Rather, the purpose was to provide insight into successful intervention techniques and the psychological mechanisms associated with them. This scientific knowledge is key to informing future evidence-based interventions, which will hopefully facilitate longitudinal research design and sustained psychological change. That being said, the meaningfulness of the significant results may be questionable, as follow-up assessments were not collected after the completion of the interventions (Chapter 2, 3, and 4). Indeed, as psychological change was measured only immediately after the interventions, our findings may lack ecological validity. The inoculation condition aimed to test the robustness of the intervention by including a misinformation "attack", which is more representative of testing real-world effects, but this would not be the same as the impact of real-world social factors and other influences, including the passage of time.

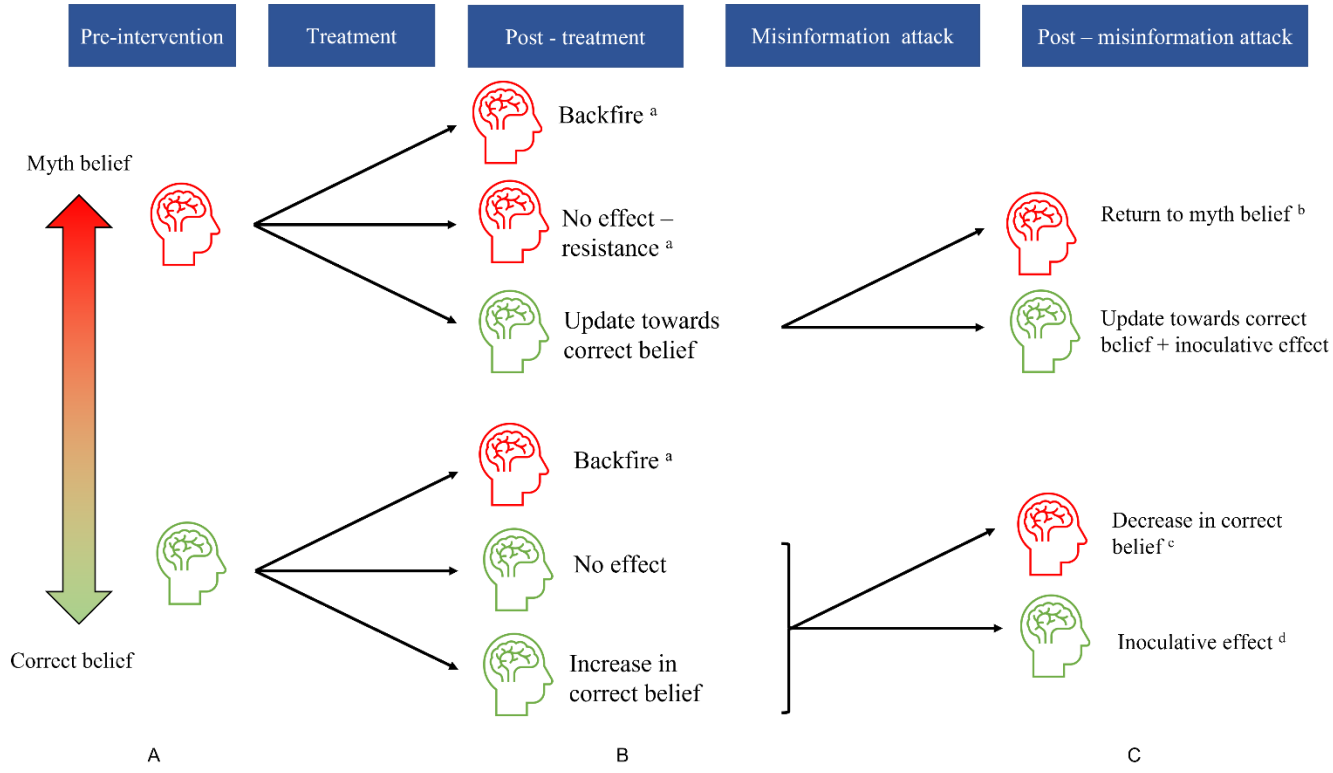
This research did not distinguish between prophylactic and therapeutic inoculation effects on gender-STEM stereotype discernment. As prior beliefs were not measured in Chapter 2 or 3, distinguishing between the two is not possible. While prior beliefs were assessed in Chapter 4, a countermessage attack was not included and therefore theoretical inoculation cannot be tested. Thus, for both theory and intervention development, further research is needed on using inoculation interventions in the context of stereotypes. Future research would benefit from the experimental design shown in Figure 6.1.

While motivation-based threat and counterarguing are evidenced elements for successful prophylactic inoculation, more corroboration on the elements involved in *therapeutic* inoculation are needed. While increasing motivation-based threat may still be relevant, it would be motivation to change one's beliefs – not protect – and therefore effective therapeutic inoculation may require psychological *safety* first. Considering that inoculation intervention materials are generally the same for traditional and therapeutic inoculation, future researchers may want to take a closer look at the balance of evoking motivational threat and psychological safety and if there are differential elements for success between the two types of inoculations. Further, researchers may want to design and test different interventions for those with the "correct" prior belief and those with the "myth"

prior belief, in order to help clarify the psychological processes. Following the inoculation metaphor, and using dynamic systems thinking, perhaps future treatments could be personalised, like precision (personal) medicine, where treatments are tailored to the individual person based on their predicted response to or risk of infection/disease. Indeed, research shows that designing targeted diversity training may improve effectiveness (Sawyerr et al., 2005). Further, personalised feedback seems to increase intervention effects (Morewedge et al., 2015; Carey et al., 2015). Therefore, a 'personalised inoculation' approach could be a fruitful direction to explore.

Figure 6.1

Schematic illustrating suggested research design for future inoculation and optimised debunking interventions



Note. A – B represents optimised debunking intervention. A – C represents inoculation intervention. Assessments to be taken at each point (A, B, C) as well as follow-ups after intervention (not represented in illustration). [a] Individuals whose beliefs in the myth do not change following "treatment", or those for which the treatment backfired (i.e. change from "correct" belief to myth belief or increase in myth belief), post-misinformation attack can experience a further increase in myth belief or maintain levels of myth belief following exposure to the counter message attack. This is not represented for ease of illustration interpretation. [b] Return to myth belief could mean resistance (if degree of belief in the myth is equal to that of pre-intervention) or backfiring (if degree of belief in the myth is higher than that of pre-intervention). [c] "Correct" belief individuals who experience a decrease effect after the misinformation attack may still maintain a "correct" belief - just to a lesser extent - or they may cross over to the "myth" belief. This is an important distinction, but is not represented for ease of illustration interpretation. *d* "Correct" belief individuals may experience either no

effect of the misinformation attack or they may experience an increase in correct belief (parallel to increase in immunity in inoculation metaphor). This is not represented for ease of illustration interpretation.

To help establish psychological safety, in advance of the main intervention, researchers could provide an initial inoculation intervention to target psychological reactance by (1) warning participants that they may experience psychological reactance in the upcoming intervention and (2) explaining why this process happens and providing them with the techniques to not be as affected by it (Compton et al., 2022, in press). As stereotypes are adaptive, core cognitive functions, the aim is to pre-emptively address the psychological reactance to having a core world concept shaken. Thus, inoculating against psychological reactance to the upcoming inoculation messages may help foster psychological shifts among participants who would not be able to otherwise. Another recommended intervention technique to establish psychological safety is self-affirmation. Self-affirmation has been found to be helpful in making individuals more open to debunking messages that promote worldviews that are inconsistent with their own (Paynter et al., 2019). It is theorised that affirming one's values increases a sense of self and therefore resilience to identity threat (Sherman & Cohen, 2002; Nyhan & Reifler, 2018; Bowen et al., 2013). In other words, establishing a sense of safety around one's self-concept reduces the perceived need for defensive biases. Finally, using non-judgmental listening in conversational interventions appears to be an effective technique for establishing psychological safety (Broockman & Kalla, 2017; Kalla & Broockman, 2020).

While cognitive and affective mechanisms were explored, it is critical to acknowledge the role of social mechanisms in gender stereotyping. From social-norm messaging to personal relationships, these social factors are shown to influence stereotyped and prejudiced beliefs, attitudes, and behaviours (Paynter et al., 2019; Hornsey & Fielding, 2017; Duguid & Thomas-Hunt, 2015; Stangor, Sechrist, & Jost, 200). For example, when participants are told that few people hold stereotypical views of others, they engaged in less stereotyping of women (as communal; on three traits: warm, family-oriented, career-oriented [reverse-coded]) than participants who were told the vast majority of people hold stereotypical views of others (Duguid & Thomas-Hunt; 2015). Another idea would be to conduct the training within established social groups and by trusted sources within one's social group to facilitate collective psychological shifts, which would likely result in more sustainable change, given the significant influence of friends, family, and source credibility on beliefs and attitudes (Traberg & van der Linden, 2022; Guillory & Geraci, 2013; Dey, 1996; Mitchell et al., 2015;

Smith & Farkas, 2022). Indeed, when individuals are faced with a persuasive message that is incongruent with their current beliefs or attitudes, they will actively recall others in their social groups (family, friends, coworkers) who share their existing belief or attitude (Zuwerink, 2003). This social element may help explain why so few stereotype-reduction programmes produce long-term effects. Paluck and Green (2009) argue that social norms and conformity – well-known forces for expressions of prejudice in history – could be explored as avenues for inducing acceptance as well. Indeed, prceived normative *appropriateness* of prejudice influences the extent to which people express prejudiced attitudes and behaviours (Duguid & Thomas-Hunt, 2015; Crandall & Stangor, 2005; Stangor et al., 2001; Crandall, Eshleman, & O'Brien, 2002; Brief, Dietz, Cohen, Pugh, & Vaslow, 2000). Thus, for individual-level inoculation to effectively occur and last, there likely needs to be group-level agreement or change, i.e., 'herd immunity'.

Beliefs, attitudes, and psychological change can be difficult to empirically quantify as they are not observable, objective constructs. Further, the self-report method is subject to social desirability bias, particularly with socially-sensitive topics like gender stereotypes and sexism. Future research may benefit from prioritising and measuring psychological safety, so that participants are able to engage as honestly and openly as possible. This would be helpful for two purposes: increasing confidence in validity of scores and, because psychologically safe environments are more conducive to belief change, improving efficacy of the research programme.

Given the significant role implicit stereotypes appear to play in gender-STEM gender disparities (Charlesworth & Banaji, 2021; Charlesworth & Banaji, 2020), it would also be interesting to explore the application of debunking and inoculation interventions to *implicit* stereotypes. It is a vicious cycle: as the target of negative stereotypes in male-dominated arenas, women are unjustly perceived as less competent, which results in fewer women receiving opportunities to enter and advance in these fields. This overrepresentation of men then reinforces the implicit association between men-STEM, which then further perpetuates the implicit system of gatekeeping women from these roles. Given the explicitness of debunking and inoculation messages, it would be interesting to better understand how implicit beliefs and attitudes may (or may not) respond. Aversive racism (Kovel, 1970; Duguid & Thomas-Hunt, 2015) represents the subtle, unintentional racism among people who value egalitarianism and, importantly, believe they are non-prejudiced. Aversive sexism, then, is particularly key to target due to its insidious nature. As it becomes less socially acceptable to explicitly endorse gender stereotypes, social desirability bias may play a more

prominent role in the measurement validity of explicit stereotypes. For example, an implicit male-agentic and female-communal stereotype is related to hiring discrimination against female applicants – but explicit (self-report) is not (Rudman & Glick, 2001). Extensive findings from the Implicit Association Test (IAT; Greenwald et al., 1998) suggest that, among both men and women, STEM faculty and general population, men are more associated with science and career, while women are more associated with the arts and home (e.g., Nosek et al., 2002; Nosek & Smyth, 2011; Charlesworth & Banaji, 2019, 2022; Cai et al., 2016; Nosek et al., 2009; Farrell & McHugh, 2017; Marini & Banaji, 2020). As our implicit beliefs are largely a social imprint of what we are exposed to, implicit stereotypes of male-science/female-arts and male-career/female-family have reduced by 13-19% over the last decade – but they are far from eradicated (Charlesworth & Banaji, 2021). Again, taking a personalised medicine approach, inoculation interventions could be tailored to the individual person based on their psychological immunity and predicted risk of infection.

Similarly, considering the moderating role of intersecting social identities on stereotype endorsement may help to improve design and targeting of interventions. A recent study reported significant differences in gender-stereotyped beliefs based on sex, age and level of education among teachers in Ghana (Osman, 2021). Female teachers, younger teachers and more educated teachers expressed significantly reduced gender-stereotypical views (Osman, 2021). No differences in gender beliefs based on religion or marital status were reported. Hentschel et al. (2019) found the same moderating effect of gender, but no moderating effect of age or education, on stereotype endorsement among a sample of mostly White American adults. Likewise, an international workplace gender-stereotype intervention found no effects on American employees' sexist attitudes, but the subgroup of workers outside of the United States demonstrated significant change in attitude (Chang et al., 2019). Additionally, across all employees, men but not women demonstrated a decrease in sexist attitudes towards women. Therefore, there may be cultural differences in the moderating effects of age and education on gender stereotype endorsement, whereas gender appears to be a salient moderating factor cross-culturally. Taking such individual differences into account by designing targeted diversity training may improve intervention success (Sawyerr et al., 2005).

Like much of psychological intervention research, bias workshops in the workplace are typically isolated, one-off, one-hour trainings. Stereotypes develop and strengthen through repeated association, so it is unlikely that any one-off intervention would be powerful enough to eradicate years of deeply ingrained associations; nevermind replace them

with new, different cognitive patterns strong enough to be maintained, especially in the face of counter information and attitudes. Not surprisingly, then, they do not appear to affect meaningful, long-lasting change (see Paluck & Green, 2021 for meta-analysis and review). This is why social- and cultural-level change is critical for sustainable individual-level psychological change. When individuals are faced with a persuasive message that is incongruent with their current beliefs or attitudes, they will actively recall others in their social groups (family, friends, coworkers) who share their existing belief or attitude (Zuwerink, 2003). This social validation process helps them protect their belief or attitude. Given the salience of gender stereotypes, repeated, long-term interventions may be necessary for both explicit and implicit reduction. From the lens of inoculation theory, these can be understood as "booster shots" to top-up psychological antibodies due to decay effects. Considering the lifetime of being exposed to and learning gender stereotypes, it will likely take consistent, concerted effort, from psychological to institutional to political interventions, to unlearn - and replace - such deeply held beliefs. Thus, longitudinal research is needed, both to include long-term interventions in the study design, as well as to measure long-term effects of the interventions. Likewise, systematically, we need to move away from addressing the effects of gender stereotypes with peripheral, tick-box exercises, like most bias trainings and quotas, and integrate gender diversity into the DNA of schools, workplaces, and other institutions.

Overall, intervening earlier in the lifespan would likely be more successful. There is substantial scope for gender-stereotype reduction at child/school level, when gender schemas and stereotypes take stronger hold. A recent meta-analysis on prejudice interventions found that prejudice interventions are more effective for young people than adults (Hsieh et al., 2022). From an inoculation theory perspective, therapeutic inoculation is found to be most effective during the incubation period (before beliefs fully settle into one's mental model); thus, this developmental period is fertile for impacting stereotype formation and application. Indeed, researchers advise early childhood teachers to make a conscious effort to select books with counter-stereotypical or non-stereotypical characters (Roberts & Hill, 2003), as typical children's books can socialise children into gender roles and sexist attitudes. Further, role modelling reduces stereotype threat among women (McIntyre et al., 2003) and increases women's aspirations (Geis et al., 1984). Role models may help to create an "identity-safe" environment and may be key to breaking the perpetuating cycle of both stereotypes (through exposure to counter-stereotypic information) and stereotype threat.

Additionally, from a systems perspective, gender-stereotype interventions at the school-level would be advantageous not just for children, but for educators too. As the pygmalion effect shows, teachers' expectations of their students directly impact their evaluations of, and engagement with, students, which in turn impacts their students' performance (Rosenthal & Jacobson, 1968; Lee et al., 2015). If young girls are discouraged from STEM subjects, it is less likely they will demonstrate interest and confidence in them later in life. Indeed, primary school teachers' gender-STEM bias has a significant, positive effect on boys' achievement and future enrollment in advanced math courses at high-school level (Lavy & Sand, 2015). Likewise, teachers' with high levels of gender bias favouring boys negatively impacts primary-school girls, resulting in lower achievement and lower future enrollment in high-school math courses. Promisingly, this gender difference in achievement and later enrollment is not found among students with non-biased teachers. These findings provide insight into the detrimental, long-term consequences of children's exposure to gender-STEM stereotypes, as well as the need for and potential of gender-STEM stereotype interventions for educators.

The current research focuses on gender stereotypes in STEM, and the intervention materials are detailed and specific to this domain. However, the general myths, their effects, and the misleading tactics may be generalizable to other male-typed domains - and not just domains, but positions too. For example, historically, females have been stereotyped as poor leaders, explained in part by the agency-communality stereotype (e.g., Eagly & Carli, 2003; Eagly & Karau, 2002). The effect of this stereotype in leadership, like STEM, is reflected in an alarming gender gap: just 5 percent of the globe's largest corporations are led by female CEOs (Fortune, 2017), and pay gap: female CEOs are paid 25% less than their male counterparts (Institute for Women's Policy Research, 2012). Beyond the corporate realm, there is a lack of gender equality in leadership positions across a variety of fields, including law, politics, medicine, and academia (Bureau of Labor Statistics, 2012; Etkowitz & Kemelgor, 2001; National Association for Law Placement, 2011; Nonnemaker, 2000). For example, students rated the exact same online course instructor significantly higher across all 12 performance evaluations when they thought the instructor was male versus female. Even on objective measures, such as how promptly marked assignments were returned, the same instructor received a 4.35 out of 5 when 'male' and a 3.55 out of 5 when 'female' (Macnell, Driscoll, & Hunt, 2014). As such, testing the 'blanket of protection' or 'umbrella' effect of interventions in reducing related but different gender stereotypes would be a valuable next-step in research.

To help achieve 'blanket of protection' conferred by inoculation interventions, it is important for future research to identify a suitable balance between "the more detailed the better" (according to inoculation interventions) and the practicality of widely disseminating evidence-based trainings. As Roozenbeek and van der Linden (2018, 2019) have suggested, focusing on the misleading techniques, rather than the inaccurate content, may help with the umbrella effects of inoculation treatments; therefore, by focusing on misleading *techniques* rather than misleading *content*, a generalised training may be able to shift beliefs, attitudes, and behaviours across multiple gender stereotypes — or even multiple domains of stereotyping (e.g., gender, race, sexuality, religion). For example, within the realm of stereotyping, this approach might look like focusing on how schemes/cognitive associations function, neuroplasticity, or the illusory truth effect. However, careful consideration would need to be given to not homogenise negative effects across social groups and miss important nuances and differences.

Moreover, the interactive effects of multiple social identities should be considered. Most research has been conducted using White targets. Asian women, for example, are perceived as more feminine than White women, and Black women are rated as more masculine than White women (Galinsky et al., 2013). This can mean that Black female leaders do not experience backlash in the same way as has been found for White women (Livingston et al., 2012; Rosette et al., 2016). Conversely, agentic Black men are not afforded the same privileges as White men – they are rated less favourably than agentic White men, because they are racially stereotyped as being aggressive and threatening (Eberhardt et al., 2004, Correll et al., 2002). Thus, an intersectional approach to designing and implementing stereotype-reduction interventions is recommended.

Conclusions

Considerable evidence suggests that gender stereotypes drive gender disparities in STEM. To improve gender diversity in STEM, finding effective ways to improve discernment and reduce endorsement and application of gender stereotypes is critical.

The brain has the ability to change and adapt throughout the lifespan in response to experiences. This, neuroplasticity, was described in the research interventions to debunk the myth that there are inborn differences in STEM aptitude: when we do STEM, our brains become better at STEM. While the concept of neuroplasticity can be used to explain why we might see gender differences in STEM, it can also be used to *treat* these gender stereotypes in STEM. Indeed, by exposing the brain to new (correct) information, it can update and create

new neural connections to integrate this learned content as new knowledge. Importantly, the more repetition, the stronger these new neural connections, i.e., learnings, will be. This means we need consistent, multi-dimensional approaches at individual-, organisational-, structural-, and systems-levels. Relatedly, incorporating behavioural intervention design to support psychological interventions may improve integration and subsequent long-term effectiveness of insights made. Thus, while psychological interventions may help to reduce the effects of gender stereotypes in STEM, the effect of a single intervention appears to be small and limited, often negligible. The findings from this thesis replicate past research, suggesting psychological interventions may be insufficient in addressing the gravity of the issue.

The current research contributes to existing literature by integrating and testing intervention techniques not previously studied in gender-stereotype reduction, examining the relative importance of psychological mechanisms involved in improving stereotype discernment, addressing methodological concerns by using a multi-dimensional approach to gender stereotyping and conducting a field study, and providing a more precise and updated understanding of current gender-trait stereotypes. Moving forward, intersectional research with longitudinal design and a multi-dimensional measurement approach is called for to improve scientific understanding of the effects of gender stereotypes and to inform more robust interventions. Together with diverse and interdisciplinary, multi-level efforts, this will help to create evidence-based change and improve gender diversity in STEM.

References

Abel, M. H., & Meltzer, A. L. (2007). Student ratings of a male and female professors' lecture on sex discrimination in the workforce. *Sex Roles*, *57*(3-4), 173–180. https://doi.org/10.1007/s11199-007-9245-x

Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology*, *85*(4), 768–776. https://doi.org/10.1037/0022-3514.85.4.768

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*(7), 1063–1065. https://doi.org/10.1002/ejsp.574

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01810

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition. *Advances in Experimental Social Psychology*, *50*(50), 195–255. https://doi.org/10.1016/b978-0-12-800284-1.00004-7

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203

Allen, S. (1998). Identity: feminist perspectives on "race", ethnicity and nationality. In N. Charles & H. Hintjens (Eds.), *Gender, Ethnicity and Political Ideologies*. Routledge.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt. *Psychological Science*, *18*(6), 524–530. https://doi.org/10.1111/j.1467-9280.2007.01933.x

An, C., & Pfau, M. (2003). *Efficacy of Inoculation Strategies in Promoting Resistance to Political Attack Messages: A Source Credibility Perspective* [Doctoral dissertation].

Anand, R., & Winters, M.-F. (2008). A retrospective view of corporate diversity training from 1964 to the present. *Academy of Management Learning & Education*, *7*(3), 356–372. https://doi.org/10.5465/amle.2008.34251673

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modelling in practice: a review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Aosved, A. C., Long, P. J., & Voller, E. K. (2009). Measuring sexism, racism, sexual prejudice, ageism, classism, and religious intolerance: the intolerant schema measure. *Journal of Applied Social Psychology*, *39*(10), 2321–2354. https://doi.org/10.1111/j.1559-1816.2009.00528.x

Appelbaum, S. H., Audet, L., & Miller, J. C. (2003). Gender and leadership? Leadership and gender? A journey through the landscape of theories. *Leadership & Organization Development Journal*, *24*(1), 43–51. https://doi.org/10.1108/01437730310457320

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, *2*(2), 81–94. https://doi.org/10.1002/bdm.3960020203

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

ASA-CAP. (2019). *Ban on harmful gender stereotypes in ads comes into force*. Asa.org.uk. https://www.asa.org.uk/news/ban-on-harmful-gender-stereotypes-in-ads-comes-into-force.html

Aube, J., Norcliffe, H., Craig, J.-A., & Koestner, R. (1995). Gender characteristics and adjustment-related outcomes: questioning the masculinity model. *Personality and Social Psychology Bulletin*, *21*(3), 284–295. https://doi.org/10.1177/0146167295213009

Baber, K. M., & Tucker, C. J. (2006). The social roles questionnaire: a new approach to measuring attitudes toward gender. *Sex Roles*, *54*(7-8), 459–467. https://doi.org/10.1007/s11199-006-9018-y

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion.* Rand Mcnally.

Balarajan, M., Gray, M., & Mitchell, M. (2011). *Monitoring equality: developing a gender identity question*. Equality & Human Rights Commission.

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*(3), 281–311. https://doi.org/10.1080/03637751003758193

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244. https://doi.org/10.1037/0022-3514.71.2.230

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Basol, M., Roozenbeek, J., & Van der Linden, S. (2020). Good news about bad news: gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1), 2. https://doi.org/10.5334/joc.91

Bassoff, E. S., & Glass, G. V. (1982). The relationship between sex roles and mental health: a meta-analysis of twenty-six studies. *The Counseling Psychologist*, *10*(4), 105–112. https://doi.org/10.1177/0011000082104019

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin*, *23*(7), 751–758. https://doi.org/10.1177/0146167297237008

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, *55*(1), 19–39. https://doi.org/10.1111/j.1467-6494.1987.tb00426.x

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein, T. R., & Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of*

Personality and Social Psychology, *72*(1), 105–118. https://doi.org/10.1037/0022-3514.72.1.105

Bearman, S., & Amrhein, M. (2013). Girls, women, and internalized sexism. *Internalized Oppression*. https://doi.org/10.1891/9780826199263.0008

Beauvais, C., & Spence, J. T. (1987). Gender, prejudice, and categorization. *Sex Roles*, *16*(1-2), 89–100. https://doi.org/10.1007/bf00302854

Beauvois, J.-L., & Dubois, N. (2009). Lay psychology and the social value of persons. *Social and Personality Psychology Compass*, *3*(6), 1082–1095. https://doi.org/10.1111/j.1751-9004.2009.00225.x

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162. https://doi.org/10.1037/h0036215

Bem, S. L. (1977). On the utility of alternative procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology*, *45*(2), 196–205. https://doi.org/10.1037/0022-006x.45.2.196

Bem, S. L. (1981a). *Bem Sex-Role Inventory: professional manual*. Consulting Psychologists Press.

Bem, S. L. (1981b). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*(4), 354–364. https://doi.org/10.1037/0033-295x.88.4.354

Bem, S. L. (1994). *Lenses of gender: transforming the debate on sexual inequality*. Yale University Press.

Bem, S. L., & Bem, D. J. (1970). Case study of a nonconscious ideology: Training the woman to know her place. In D. J. Bem (Ed.), *Beliefs, attitudes, and human affairs* (pp. 89–99). Brooks/Cole.

Benokraitis, N. V. (Ed.). (1997). *Subtle sexism: Current practice and prospects for change*. Sage Publications, Inc.

Bensley, D. A., & Lilienfeld, S. O. (2017). Psychological misconceptions: recent scientific advances and unresolved issues. *Current Directions in Psychological Science*, *26*(4), 377–382. https://doi.org/10.1177/0963721417699026

Berkowitz, A. D. (2004). An Overview of the Social Norms Approach. In L. Lederman, L. Stewart, F. Goodhart, & L. Laitman (Eds.), *Changing the Culture of College Drinking: A Socially Situated Prevention Campaign, Hampton Press*. http://www.alanberkowitz.com/articles/social_norms.pdf

Bilimoria, D., & Liang, X. (2013). State of knowledge about the workface participation, equity, and inclusion of women in academic science and engineering. In M. Wyer, M. Barbercheck, D. Cookmeyer, O. Ozturk, & M. Wayner (Eds.), *Women, science and technology: A reader in feminist science studies* (pp. 21–50). Routledge.

Blackburn, R., Donnelly, J. P., Logan, C., & Renwick, S. J. D. (2004). Convergent and Discriminative Validity of Interview and Questionnaire Measures of Personality Disorder in Mentally Disordered Offenders: A Multitrait-Multimethod Analysis Using Confirmatory Factor Analysis. *Journal of Personality Disorders*, *18*(2), 129–150. https://doi.org/10.1521/pedi.18.2.129.32779

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, *79*(6), 993–997. https://doi.org/10.1037/0021-9010.79.6.993

Blanchard, F. A., Lilly, T., & Vaughn, L. A. (1991). Reducing the Expression of Racial Prejudice. *Psychological Science*, *2*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1991.tb00108.x

Bohannon, J. (2016). Mechanical Turk upends social sciences. *Science*, *352*(6291), 1263–1264. https://doi.org/10.1126/science.352.6291.1263

Boldizar, J. P. (1991). Assessing sex typing and androgyny in children: The Children's Sex Role Inventory. *Developmental Psychology*, *27*(3), 505–515. https://doi.org/10.1037/0012-1649.27.3.505

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*(2), 135–143. https://doi.org/10.1002/ejsp.744

Brehm, J. W. (1966). *A Theory of Psychological Reactance*. Academic Press.

Brehm, S. S., & Weinraub, M. (1977). Physical barriers and psychological reactance: 2-yr-olds' responses to threats to freedom. *Journal of Personality and Social Psychology*, *35*(11), 830–836. https://doi.org/10.1037/0022-3514.35.11.830

Brewe, E., Sawtelle, V., Kramer, L. H., O'Brien, G. E., Rodriguez, I., & Pamelá, P. (2010). Toward equity through participation in Modeling Instruction in introductory university physics. *Physical Review Special Topics - Physics Education Research*, *6*(1). https://doi.org/10.1103/physrevstper.6.010106

Briñol, P., Petty, R. E., & Tormala, Z. L. (2004). Self-Validation of Cognitive Responses to Advertisements. *Journal of Consumer Research*, *30*(4), 559–573. https://doi.org/10.1086/380289

Bromberger, J. T., & Matthews, K. A. (1996). A "feminine" model of vulnerability to depressive symptoms: A longitudinal investigation of middle-aged women.. *Journal of Personality and Social Psychology*, *70*(3), 591–598. https://doi.org/10.1037/0022-3514.70.3.591

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), 220–224. https://doi.org/10.1126/science.aad9713

Brown, M. M., Thibodeau, R. B., Pierucci, J. M., & Gilpin, A. T. (2017). Supporting the development of empathy: The role of theory of mind and fantasy orientation. *Social Development*, *26*(4), 951–964. https://doi.org/10.1111/sode.12232

Brown, P. (1987). *Politeness: Some Universals in Language Usage*. Cambridge University Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Bruckmüller, S., & Abele, A. E. (2013). The Density of the Big Two. *Social Psychology*, *44*(2), 63–74. https://doi.org/10.1027/1864-9335/a000145

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk. *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Buller, D. B., Burgoon, M., Hall, J. R., Levine, N., Taylor, A. M., & Beach, B. (2000). Long-term effects of language intensity in preventive messages on planned family solar protection. *Health Communication, 12*, 261–275.

Burnett, J. W., Anderson, W. P., & Heppner, P. P. (1995). Gender Roles and Self-Esteem: A Consideration of Environmental Factors. *Journal of Counseling & Development, 73*(3), 323–326. https://doi.org/10.1002/j.1556-6676.1995.tb01757.x

Burns, M. D., Monteith, M. J., & Parker, L. R. (2017). Training away bias: The differential effects of counterstereotype training and self-regulation on stereotype activation and application. *Journal of Experimental Social Psychology, 73*, 97–110. https://doi.org/10.1016/j.jesp.2017.06.003

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity.* Routledge.

Butler, J. (2004). *Undoing gender*. Routledge.

Cai, H., Luo, Y. L. L., Shi, Y., Liu, Y., & Yang, Z. (2016). Male = Science, Female = Humanities. *Social Psychological and Personality Science*, *7*(5), 412–419. https://doi.org/10.1177/1948550615627367

Cameron, J. J., & Stinson, D. A. (2019). Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. *Social and Personality Psychology Compass*, *13*(11). https://doi.org/10.1111/spc3.12506

Cameron, L., Rutland, A., Hossain, R., & Petley, R. (2011). When and why does extended contact work? *Group Processes & Intergroup Relations*, *14*(2), 193–206. https://doi.org/10.1177/1368430210390535

Campbell, B., Schellenberg, E. G., & Senn, C. Y. (1997). Evaluating Measures of Contemporary Sexism. *Psychology of Women Quarterly, 21*, 89–102.

Canetto, S. S., Kaminski, P. L., & Felicio, D. M. (1995). Typical and Optimal Aging in Women and Men: Is There a Double Standard? *The International Journal of Aging and Human Development*, *40*(3), 187–207. https://doi.org/10.2190/rx0u-t56b-1g0f-266u

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, *35*(1), 38–48. https://doi.org/10.1037/0022-3514.35.1.38

Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes About Gender and Science. *Psychology of Women Quarterly*, *40*(2), 244–260. https://doi.org/10.1177/0361684315622645

Carli, L. L., LaFleur, S. J., & Loeber, C. C. (1995). Nonverbal behavior, gender, and influence. *Journal of Personality and Social Psychology*, *68*(6), 1030–1041. https://doi.org/10.1037/0022-3514.68.6.1030

Carnes, M., Devine, P., Baier Manwell, L., Sheridan, J., Ford, C., Byars-Winston, A., Isaac, C., Fine, E., & Nelson Savoy, J. (2015). *Breaking the Bias Habit: A Workshop to Promote Gender Equity*. WISELI & the University of Wisconsin-Madison.

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The Primacy of Agency Over Competence in Status Perception. *Social Psychology*, *45*(5), 347–356. https://doi.org/10.1027/1864-9335/a000176

Cejka, M. A., & Eagly, A. H. (1999). Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment. *Personality and Social Psychology Bulletin*, *25*(4), 413–423. https://doi.org/10.1177/0146167299025004002

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, *28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Chapman, B. P., Duberstein, P. R., Sörensen, S., & Lyness, J. M. (2007). Gender differences in Five Factor Model personality traits in an elderly cohort. *Personality and Individual Differences*, *43*(6), 1594–1603. https://doi.org/10.1016/j.paid.2007.04.028

Charlesworth, T. E. S., & Banaji, M. R. (2019). Gender in Science, Technology, Engineering, and Mathematics: Issues, Causes, Solutions. *Journal of Neuroscience*, *39*(37), 7228–7243. https://doi.org/10.1523/JNEUROSCI.0475-18.2019

Charlesworth, T. E. S., & Banaji, M. R. (2021a). Patterns of Implicit and Explicit Stereotypes III: Long-Term Change in Gender Stereotypes. *Social Psychological and Personality Science*, 194855062098842. https://doi.org/10.1177/1948550620988425

Charlesworth, T. E. S., & Banaji, M. R. (2021b). Patterns of implicit and explicit attitudes II. Long-term change and stability, regardless of group membership. *American Psychologist*, *76*(6), 851–869. https://doi.org/10.1037/amp0000810

Cheng, Y., Chen, C., Lin, C.-P., Chou, K.-H., & Decety, J. (2010). Love hurts: An fMRI study. *NeuroImage*, *51*(2), 923–929. https://doi.org/10.1016/j.neuroimage.2010.02.047

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00049

Choi, N. (2004). Sex Role Group Differences in Specific, Academic, and General Self-Efficacy. *The Journal of Psychology*, *138*(2), 149–159. https://doi.org/10.3200/jrlp.138.2.149-159

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews*, *46*(4), 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001

Chung, A. H., & Slater, M. D. (2013). Reducing Stigma and Out-Group Distinctions Through Perspective-Taking in Narratives. *Journal of Communication*, n/a-n/a. https://doi.org/10.1111/jcom.12050

Coffé, H. (2018). Gender, gendered personality traits and radical right populist voting. *Politics*, 026339571774547. https://doi.org/10.1177/0263395717745476

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap. *Science*, *324*(5925), 400–403. https://doi.org/10.1126/science.1170769

Compton, J. (2013). Inoculation theory. In J. P. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 220–236). Sage Publications, Inc.

Compton, J., & Ivanov, B. (2013). Vaccinating Voters: Surveying Political Campaign Inoculation Scholarship. *Annals of the International Communication Association*, *37*(1), 251–283. https://doi.org/10.1080/23808985.2013.11679152

Compton, J., Jackson, B., & Dimmock, J. A. (2016). Persuading Others to Avoid Persuasion: Inoculation Theory and Resistant Health Attitudes. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00122

Compton, J., & Pfau, M. (2004). Use of inoculation to foster resistance to credit card marketing targeting college students. *Journal of Applied Communication Research*, *32*(4), 343–364. https://doi.org/10.1080/0090988042000276014

Compton, J., & Pfau, M. (2009). Spreading Inoculation: Inoculation, Resistance to Influence, and Word-of-Mouth Communication. *Communication Theory*, *19*(1), 9–28. https://doi.org/10.1111/j.1468-2885.2008.01330.x

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum Associates, Inc.

Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, *71*(1), 25–38. https://doi.org/10.1037/0022-3514.71.1.25

Conway, M., & Vartanian, L. R. (2000). A status account of gender stereotypes: Beyond communality and agency. *Sex Roles*, *43*(3/4), 181–199. https://doi.org/10.1023/a:1007076813819

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, *12*(5), e0175799. https://doi.org/10.1371/journal.pone.0175799

Cook, R., & Cusack, S. (2011). *Gender stereotyping - transnational legal perspectives.* University Of Pennsylvania Pre.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Houghton Mifflin Company.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. https://doi.org/10.1037/0022-3514.81.2.322

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7).

Cota, A. A., & Fekken, G. C. (1988). Dimensionality of the Personal Attributes Questionnaire: An empirical replication. *Journal of Social Behavior & Personality*, *3*(1), 135–140.

Coury, S., Huang, J., Kumar, A., Prince, S., Krivkovich, A., & Yee, L. (2020, September 30). *Women in the Workplace | McKinsey*. Www.mckinsey.com. https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*(4), 608–630.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2004). When professionals become Mothers, Warmth Doesn't Cut the Ice. *Journal of Social Issues*, *60*(4), 701–718. https://doi.org/10.1111/j.0022-4537.2004.00381.x

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology*, *40*, 61–149.

Cuddy, A. J. C., Wolf, E. B., Glick, P., Crotty, S., Chong, J., & Norton, M. I. (2015). Men as cultural ideals: Cultural values moderate gender stereotype content. *Journal of Personality and Social Psychology*, *109*(4), 622–635. https://doi.org/10.1037/pspi0000027

Cuff, B., Brown, S. J., Taylor, L., & Howat, D. (2015). Empathy: A review of the concept. *Emotion Review*, *8*(2). https://doi.org/10.1177/1754073914558466

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104. https://doi.org/10.1016/j.tics.2006.12.005

Curran, M. D., & Warber, K. M. (2011). Examining the Factor Structure of the Personality Attributes Questionnaire. *Communication Research Reports*, *28*(1), 86–96. https://doi.org/10.1080/08824096.2011.541366

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

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the Air: Identity Safety Moderates the Effects of Stereotype Threat on Women's Leadership Aspirations. *Journal of Personality and Social Psychology*, *88*(2), 276–287. https://doi.org/10.1037/0022-3514.88.2.276

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441. https://doi.org/10.1016/j.tics.2006.08.008

Deaux, K., & Kite, M. E. (1987). Thinking about gender. In B. B. Hess & M. M. Ferree (Eds.), *Analyzing gender: A handbook of social science research*. Sage.

Deaux, K., Kite, M. E., & Lewis, L. L. (1985). Clustering and Gender Schemata. *Personality and Social Psychology Bulletin*, *11*(4), 387–397. https://doi.org/10.1177/0146167285114005

Decety, J., & Jackson, P. L. (2004). The Functional Architecture of Human Empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. https://doi.org/10.1177/1534582304267187

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

del Rey, P., & Sheppard, S. (1981). Relationship of psychological androgyny in female athletes to self-esteem. *International Journal of Sport Psychology*, *12*(3), 165–175.

Deloitte. (2016). *Women in STEM: how can businesses tackle the gender pay gap?* Deloitte United Kingdom. https://www2.deloitte.com/uk/en/pages/growth/articles/technology-career-pathways-gender-pay-gap.html

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as Dynamic Constructs: Women and Men of the Past, Present, and Future. *Personality and Social Psychology Bulletin*, *26*(10), 1171–1188. https://doi.org/10.1177/0146167200262001

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*(6), 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246

Dillard, J. P., & Shen, L. (2005). On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs*, *72*(2), 144–168. https://doi.org/10.1080/03637750500111815

Dixon, G. N., McKeever, B. W., Holton, A. E., Clarke, C., & Eosco, G. (2015). The Power of a Picture: Overcoming Scientific Misinformation by Communicating Weight-of-Evidence Information with Visual Exemplars. *Journal of Communication*, *65*(4), 639–659. https://doi.org/10.1111/jcom.12159

Dobbin, F., Schrage, D., & Kalev, A. (2015). Rage against the Iron Cage. *American Sociological Review*, *80*(5), 1014–1044. https://doi.org/10.1177/0003122415596416

Donnelly, K., & Twenge, J. M. (2017). Masculine and Feminine Traits on the Bem Sex-Role Inventory, 1993–2012: a Cross-Temporal Meta-Analysis. *Sex Roles*, *76*(9-10), 556–565. https://doi.org/10.1007/s11199-016-0625-y

Dovidio, J. F., ten Vergert, M., Stewart, T. L., Gaertner, S. L., Johnson, J. D., Esses, V. M., Riek, B. M., & Pearson, A. R. (2004). Perspective and Prejudice: Antecedents and Mediating Mechanisms. *Personality and Social Psychology Bulletin*, *30*(12), 1537–1549. https://doi.org/10.1177/0146167204271177

Eagly, A. H. (1987). *Sex differences in social behavior : a social-role interpretation*. Lawrence Erlbaum Associates.

Eagly, A. H. (2020). Do the social roles that women and men occupy in science allow equal access to publication? *Proceedings of the National Academy of Sciences*, *117*(11), 5553–5555. https://doi.org/10.1073/pnas.2001684117

Eagly, A. H., & Chaiken, S. (2007). The Advantages of an Inclusive Definition of Attitude. *Social Cognition*, *25*(5), 582–602. https://doi.org/10.1521/soco.2007.25.5.582

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295x.109.3.573

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, *111*(1), 3–22. https://doi.org/10.1037/0033-2909.111.1.3

Eagly, A. H., & Mladinic, A. (1989). Gender Stereotypes and Attitudes Toward Women and Men. *Personality and Social Psychology Bulletin*, *15*(4), 543–558. https://doi.org/10.1177/0146167289154008

Eagly, A. H., & Mladinic, A. (1994). Are People Prejudiced Against Women? Some Answers From Research on Attitudes, Gender Stereotypes, and Judgments of Competence. *European Review of Social Psychology*, *5*(1), 1–35. https://doi.org/10.1080/14792779543000002

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2019). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*(4), 735–754. https://doi.org/10.1037/0022-3514.46.4.735

Eagly, A. H., & Wood, W. (1991). Explaining Sex Differences in Social Behavior: A Meta-Analytic Perspective. *Personality and Social Psychology Bulletin*, *17*(3), 306–315. https://doi.org/10.1177/0146167291173011

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*(6), 408–423. https://doi.org/10.1037/0003-066x.54.6.408

Eagly, A. H., & Wood, W. (2012). Social Role Theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (Vol. 2, pp. 458–476). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n49

Eagly, A. H., & Wood, W. (2013). The Nature–Nurture Debates. *Perspectives on Psychological Science*, *8*(3), 340–357. https://doi.org/10.1177/1745691613484767

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). Lawrence Erlbaum Associates.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, Crime, and Visual Processing. *Journal of Personality and Social Psychology*, *87*(6), 876–893. https://doi.org/10.1037/0022-3514.87.6.876

Echabe, A. E. (2010). Role identities versus social identities: Masculinity, femininity, instrumentality and communality. *Asian Journal of Social Psychology*, *13*(1), 30–43. https://doi.org/10.1111/j.1467-839x.2010.01298.x

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185–192. https://doi.org/10.1016/j.jarmac.2017.01.014

Ecker, U. K. H., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, *18*(3), 570–578. https://doi.org/10.3758/s13423-011-0065-1

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, *37*(4), 451–463. https://doi.org/10.1037/0012-1649.37.4.451

Eisend, M. (2019). Gender Roles. *Journal of Advertising*, *48*(1), 1–9. https://doi.org/10.1080/00913367.2019.1566103

Eisinga, R., Grotenhuis, M. te, & Pelzer, B. (2012). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Ellemers, N. (2018). Gender Stereotypes. *Annual Review of Psychology*, *69*(1), 275–298. https://doi.org/10.1146/annurev-psych-122216-011719

Entringer, T. M., Gebauer, J. E., & Paulhus, D. L. (2021). Extracting Agency and Communion From the Big Five: A Four-Way Competition. *Assessment*, 107319112110039. https://doi.org/10.1177/10731911211003978

European Commission. (2020). Striving for a Union of Equality. The Gender Equality Strategy 2020-2025. In *https://ec.europa.eu/info/sites/default/files/aid_development_cooperation_fundamental_rights/gender_equality_strategy_factsheet_en.pdf*.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989x.4.3.272

Factor, R., & Rothblum, E. (2008). Exploring gender identity and community among three groups of transgender individuals in the United States: MTFs, FTMs, and genderqueers. *Health Sociology Review*, *17*(3), 235–253. https://doi.org/10.5172/hesr.451.17.3.235

Farrell, L., & McHugh, L. (2017). Examining gender-STEM bias among STEM and non-STEM students using the Implicit Relational Assessment Procedure (IRAP). *Journal of Contextual Behavioral Science*, *6*(1), 80–90. https://doi.org/10.1016/j.jcbs.2017.02.001

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Festinger, L. A. (1957). *Theory of cognitive dissonance*. Stanford University Press.

Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications.

Fischer, A. H. (Ed.). (2000). *Gender and Emotion*. Cambridge University Press. https://doi.org/10.1017/cbo9780511628191

Fishbein, M. (1967). Attitude and prediction of behavior. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. Wiley.

Fisher, F. L. (1968). Influences of Reading and Discussion on the Attitudes of Fifth Graders Toward American Indians. *The Journal of Educational Research*, *62*(3), 130–134. https://doi.org/10.1080/00220671.1968.10883788

Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. A. Uleman & J. S. Bargh (Eds.), *Unintended thought* (pp. 253–283). The Guilford Press.

Fiske, S. T. (2017). Prejudices in Cultural Contexts: Shared Stereotypes (Gender, Age) Versus Variable Stereotypes (Race, Ethnicity, Religion). *Perspectives on Psychological Science*, *12*(5), 791–799. https://doi.org/10.1177/1745691617708204

Fiske, S. T., Bersoff, D. N., Borgida, E., Deaux, K., & et al. (1991). Social science research on trial: Use of sex stereotyping research in Price Waterhouse v. Hopkins. *American Psychologist*, *46*(10), 1049–1060. https://doi.org/10.1037/0003-066x.46.10.1049

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25–44. https://doi.org/10.1016/j.jsp.2014.10.002

Fontana, M.-C., Sidler, A., & Hardmeier, S. (2006). The "New Right" Vote: An Analysis of the Gender Gap in the Vote Choice for the SVP. *Swiss Political Science Review*, *12*(4), 243–271. https://doi.org/10.1002/j.1662-6370.2006.tb00067.x

Ford, R., & Goodwin, M. J. (2010). Angry White Men: Individual and Contextual Predictors of Support for the British National Party. *Political Studies*, *58*(1), 1–25. https://doi.org/10.1111/j.1467-9248.2009.00829.x

Forscher, P. S., & Devine, P. G. (2016). The role of intentions in conceptions of prejudice: An historical perspective. *Handbook of Prejudice, Stereotyping, and Discrimination, 2*.

Foushee, H. C., Helmreich, R. L., & Spence, J. T. (1979). Implicit theories of masculinity and femininity: dualistic or bipolar? *Psychology of Women Quarterly, 3*, 259–269.

Frable, D. E. (1989). Sex typing and gender ideology: Two facets of the individual's gender psychology that go together. *Journal of Personality and Social Psychology*, *56*(1), 95–108. https://doi.org/10.1037/0022-3514.56.1.95

Frable, D. E. S., & Bem, S. L. (1985). If you are gender schematic, all members of the opposite sex look alike. *Journal of Personality and Social Psychology*, *49*(2), 459–468. https://doi.org/10.1037/0022-3514.49.2.459

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–228. https://doi.org/10.1037/0022-3514.57.2.212

Gaa, J. P., Liberman, D., & Edwards, T. A. (1979). A comparative factor analysis of the bem sex role inventory and the personality attributes questionnaire. *Journal of Clinical Psychology*, *35*(3), 592–598. https://doi.org/3.0.co;2-w">10.1002/1097-4679(197907)35:3<592::aid-jclp2270350322>3.0.co;2-w

Gagne, J. R., Miller, M. M., & Goldsmith, H. H. (2013). Early—but modest—gender differences in focal aspects of childhood temperament. *Personality and Individual Differences*, *55*(2), 95–100. https://doi.org/10.1016/j.paid.2013.02.006

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. C. (2013). Gendered Races. *Psychological Science*, *24*(4), 498–506. https://doi.org/10.1177/0956797612457783

Galinsky, A. D., Ku, G., & Wang, C. S. (2005). Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination. *Group Processes & Intergroup Relations*, *8*(2), 109–124. https://doi.org/10.1177/1368430205051060

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*(4), 708–724. https://doi.org/10.1037/0022-3514.78.4.708

Ganley, C. M., George, C. E., Cimpian, J. R., & Makowski, M. B. (2017). Gender Equity in College Majors: Looking Beyond the STEM/Non-STEM Dichotomy for Answers Regarding Female Participation. *American Educational Research Journal*, *55*(3), 453–487. https://doi.org/10.3102/0002831217740221

García-Sánchez, R., Almendros, C., Aramayona, B., Martín, M. J., Soria-Oliver, M., López, J. S., & Martínez, J. M. (2019). Are Sexist Attitudes and Gender Stereotypes Linked? A Critical Feminist Approach With a Spanish Sample. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02410

Gardner, P. L. (1995). Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education*, *25*(3), 283–289. https://doi.org/10.1007/bf02357402

Gardner, R. C. (1973). Ethnic stereotypes: The traditional approach, a new look. *Canadian Psychologist/Psychologie Canadienne*, *14*(2), 133–148. https://doi.org/10.1037/h0082215

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gebauer, J. E., Sedikides, C., Verplanken, B., & Maio, G. R. (2012). Communal narcissism. *Journal of Personality and Social Psychology*, *103*(5), 854–878. https://doi.org/10.1037/a0029629

Gehlbach, H., & Brinkworth, M. E. (2012). The Social Perspective Taking Process: Strategies and Sources of Evidence in Taking Another's Perspective. *Teachers College Record: The Voice of Scholarship in Education*, *114*(1), 1–29. https://doi.org/10.1177/016146811211400107

Geis, F. L., Brown, V., Jennings (Walstedt), J., & Porter, N. (1984). TV commercials as achievement scripts for women. *Sex Roles*, *10*(7-8). https://doi.org/10.1007/bf00287260

Gigerenzer, G. (1984). External Validity of Laboratory Experiments: The Frequency-Validity Relationship. *The American Journal of Psychology*, *97*(2), 185. https://doi.org/10.2307/1422594

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

Glick, P., & Fiske, S. T. (1997). Hostile and Benevolent Sexism. *Psychology of Women Quarterly*, *21*(1), 119–135. https://doi.org/10.1111/j.1471-6402.1997.tb00104.x

Glick, P., & Fiske, S. T. (1999). Sexism and Other "Isms": Interdependence, Status, and the Ambivalent Content of Stereotypes. In W. B. Swan, J. H. Langlois, & L. A. Gilbert (Eds.), *Sexism and Stereotypes in Modern Society*. American Psychological Association.

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

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

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

Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, *55*(2), 178–186. https://doi.org/10.1037/0022-3514.55.2.178

Godbold, L. C., & Pfau, M. (2000). Conferring Resistance to Peer Pressure Among Adolescents. *Communication Research*, *27*(4), 411–437. https://doi.org/10.1177/009365000027004001

Grau, S. L., & Zotos, Y. C. (2016). Gender stereotypes in advertising: a review of current research. *International Journal of Advertising*, *35*(5), 761–770. https://doi.org/10.1080/02650487.2016.1203556

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

Guilford, J. P. (1954). *Psychometric Methods*. McGraw-Hill.

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, *2*(4), 201–209. https://doi.org/10.1016/j.jarmac.2013.10.001

Gygax, P. M., Garnham, A., & Doehren, S. (2016). What Do True Gender Ratios and Stereotype Norms Really Tell Us? *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01036

Haimo, S., & Blitman, F. (1985). The Effects of Assertive Training on Sex Role Concept in Female Agoraphobics. *Women & Therapy*, *4*(2), 53–61. https://doi.org/10.1300/j015v04n02_08

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

Hair, J. F., Black, B., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis: Global Edition* (7th ed.). Pearson Education.

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, *77*(2), 350–359. https://doi.org/10.1037/0022-3514.77.2.350

Hall, J. A., & Halberstadt, A. G. (1980). Masculinity and femininity in children: Development of the Children's Personal Attributes Questionnaire. *Developmental Psychology*, *16*(4), 270–280. https://doi.org/10.1037/0012-1649.16.4.270

Handley, I. M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences*, *112*(43), 13201–13206. https://doi.org/10.1073/pnas.1510649112

Hansen, F. (2003). *Diversity's Business Case Doesn't Add Up | Workforce.com>*. Workforce.com. https://workforce.com/uk/news/diversitys-business-case-doesnt-add-up

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/s0022-5371(77)80012-1

Hayes, E.-R., & Swim, J. K. (2013). African, Asian, Latina/o, and European Americans' Responses to Popular Measures of Sexist Beliefs. *Psychology of Women Quarterly*, *37*(2), 155–166. https://doi.org/10.1177/0361684313480044

Heilman, M. E. (1983). Sex bias in work settings: The Lack of Fit model. *Research in Organizational Behavior*, *5*, 269–298.

Heilman, M. E. (1984). Information as a deterrent against sex discrimination: The effects of applicant sex and information type on preliminary employment decisions. *Organizational Behavior and Human Performance*, *33*(2), 174–186. https://doi.org/10.1016/0030-5073(84)90019-9

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135. https://doi.org/10.1016/j.riob.2012.11.003

Heilman, M. E., & Chen, J. J. (2005). Same Behavior, Different Consequences: Reactions to Men's and Women's Altruistic Citizenship Behavior. *Journal of Applied Psychology*, *90*(3), 431–441. https://doi.org/10.1037/0021-9010.90.3.431

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles*, *7*(11), 1097–1108. https://doi.org/10.1007/bf00287587

Hennes, E. P., Pietri, E. S., Moss-Racusin, C. A., Mason, K. A., Dovidio, J. F., Brescoll, V. L., H. Bailey, A., & Handelsman, J. (2018). Increasing the perceived malleability of

gender bias using a modified Video Intervention for Diversity in STEM (VIDS). *Group Processes & Intergroup Relations*, *21*(5), 788–809. https://doi.org/10.1177/1368430218755923

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

Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The Multiple Dimensions of Gender Stereotypes: A Current Look at Men's and Women's Characterizations of Others and Themselves. *Frontiers in Psychology*, *10*(11). https://doi.org/10.3389/fpsyg.2019.00011

Hill, M. E., & Augoustinos, M. (2001). Stereotype change and prejudice reduction: short- and long-term evaluation of a cross-cultural awareness programme. *Journal of Community & Applied Social Psychology*, *11*(4), 243–262. https://doi.org/10.1002/casp.629

Hill, S. A., Fekken, G. C., & Bond, S. L. (2000). Factor structure integrity of the Personal Attributes Questionnaire: An English–French comparison. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *32*(4), 234–242. https://doi.org/10.1037/h0087120

Hines, M. (2011). Gender Development and the Human Brain. *Annual Review of Neuroscience*, *34*(1), 69–88. https://doi.org/10.1146/annurev-neuro-061010-113654

Hines, M. (2015). Gendered development. In *Handbook of child development and developmental science*. Wiley.

Holt, C. L., & Ellis, Jon. B. (1998). Assessing the Current Validity of the Bem Sex-Role Inventory. *Sex Roles*, *39*(11/12), 929–941. https://doi.org/10.1023/a:1018836923919

Hornsey, M. J., & Fielding, K. S. (2017). Attitude roots and Jiu Jitsu persuasion: Understanding and overcoming the motivated rejection of science. *American Psychologist*, *72*(5), 459–473. https://doi.org/10.1037/a0040437

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation*

Modeling: A Multidisciplinary Journal, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hunter, J. E., & Gerbing, D. W. (1982). Unidimensional Measurement, Second Order Factor Analysis, and Causal Models. In B. M. Straw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (pp. 267–299).

Hutcheson, G., & Sofroniou, N. (1999). *he Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models*. Sage Publication. https://doi.org/10.4135/9780857028075

Hyde, J. S. (2014). Gender Similarities and Differences. *Annual Review of Psychology*, *65*(1), 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

International Labour Organization. (2020). *Empowering Women at Work*.

Isaac, C., Lee, B., & Carnes, M. (2009). Interventions That Affect Gender Bias in Hiring: A Systematic Review. *Academic Medicine*, *84*(10), 1440–1446. https://doi.org/10.1097/acm.0b013e3181b6ba00

Ishii, K., & Kitayama, S. (2010). Outgroup homogeneity effect in perception: An exploration with Ebbinghaus illusion. *Asian Journal of Social Psychology*, *14*(2), 159–163. https://doi.org/10.1111/j.1467-839x.2010.01339.x

Ivanov, B., Burns, W. J., Sellnow, T. L., Petrun Sayers, E. L., Veil, S. R., & Mayorga, M. W. (2016). Using an inoculation message approach to promote public confidence in protective agencies. *Journal of Applied Communication Research*, *44*(4), 381–398. https://doi.org/10.1080/00909882.2016.1225165

Ivanov, B., Parker, K. A., & Compton, J. (2011). The potential of inoculation in reducing post-purchase dissonance: Reinforcement of purchase behavior. *Central Business Review*, *30*, 10–16.

Ivanov, B., Pfau, M., & Parker, K. A. (2009). Can Inoculation Withstand Multiple Attacks? *Communication Research*, *36*(5), 655–676. https://doi.org/10.1177/0093650209338909

Ivanov, B., Sims, J. D., Compton, J., Miller, C. H., Parker, K. A., Parker, J. L., Harrison, K. J., & Averbeck, J. M. (2015). The General Content of Postinoculation Talk: Recalled

Issue-Specific Conversations Following Inoculation Treatments. *Western Journal of Communication*, *79*(2), 218–238. https://doi.org/10.1080/10570314.2014.943423

Jackman, M. R. (1994). *he Velvet Glove: Paternalism and Conflict in Gender, Class, and Race Relations*. University of California Press.

Jackson, S. M., Hillard, A. L., & Schneider, T. R. (2014). Using implicit bias training to improve attitudes toward women in STEM. *Social Psychology of Education*, *17*(3), 419–438. https://doi.org/10.1007/s11218-014-9259-5

Jaxon, J., Lei, R. F., Shachnai, R., Chestnut, E. K., & Cimpian, A. (2019). The Acquisition of Gender Stereotypes about Intellectual Ability: Intersections with Race. *Journal of Social Issues*, *75*(4), 1192–1215. https://doi.org/10.1111/josi.12352

Johnson, D. R. (2013). Transportation into literary fiction reduces prejudice against and increases empathy for Arab-Muslims. *Scientific Study of Literature*, *3*(1), 77–92. https://doi.org/10.1075/ssol.3.1.08joh

Johnson, D. W. (1975). Cooperativeness and social perspective taking. *Journal of Personality and Social Psychology*, *31*(2), 241–244. https://doi.org/10.1037/h0076285

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

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

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*(6), 899–913. https://doi.org/10.1037/0022-3514.89.6.899

Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology*, *54*(5), 778–788. https://doi.org/10.1037/0022-3514.54.5.778

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/bf02291575

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kalla, J. L., & Broockman, D. E. (2020). Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments. *American Political Science Review*, *114*(2), 410–425. https://doi.org/10.1017/s0003055419000923

Kanske, P. (2018). The social mind: disentangling affective and cognitive routes to understanding others. *Interdisciplinary Science Reviews*, *43*(2), 115–124. https://doi.org/10.1080/03080188.2018.1453243

Kant, I. (1929). *Critique of Pure Reason (1781-1787)*. Trans. Kemp Smith.

Katz, R. S. (2000). The United States of America. *European Journal of Political Research*, *38*(3-4), 552–556. https://doi.org/10.1111/j.1475-6765.2000.tb01165.x

Kaufman, D. Q., Stasson, M. F., & Hart, J. W. (1999). Are the Tabloids Always Wrong or Is That Just What We Think? Need for Cognition and Perceptions of Articles in Print Media. *Journal of Applied Social Psychology*, *29*(9), 1984–2000. https://doi.org/10.1111/j.1559-1816.1999.tb00160.x

Kaufman, E. (2021). RBG and Gender Discrimination. *Touro Law Review*, *37*(2). https://digitalcommons.tourolaw.edu/cgi/viewcontent.cgi?article=3311&context=lawreview

Kay, A. C., & Jost, J. T. (2003). Complementary Justice: Effects of "Poor but Happy" and "Poor but Honest" Stereotype Exemplars on System Justification and Implicit Activation of the Justice Motive.. *Journal of Personality and Social Psychology*, *85*(5), 823–837. https://doi.org/10.1037/0022-3514.85.5.823

Kelly, C. (1989). Political identity and perceived intragroup homogeneity. *British Journal of Social Psychology*, *28*(3), 239–250. https://doi.org/10.1111/j.2044-8309.1989.tb00866.x

Kervyn, N., Yzerbyt, V., & Judd, C. M. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two fundamental dimensions of social perception. *European Review of Social Psychology*, *21*(1), 155–187. https://doi.org/10.1080/13546805.2010.517997

Khan, D., & Katz, R. (1978). *The social psychology of organizations*. Wiley.

Kite, M. E. (2001). Changing times, changing gender roles: Who do we want women and men to be? In R. E. Unger (Ed.), *Handbook of the psychology of women and gender* (pp. 215–227). John Wiley & Sons, Inc.

Kite, M. E., Deaux, K., & Haines, E. L. (2008). Psychology of women: Handbook of issues and theories. In F. Denmark & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories*(pp. 205–236). Praeger Publishers/Greenwood Publishing Group.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. The Guilford Press.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York Guilford.

Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01086

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371–392. https://doi.org/10.1037/a0037215

Kowalski, P., & Taylor, A. K. (2009). The Effect of Refuting Misconceptions in the Introductory Psychology Class. *Teaching of Psychology*, *36*(3), 153–159. https://doi.org/10.1080/00986280902959986

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Krueger, J. I., Hasman, J. F., Acevedo, M., & Villano, P. (2003). Perceptions of Trait Typicality in Gender Stereotypes: Examining the Role of Attribution and Categorization Processes. *Personality and Social Psychology Bulletin*, *29*(1), 108–116. https://doi.org/10.1177/0146167202238376

Kurt, B. M., & Buckley, W. (1970). *Class and Society*. Random House.

LaFrance, M., & Banaji, M. (1992). Toward a reconsideration of the gender–emotion relationship. In M. S. Clark (Ed.), *Emotion and social behavior* (pp. 178–201). Sage Publications.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., & Hu, X. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001–1016. https://doi.org/10.1037/xge0000179

Landrine, H. (1985). Race x class stereotypes of women. *Sex Roles*, *13*(1-2), 65–75. https://doi.org/10.1007/bf00287461

Latrofa, M., Vaes, J., & Cadinu, M. (2012). Self-stereotyping: the central role of an ingroup threatening identity. *The Journal of Social Psychology*, *152*(1), 92–111. https://doi.org/10.1080/00224545.2011.565382

Lavy, V., & Sand, E. (2015). On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases. *National Bureau of Economic Research*. https://doi.org/10.3386/w20909

Leaper, C., & Bigler, R. (2011). *Social development : relationships in infancy, childhood, and adolescence* (M. K. Underwood & L. H. Rosen, Eds.; pp. 289–315). The Guilford Press.

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. *Advances in Experimental Social Psychology Volume 32*, *32*, 1–62. https://doi.org/10.1016/s0065-2601(00)80003-9

Lee, S. W., Min, S., & Mamerow, G. P. (2015). Pygmalion in the Classroom and the Home: Expectation's Role in the Pipeline to STEMM. *Teachers College Record: The Voice of Scholarship in Education*, *117*(9), 1–40. https://doi.org/10.1177/016146811511700907

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of mechanical Turk samples. *SAGE Open*, *6*(1), 215824401663643. https://doi.org/10.1177/2158244016636433

Levine, T., Hullett, C. R., Turner, M. M., & Lapinski, M. K. (2006). The Desirability of Using Confirmatory Factor Analysis on Published Scales. *Communication Research Reports*, *23*(4), 309–314. https://doi.org/10.1080/08824090600962698

Levy, S. R., & Dweck, C. S. (1999). The Impact of Children's Static versus Dynamic Conceptions of People on Stereotype Formation. *Child Development*, *70*(5), 1163–1180. https://doi.org/10.1111/1467-8624.00085

Levy, S. R., West, T. L., Ramirez, L. F., & Pachankis, J. E. (2004). Racial and ethnic prejudice among children. In *The psychology of prejudice and discrimination: Racism in America* (pp. 37–60). Praeger.

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., & Vraga, E. K. (2020). *The Debunking Handbook 2020*. https://www.climatechangecommunication.org/wp-content/uploads/2020/10/DebunkingHandbook2020.pdf

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lieberman, J. D., & Arndt, J. (2000). Understanding the limits of limiting instructions: Social psychological explanations for the failures of instructions to disregard pretrial

publicity and other inadmissible evidence. *Psychology, Public Policy, and Law*, *6*(3), 677–711. https://doi.org/10.1037/1076-8971.6.3.677

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, *57*(2), 165–188. https://doi.org/10.1037/0022-3514.57.2.165

Lips, H. M. (1991). *Women, men, and power*. Mayfield.

Lishner, D. A., Stocks, E. L., & Steinert, S. W. (2017). Empathy. *Encyclopedia of Personality and Individual Differences*, 1–8. https://doi.org/10.1007/978-3-319-28099-8_513-1

Livingston, J. D., Milne, T., Fang, M. L., & Amari, E. (2012). The effectiveness of interventions for reducing stigma related to substance use disorders: a systematic review. *Addiction*, *107*(1), 39–50. https://doi.org/10.1111/j.1360-0443.2011.03601.x

Livingston, R. (2013). *Gender, race and leadership: An Examination of the Challenges Facing Non-prototypical Leaders*. Harvard Business School. https://www.hbs.edu/faculty/Shared%20Documents/conferences/2013-w50-research-symposium/livingston.pdf

Livingston, R. W., Rosette, A. S., & Washington, E. F. (2012). Can an Agentic Black Woman Get Ahead? The Impact of Race and Interpersonal Dominance on Perceptions of Female Leaders. *Psychological Science*, *23*(4), 354–358. https://doi.org/10.1177/0956797611428079

Löffler, C. S., & Greitemeyer, T. (2021). Are women the more empathetic gender? The effects of gender role expectations. *Current Psychology*. https://doi.org/10.1007/s12144-020-01260-8

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of*

Personality and Social Psychology, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Lubinski, D., Tellegen, A., & Butcher, J. N. (1983). Masculinity, femininity, and androgyny viewed and assessed as distinct concepts. *Journal of Personality and Social Psychology, 44*(2), 428–439. https://doi.org/10.1037/0022-3514.44.2.428

Lustig, I. (2003). *The effects of studying distal conflicts on the perception of a proximal one* [Unpublished thesis].

Luzzo, D. A., & McWhirter, E. H. (2001). Sex and Ethnic Differences in the Perception of Educational and Career-Related Barriers and Levels of Coping Efficacy. *Journal of Counseling & Development*, *79*(1), 61–67. https://doi.org/10.1002/j.1556-6676.2001.tb01944.x

Major, B., Quinton, W. J., & Schmader, T. (2003). Attributions to discrimination and self-esteem: Impact of group identification and situational ambiguity. *Journal of Experimental Social Psychology*, *39*(3), 220–231. https://doi.org/10.1016/s0022-1031(02)00547-4

Marini, M., & Banaji, M. R. (2020). An implicit gender sex-science association in the general population and STEM faculty. *The Journal of General Psychology*, 1–28. https://doi.org/10.1080/00221309.2020.1853502

Markus, H., Crane, M., Bernstein, S., & Siladi, M. (1982). Self-schemas and gender. *Journal of Personality and Social Psychology*, *42*(1), 38–50. https://doi.org/10.1037/0022-3514.42.1.38

Marshall, S. P. (1995). *Schemas in problem solving.* Cambridge University Press.

Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology, 52*(3), 489–499. https://doi.org/10.1037/0022-3514.52.3.489

Martin, C. L. (1995). Stereotypes about children with traditional and nontraditional gender roles. *Sex Roles*, *33*(11-12), 727–751. https://doi.org/10.1007/bf01544776

Martin, C. L., & Ruble, D. (2004). Children's Search for Gender Cues. *Current Directions in Psychological Science, 13*(2), 67–70. https://doi.org/10.1111/j.0963-7214.2004.00276.x

Martinez, E. D., Botos, J., Dohoney, K. M., Geiman, T. M., Kolla, S. S., Olivera, A., Qiu, Y., Rayasam, G. V., Stavreva, D. A., & Cohen-Fix, O. (2007). Falling off the academic bandwagon. *EMBO Reports*, *8*(11), 977–981. https://doi.org/10.1038/sj.embor.7401110

Mastari, L., Spruyt, B., & Siongers, J. (2019). Benevolent and Hostile Sexism in Social Spheres: The Impact of Parents, School and Romance on Belgian Adolescents' Sexist Attitudes. *Frontiers in Sociology*, *4*. https://doi.org/10.3389/fsoc.2019.00047

Mathews, G. A., Fane, B. A., Conway, G. S., Brook, C. G. D., & Hines, M. (2009). Personality and congenital adrenal hyperplasia: Possible effects of prenatal androgen exposure. *Hormones and Behavior*, *55*(2), 285–291. https://doi.org/10.1016/j.yhbeh.2008.11.007

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, *36*(9), 929–940. https://doi.org/10.1037/0022-3514.36.9.929

McCrae, R. R., Terracciano, A., & Personality Profiles of Cultures Project. (2005). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology*, *88*(3), 547–561. https://doi.org/10.1037/0022-3514.88.3.547

McCreary, D. R., & Steinberg, M. (1992). The Personal Attributes Questionnaire in Britain: Establishing construct validity. *British Journal of Social Psychology*, *31*(4), 369–378. https://doi.org/10.1111/j.2044-8309.1992.tb00979.x

McDermott, M. L. (2016). *Masculinity and Femininity in American Political Behavior*. Oxford Univ. Press.

McGuire, W. J. (1964). Inducing resistance to persuasion: some contemporary approaches. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*. Academic Press.

McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *The Journal of Abnormal and Social Psychology*, *62*(2), 327–337. https://doi.org/10.1037/h0042026

McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, *39*(1), 83–90. https://doi.org/10.1016/s0022-1031(02)00513-9

Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F., & Rothermund, K. (2019). Predicting Behavior With Implicit Measures: Disillusioning Findings, Reasonable Explanations, and Sophisticated Solutions. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02483

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, *100*(6), 1678–1712. https://doi.org/10.1037/apl0000022

Miller, C. H., Ivanov, B., Sims, J., Compton, J., Harrison, K. J., Parker, K. A., Parker, J. L., & Averbeck, J. M. (2013). Boosting the Potency of Resistance: Combining the Motivational Forces of Inoculation and Psychological Reactance. *Human Communication Research*, *39*(1), 127–155. https://doi.org/10.1111/j.1468-2958.2012.01438.x

Miner, K. N., Walker, J. M., Bergman, M. E., Jean, V. A., Carter-Sowell, A., January, S. C., & Kaunas, C. (2018). From "Her" Problem to "Our" Problem: Using an Individual Lens Versus a Social-Structural Lens to Understand Gender Inequity in STEM. *Industrial and Organizational Psychology*, *11*(2), 267–290. https://doi.org/10.1017/iop.2018.7

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*(3), 469–485. https://doi.org/10.1037/0022-3514.65.3.469

Montes, A. (2021). Gender stereotyping and retro sexism in advertising discourse from a postfeminist perspective. *Feminismo*, *38*, 107–202.

Moors, G., Kieruj, N. D., & Vermunt, J. K. (2014). The Effect of Labeling and Numbering of Response Scales on the Likelihood of Response Bias. *Sociological Methodology*, *44*(1), 369–399. https://doi.org/10.1177/0081175013516114

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Moss-Racusin, C. A., Molenda, A. K., & Cramer, C. R. (2015). Can Evidence Impact Attitudes? Public Reactions to Evidence of Gender Bias in STEM Fields. *Psychology of Women Quarterly*, *39*(2), 194–209. https://doi.org/10.1177/0361684314565777

Moss-Racusin, C. A., Pietri, E. S., Hennes, E. P., Dovidio, J. F., Brescoll, V. L., Roussos, G., & Handelsman, J. (2018). Reducing STEM gender bias with VIDS (video interventions for diversity in STEM). *Journal of Experimental Psychology: Applied*, *24*(2), 236–260. https://doi.org/10.1037/xap0000144

Moss-Racusin, C. A., & Rabasco, H. (2018). Reducing gender identity bias through imagined intergroup contact. *Journal of Applied Social Psychology*, *48*(8), 457–474. https://doi.org/10.1111/jasp.12525

Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2014). Scientific Diversity Interventions. *Science*, *343*(6171), 615–616. https://doi.org/10.1126/science.1245936

Moya, M., Glick, P., Expósito, F., de Lemus, S., & Hart, J. (2007). It's for Your Own Good: Benevolent Sexism and Women's Reactions to Protectively Justified Restrictions. *Personality and Social Psychology Bulletin*, *33*(10), 1421–1434. https://doi.org/10.1177/0146167207304790

Nadelson, C., Salt, P., & Notman, M. T. (1983). Evidence of Physician Adaptability During the Medical Training Period. *Academic Psychiatry*, *7*(3), 167–182. https://doi.org/10.1007/bf03399884

Nagoshi, J. L., Brzuzy, S., & Terrell, H. K. (2012). Deconstructing the complex perceptions of gender roles, gender identity, and sexual orientation among transgender individuals. *Feminism & Psychology*, *22*(4), 405–422. https://doi.org/10.1177/0959353512461929

Nagoshi, J. L., Nagoshi, C. T., & Brzuzy, S. (2014). *Gender and sexual identity : transcending feminist and queer theory*. Springer.

Nanda, S. (2015). Hijras. In P. Whelehan & A. Bolin (Eds.), *The International Encyclopedia of Human Sexuality* (pp. 501–581). Wiley Blackwell.

Napholz, L. (1995). Indexes of Psychological Well-Being and Role Commitment Among Working Women. *Journal of Employment Counseling*, *32*(1), 22–31. https://doi.org/10.1002/j.2161-1920.1995.tb00421.x

National Coalition for Women and Girls in Education (NCWGE). (2017). *Title IX at 45: Advancing Opportunity through Equity in Education*. NCWGE.

National Science Board. (2016). *S&E Indicators 2016 | NSF - National Science Foundation*. Www.nsf.gov. https://www.nsf.gov/statistics/2016/nsb20161/#/

National Science Foundation. (2008). *Thirty-Three Years of Women in S&E Faculty Positions*. https://www.nsf.gov/statistics/infbrief/nsf08308/?org=NSF

National Science Foundation. (2012). *Science and Engineering Indicators*. http://www.scifun.org/readings/SEind12.pdf

Nelson, D. J. (2007). *A National Analysis of Minorities in Science and Engineering Faculties at Research Universities*. https://w.astro.berkeley.edu/~kalas/ethics/documents/diversity/nelson07.pdf

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

Nosek, B. A., & Smyth, F. L. (2011). Implicit Social Cognitions Predict Sex Differences in Math Engagement and Achievement. *American Educational Research Journal*, *48*(5), 1125–1156. https://doi.org/10.3102/0002831211410683

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88. https://doi.org/10.1080/10463280701489053

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., & Somogyi, M. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., & Reifler, J. (2018). The roles of information deficits and identity threat in the prevalence of misperceptions. *Journal of Elections, Public Opinion and Parties*, *29*(2), 222–244. https://doi.org/10.1080/17457289.2018.1465061

Oldmeadow, J. A., & Fiske, S. T. (2010). Social status and the pursuit of positive social identity: Systematic domains of intergroup differentiation and discrimination for high- and low-status groups. *Group Processes & Intergroup Relations*, *13*(4), 425–444. https://doi.org/10.1177/1368430209355650

Ong, M., Wright, C., Espinosa, L., & Orfield, G. (2011). Inside the Double Bind: A Synthesis of Empirical Research on Undergraduate and Graduate Women of Color in Science, Technology, Engineering, and Mathematics. *Harvard Educational Review*, *81*(2), 172–209. https://doi.org/10.17763/haer.81.2.t022245n7x4752v2

Oreg, S. (2003). Resistance to change: Developing an individual differences measure. *Journal of Applied Psychology*, *88*(4), 680–693. https://doi.org/10.1037/0021-9010.88.4.680

Orue, I., Calvete, E., & Gamez-Guadix, M. (2016). Gender moderates the association between psychopathic traits and aggressive behavior in adolescents. *Personality and Individual Differences*, *94*, 266–271. https://doi.org/10.1016/j.paid.2016.01.043

Osman, S. (2021). Preservice Teachers' Perception of Gender Stereotype Beliefs and Practices at Tumu College of Education, Ghana. *Journal of Advances in Education and Philosophy*, *5*(11), 340–349.

Oswald, P. A. (2002). The Interactive Effects of Affective Demeanor, Cognitive Processes, and Perspective-Taking Focus on Helping Behavior. *The Journal of Social Psychology*, *142*(1), 120–132. https://doi.org/10.1080/00224540209603890

Padesky, C. A. (1994). Schema change processes in cognitive therapy. *Clinical Psychology & Psychotherapy*, *1*(5), 267–278. https://doi.org/10.1002/cpp.5640010502

Paluck, E. L. (2006). Diversity Training and Intergroup Contact: A Call to Action Research. *Journal of Social Issues*, *62*(3), 577–595. https://doi.org/10.1111/j.1540-4560.2006.00474.x

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology*, *96*(3), 574–587. https://doi.org/10.1037/a0011989

Paluck, E. L., & Green, D. P. (2009). Prejudice Reduction: What Works? A Review and Assessment of Research and Practice. *Annual Review of Psychology*, *60*(1), 339–367. https://doi.org/10.1146/annurev.psych.60.110707.163607

Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, *113*(3), 566–571. https://doi.org/10.1073/pnas.1514483113

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology*, *42*(6), 1051–1068. https://doi.org/10.1037/0022-3514.42.6.1051

Parker, K. A., Ivanov, B., & Compton, J. (2012). Inoculation's Efficacy With Young Adults' Risky Behaviors: Can Inoculation Confer Cross-Protection Over Related but Untreated Issues? *Health Communication*, *27*(3), 223–233. https://doi.org/10.1080/10410236.2011.575541

Parry, D. C. (2014). The Relational Politics of Gender and Leisure. *Semantic Scholar*.

Paulhus, D. L., & Trapnell, P. D. (2008). Self-presentation of personality: An agency-communion framework. In O. P. John, R. W. Robins, & L. A. Pervin

(Eds.), *Handbook of personality: Theory and research* (pp. 492–517). The Guilford Press.

Payne, T. J., Connor, J. M., & Colletti, G. (1987). Gender-based schematic processing: An empirical investigation and reevaluation. *Journal of Personality and Social Psychology, 52*(5), 937–945. https://doi.org/10.1037/0022-3514.52.5.937

Paynter, J., Luskin-Saxby, S., Keen, D., Fordyce, K., Frost, G., Imms, C., Miller, S., Trembath, D., Tucker, M., & Ecker, U. (2019). Evaluation of a template for countering misinformation—Real-world Autism treatment myth debunking. *PLOS ONE*, *14*(1), e0210746. https://doi.org/10.1371/journal.pone.0210746

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Petty, R. E., & Briñol, P. (2010). Attitude change. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of the science* (pp. 217–259). Oxford University Press.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The Meta–Cognitive Model (MCM) of Attitudes: Implications for Attitude Measurement, Change, and Strength. *Social Cognition*, *25*(5), 657–686. https://doi.org/10.1521/soco.2007.25.5.657

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Communication and Persuasion*, 1–24. https://doi.org/10.1007/978-1-4612-4964-1_1

Pew Research Center. (2021). *Pew Research Center Analysis of 2017 - 2019 American Community Survey (IPUMS) "STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity."*

Pfau, M. (1992). The potential of inoculation in promoting resistance to the effectiveness of comparative advertising messages. *Communication Quarterly*, *40*(1), 26–44. https://doi.org/10.1080/01463379209369818

Pfau, M. (1995). Designing messages for behavioral inoculation. In E. Maibach & R. L. Parrott (Eds.), *Designing Health Messages: Approaches from Communication Theory and Public Health Practice* (pp. 99–113). Sage.

Pfau, M. (1997). Inoculation model of resistance to influence. In G. A. Barnett & F. J. Boster (Eds.), *Progress in communication sciences: Advances in persuasion* (pp. 133–171). Ablex.

Pfau, M., Bockern, S. V., & Kang, J. G. (1992). Use of inoculation to promote resistance to smoking initiation among adolescents. *Communication Monographs*, *59*(3), 213–230. https://doi.org/10.1080/03637759209376266

Pfau, M., Compton, J., Parker, K. A., Wittenberg, E. M., An, C., Ferguson, M., Horton, H., & Malyshev, Y. (2004). The Traditional Explanation for Resistance Versus Attitude Accessibility. *Human Communication Research*, *30*(3), 329–360. https://doi.org/10.1111/j.1468-2958.2004.tb00735.x

Pfau, M., Haigh, M. M., Shannon, T., Tones, T., Mercurio, D., Williams, R., Binstock, B., Diaz, C., Dillard, C., Browne, M., Elder, C., Reed, S., Eggers, A., & Melendez, J. (2008). The Influence of Television News Depictions of the Images of War on Viewers. *Journal of Broadcasting & Electronic Media*, *52*(2), 303–322. https://doi.org/10.1080/08838150801992128

Pfau, M., Ivanov, B., Houston, B., Haigh, M., Sims, J., Gilchrist, E., Russell, J., Wigley, S., Eckstein, J., & Richert, N. (2005). Inoculation and Mental Processing: The Instrumental Role of Associative Networks in the Process of Resistance to Counterattitudinal Influence. *Communication Monographs*, *72*(4), 414–441. https://doi.org/10.1080/03637750500322578

Pietri, E. S., Moss-Racusin, C. A., Dovidio, J. F., Guha, D., Roussos, G., Brescoll, V. L., & Handelsman, J. (2017). Using Video to Increase Gender Bias Literacy Toward

Women in Science. *Psychology of Women Quarterly*, *41*(2), 175–196. https://doi.org/10.1177/0361684316674721

Plakoyiannaki, E., & Zotos, Y. (2009). Female role stereotypes in print advertising. *European Journal of Marketing*, *43*(11/12), 1411–1434. https://doi.org/10.1108/03090560910989966

Plant, E. A., & Devine, P. G. (2003). The Antecedents and Implications of Interracial Anxiety. *Personality and Social Psychology Bulletin*, *29*(6), 790–801. https://doi.org/10.1177/0146167203029006011

Plous, S. (1991). Biases in the Assimilation of Technological Breakdowns: Do Accidents Make Us Safer? *Journal of Applied Social Psychology*, *21*(13), 1058–1082. https://doi.org/10.1111/j.1559-1816.1991.tb00459.x

Pluviano, S., Watt, C., & Della Sala, S. (2017). Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLOS ONE*, *12*(7), e0181640. https://doi.org/10.1371/journal.pone.0181640

Prasad, M., Perrin, A. J., Bezila, K., Hoffman, S. G., Kindleberger, K., Manturuk, K., & Powers, A. S. (2009). "There Must Be a Reason": Osama, Saddam, and Inferred Justification. *Sociological Inquiry*, *79*(2), 142–162. https://doi.org/10.1111/j.1475-682x.2009.00280.x

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: a meta-analysis. *Criminology*, *38*(3), 931–964. https://doi.org/10.1111/j.1745-9125.2000.tb00911.x

Preciado, P. B. (2008). *Testo Yonqui*. Espasa Calpe.

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

Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology*, *38*(1), 141–152. https://doi.org/10.1037/0022-3514.38.1.141

Ramsey, L. R. (2017). Agentic traits are associated with success in science more than communal traits. *Personality and Individual Differences*, *106*, 6–9.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, *111*(12), 4403–4408. https://doi.org/10.1073/pnas.1314788111

Rich, P. R., & Zaragoza, M. S. (2020). Correcting Misinformation in News Stories: An Investigation of Correction Timing and Correction Durability. *Journal of Applied Research in Memory and Cognition*, *9*(3). https://doi.org/10.1016/j.jarmac.2020.04.001

Ridgeway, Cecilia L., & Erickson, K. (2000). Creating and Spreading Status Beliefs. *American Journal of Sociology*, *106*(3), 579–615. https://doi.org/10.1086/318966

Riggle, E. D. B., Ellis, A. L., & Crawford, A. M. (1996). The Impact of "'Media Contact'" on Attitudes Toward Gay Men. *Journal of Homosexuality*, *31*(3), 55–69. https://doi.org/10.1300/j082v31n03_04

Rizzolatti, G., & Craighero, L. (2005). Mirror neuron: a neurological approach to empathy. *Research and Perspectives in Neurosciences*, 107–123. https://doi.org/10.1007/3-540-29803-7_9

Roberts, L. C., & Hill, H. T. (2003). Come and Listen to a Story about a Girl Named Rex: Using Children's Literature To Debunk Gender Stereotypes. *Young Children*, *58*, 39–42.

Robinson, E. A. R., & Doueck, H. J. (1994). Implications of the Pre/Post/Then Design for Evaluating Social Group Work. *Research on Social Work Practice*, *4*(2), 224–239. https://doi.org/10.1177/104973159400400207

Robnett, R. D. (2016). Gender Bias in STEM Fields. *Psychology of Women Quarterly*, *40*(1), 65–79. https://doi.org/10.1177/0361684315596162

Roozenbeek, J., Maertens, R., McClanahan, W., & van der Linden, S. (2020). Disentangling Item and Testing Effects in Inoculation Research on Online Misinformation: Solomon

Revisited. *Educational and Psychological Measurement*, 001316442094037. https://doi.org/10.1177/0013164420940378

Roozenbeek, J., & van der Linden, S. (2018). The fake news game: actively inoculating against the risk of misinformation. *Journal of Risk Research*, *22*(5), 570–580. https://doi.org/10.1080/13669877.2018.1443491

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0279-9

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1), 16–20. https://doi.org/10.1007/bf02322211

Rosette, A. S., Koval, C. Z., Ma, A., & Livingston, R. (2016). Race matters for women leaders: Intersectional effects on agentic deficiencies and penalties. *The Leadership Quarterly*, *27*(3), 429–445. https://doi.org/10.1016/j.leaqua.2016.01.008

Rosette, A. S., & Livingston, R. W. (2012). Failure is not an option for Black women: Effects of organizational performance on leaders with single versus dual-subordinate identities. *Journal of Experimental Social Psychology*, *48*(5), 1162–1167. https://doi.org/10.1016/j.jesp.2012.05.002

Rubin, M., & Badea, C. (2007). Why Do People Perceive Ingroup Homogeneity on Ingroup Traits and Outgroup Homogeneity on Outgroup Traits? *Personality and Social Psychology Bulletin*, *33*(1), 31–42. https://doi.org/10.1177/0146167206293190

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). *Handbook of child psychology* (N. Eisenberg, W. Damon, & R. M. Lerner, Eds.; pp. 858–932). John Wiley & Sons.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*(3), 629–645. https://doi.org/10.1037/0022-3514.74.3.629

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004–1010. https://doi.org/10.1037/0022-3514.77.5.1004

Rudman, L. A., & Glick, P. (2001). Prescriptive Gender Stereotypes and Backlash Toward Agentic Women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Rudman, L. A., & Glick, P. (2008). *The social psychology of gender : how power and intimacy shape gender relations*. Guilford Press.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179. https://doi.org/10.1016/j.jesp.2011.10.008

Ryan, C. (2003). Stereotype accuracy. *European Review of Social Psychology*, *13*(1), 75–109. https://doi.org/10.1080/10463280240000037

Ryan, C., Park, B., & Judd, C. M. (1996). Assessing stereotype accuracy: Implications for understanding the stereotyping process. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 121–157). Guilford Press.

Sachdev, A. R. (2018). Gender Disparity in STEM Across Cultures. *Industrial and Organizational Psychology*, *11*(2), 309–313. https://doi.org/10.1017/iop.2018.20

Saragovi, C., Aubé, J., Koestner, R., & Zuroff, D. (2002). Traits, Motives, and Depressive Styles as Reflections of Agency and Communion. *Personality and Social Psychology Bulletin*, *28*(5), 563–577. https://doi.org/10.1177/0146167202288001

Sawyerr, O. O., Strauss, J., & Yan, J. (2005). Individual value structure and diversity attitudes. *Journal of Managerial Psychology*, *20*(6), 498–521. https://doi.org/10.1108/02683940510615442

Schaumberg, R. L., & Flynn, F. J. (2017). Self-reliance: A Gender Perspective on its Relationship to Communality and Leadership Evaluations. *Academy of Management Journal*, *60*(5), 1859–1881. https://doi.org/10.5465/amj.2015.0018

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The Parasocial Contact Hypothesis. *Communication Monographs*, *72*(1), 92–115. https://doi.org/10.1080/0363775052000342544

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*(3), 440–452. https://doi.org/10.1037/0022-3514.85.3.440

Schmidsberger, F., & Löffler-Stastka, H. (2018). Empathy is proprioceptive: the bodily fundament of empathy – a philosophical contribution to medical education. *BMC Medical Education*, *18*(1). https://doi.org/10.1186/s12909-018-1161-y

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schneider, M. B. (2001). Encouragement of women physics majors at Grinnell College: A case study. *The Physics Teacher*, *39*(5), 280–282. https://doi.org/10.1119/1.1375465

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *NeuroImage*, *42*(1), 393–403. https://doi.org/10.1016/j.neuroimage.2008.04.180

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105. https://doi.org/10.1037/0003-066x.54.2.93

Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, *2*(1), 85–95.

Schwarz, N., Sanna, L. J., Skurnik, I. W., & Yoon, C. (2007). Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns. *Semantic Scholar*. https://doi.org/10.1016/S0065-2601(06)39003-X

Sherman, D. K., & Cohen, G. L. (2002). Accepting Threatening Information: Self–Affirmation and the Reduction of Defensive Biases. *Current Directions in Psychological Science*, *11*(4), 119–123. https://doi.org/10.1111/1467-8721.00182

Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology*, *104*(12), 1514–1534. https://doi.org/10.1037/apl0000420

Shifren, K., & Bauserman, R. L. (1996). The relationship between instrumental and expressive traits, health behaviors, and perceived physical health. *Sex Roles*, *34*(11-12), 841–864. https://doi.org/10.1007/bf01544319

Shih, M., Wang, E., Trahan Bucher, A., & Stotzer, R. (2009). Perspective Taking: Reducing Prejudice Towards General Outgroups and Specific Individuals. *Group Processes & Intergroup Relations*, *12*(5), 565–577. https://doi.org/10.1177/1368430209337463

Sibley, C. G., Wilson, M. S., & Duckitt, J. (2007). Antecedents of Men's Hostile and Benevolent Sexism: The Dual Roles of Social Dominance Orientation and Right-Wing Authoritarianism. *Personality and Social Psychology Bulletin*, *33*(2), 160–172. https://doi.org/10.1177/0146167206294745

Silverman, B. W. (1986). *Density Estimation*. Chapman and Hall.

Simon, B., & Pettigrew, T. F. (1990). Social identity and perceived group homogeneity: Evidence for the ingroup homogeneity effect. *European Journal of Social Psychology*, *20*(4), 269–286. https://doi.org/10.1002/ejsp.2420200402

Sin, I., Stillman, S., & Fabling, R. (2017). What drives the gender wage gap? Examining the roles of sorting, productivity differences, and discrimination. *Motu Working Paper*. https://doi.org/10.29310/wp.2017.15

Singer, T., & Lamm, C. (2009). The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*(7075), 466–469. https://doi.org/10.1038/nature04271

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Smith, V., Devane, D., Begley, C. M., & Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Medical Research Methodology*, *11*(1). https://doi.org/10.1186/1471-2288-11-15

Snell, W. E. (Ed.). (2002). *New directions in the psychology of gender roles: Research and theory*. Snell Publications.

Spence, J. T. (1984). Masculinity, Femininity, and Gender-Related Traits: A Conceptual Analysis and Critique of Current Research1 1Preparation of this article was facilitated by NIMH Grant 32066 (Janet T. Spence and Robert L. Helmreich, Principal Investigators). Thanks are due to Lucia A. Gilbert and Robert L. Helmreich for their reading of the manuscript. *Normal Personality Processes*, *13*, 1–97. https://doi.org/10.1016/b978-0-12-541413-5.50006-1

Spence, J. T. (1985). Gender identity and its implications for the concepts of masculinity and femininity. In *Nebraska Symposium on Motivation: Psychology and gender*. University of Nebraska Press.

Spence, J. T. (1991). Do The Bsri and Paq Measure the Same or Different Concepts? *Psychology of Women Quarterly*, *15*(1), 141–165. https://doi.org/10.1111/j.1471-6402.1991.tb00483.x

Spence, J. T., & Buckner, C. E. (2000). Instrumental and Expressive Traits, Trait Stereotypes, and Sexist Attitudes: What Do They Signify? *Psychology of Women Quarterly*, *24*(1), 44–53. https://doi.org/10.1111/j.1471-6402.2000.tb01021.x

Spence, J. T., & Hahn, E. D. (1997). The Attitudes Toward Women Scale and Attitude Change in College Students. *Psychology of Women Quarterly*, *21*(1), 17–34. https://doi.org/10.1111/j.1471-6402.1997.tb00098.x

Spence, J. T., & Helmreich, R. (1972). The Attitudes toward Women Scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. *JSAS Catalog of Selected Documents in Psychology*, *2*(66).

Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* University of Texas Press.

Spence, J. T., & Helmreich, R. L. (1980). Masculine Instrumentality and Feminine Expressiveness: Their Relationships with Sex Role Attitudes and Behaviors. *Psychology of Women Quarterly*, *5*(2), 147–163. https://doi.org/10.1111/j.1471-6402.1980.tb00951.x

Spence, J. T., & Helmreich, R. L. (1981). Androgyny versus gender schema: A comment on Bem's gender schema theory. *Psychological Review*, *88*(4), 365–368. https://doi.org/10.1037/0033-295x.88.4.365

Spence, J. T., Helmreich, R. L., & Holahan, C. K. (1979). Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors. *Journal of Personality and Social Psychology*, *37*(10), 1673–1682. https://doi.org/10.1037/0022-3514.37.10.1673

Spence, J. T., Helmreich, R. L., & Sawin, L. L. (1980). The Male-Female Relations Questionnaire: A self-report inventory of sex role behaviors and preferences and their relationships to masculine and feminine personality traits, sex role attitudes, and other measures. *JSAS Catalog of Selected Documents in Psychology*, *10*(87).

Spence, J. T., Helmreich, R., & Stapp, J. (1973). A short version of the Attitudes toward Women Scale (AWS). *Bulletin of the Psychonomic Society*, *2*(4), 219–220. https://doi.org/10.3758/bf03329252

Spence, J. T., Helmreich, R., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology*, *32*(1), 29–39. https://doi.org/10.1037/h0076857

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373

Stake, J. E., Zand, D., & Smalley, R. (1996). The Relation of Instrumentality and Expressiveness to Self-Concept and Adjustment: A Social Context Perspective. *Journal of Social and Clinical Psychology*, *15*(2), 167–190. https://doi.org/10.1521/jscp.1996.15.2.167

Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Social influence and intergroup beliefs: The role of perceived social consensus. In J. P. Forgas & K. D. Williams (Eds.), *Social influence: Direct and indirect processes* (pp. 235–252). Psychology Press.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study. *Sex Roles*, *41*(7/8), 509–528. https://doi.org/10.1023/a:1018839203698

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Hillsdale.

Storage, D., Charlesworth, T. E. S., Banaji, M. R., & Cimpian, A. (2020). Adults and children implicitly associate brilliance with men more than women. *Journal of Experimental Social Psychology*, *90*, 104020. https://doi.org/10.1016/j.jesp.2020.104020

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

Swim, J. K., & Cohen, L. L. (1997). Overt, Covert, And Subtle Sexism. *Psychology of Women Quarterly*, *21*(1), 103–118. https://doi.org/10.1111/j.1471-6402.1997.tb00103.x

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1948–1961. https://doi.org/10.1037/xlm0000422

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). Harper Collins.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Allyn and Bacon.

Taber, K. S. (2017). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

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

Taylor, M. C., & Hall, J. A. (1982). Psychological androgyny: Theories, methods, and conclusions. *Psychological Bulletin*, *92*(2), 347–366. https://doi.org/10.1037/0033-2909.92.2.347

Taylor, S. E., & Falcone, H.-T. (1982). Cognitive Bases of Stereotyping. *Personality and Social Psychology Bulletin*, *8*(3), 426–432. https://doi.org/10.1177/0146167282083006

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778–793. https://doi.org/10.1037/0022-3514.36.7.778

Terman, L. M., & Miles, C. C. (1936a). *Sex and Personality. Studies in masculinity and femininity*. McGraw-Hill.

Terman, L. M., & Miles, C. C. (1936b). *Sex and personality: studies in masculinity and femininity*. McGraw-Hill.

Theory of cognitive dissonance, L. A. (1957). *Theory of cognitive dissonance*. Stanford University Press.

Thomas, D. A., & Reznikoff, M. (1984). Sex Role Orientation, Personality Structure, and Adjustment in Women. *Journal of Personality Assessment*, *48*(1), 28–36. https://doi.org/10.1207/s15327752jpa4801_6

Thorndyke, P. W., & Yekovich, F. R. (1979). *A Critique of Schemata as a Theory of Human Story Memory*. Rand Corporation.

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology*, *100*(6), 1027–1042. https://doi.org/10.1037/a0022308

Tormala, Z. L., & Petty, R. E. (2007). Contextual contrast and perceived knowledge: Exploring the implications for persuasion. *Journal of Experimental Social Psychology*, *43*(1), 17–30. https://doi.org/10.1016/j.jesp.2005.11.007

Tougas, F., Brown, R., Beaton, A. M., & St-Pierre, L. (1999). Neosexism among women: the role of personally experienced social mobility attempts. *Personality and Social Psychology Bulletin*, *25*(12), 1487–1497. https://doi.org/10.1177/01461672992510005

Turner, C. S. V. (2002). Women of color in academe: living with multiple marginality. *The Journal of Higher Education*, *73*(1), 74–93. https://doi.org/10.1353/jhe.2002.0013

Tusche, A., Bockler, A., Kanske, P., Trautwein, F.-M., & Singer, T. (2016). Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *Journal of Neuroscience*, *36*(17), 4719–4732. https://doi.org/10.1523/jneurosci.3392-15.2016

Twenge, J. M. (1997). Attitudes toward women, 1970–1995. *Psychology of Women Quarterly*, *21*(1), 35–51. https://doi.org/10.1111/j.1471-6402.1997.tb00099.x

Twenge, J. M., Baumeister, R. F., Tice, D. M., & Stucke, T. S. (2001). If you can't join them, beat them: Effects of social exclusion on aggressive behavior. *Journal of Personality and Social Psychology*, *81*(6), 1058–1069. https://doi.org/10.1037/0022-3514.81.6.1058

UNDP. (2020). Tackling social norms. A game changer for gender inequalities. In *https://hdr.undp.org/sites/default/files/hd_perspectives_gsni.pdf*. United Nations Development Program.

United States Bureau of Labor Statistics. (2019). *American Time Use Survey Summary*. Bls.gov. https://www.bls.gov/news.release/atus.nr0.htm

van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2014). How to communicate the scientific consensus on climate change: plain facts, pie charts or metaphors? *Climatic Change*, *126*(1-2), 255–262. https://doi.org/10.1007/s10584-014-1190-4

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

van der Linden, S., Panagopoulos, C., & Roozenbeek, J. (2019). You are fake news: political bias in perceptions of fake news. *Media, Culture & Society*, *42*(3), 016344372090699. https://doi.org/10.1177/0163443720906992

Van Veelen, R., & Derks, B. (2021). Academics as agentic superheroes: female academics' lack of fit with the agentic stereotype of success limits their career advancement. *British Journal of Social Psychology*. https://doi.org/10.1111/bjso.12515

Veil, S. R., & Kent, M. L. (2008). Issues management and inoculation: Tylenol's responsible dosing advertising. *Public Relations Review*, *34*(4), 399–402. https://doi.org/10.1016/j.pubrev.2008.08.004

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*(1), 1–28. https://doi.org/10.1207/s15327906mbr2501_1

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: the mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology*, *33*(4), 455–472. https://doi.org/10.1002/ejsp.163

Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. A. (2013). Gender differences in implicit and explicit personality traits. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2249080

Vinacke, W. E. (1957). Stereotypes as social concepts. *The Journal of Social Psychology*, *46*(2), 229–243. https://doi.org/10.1080/00224545.1957.9714322

Vollmer, F. (1984). Sex differences in personality and expectancy. *Sex Roles*, *11*(11-12), 1121–1139. https://doi.org/10.1007/bf00288138

Vorauer, J. D., & Sasaki, S. J. (2009). Helpful only in the abstract? *Psychological Science*, *20*(2), 191–197. https://doi.org/10.1111/j.1467-9280.2009.02265.x

Ward, L. C., Thorn, B. E., Clements, K. L., Dixon, K. E., & Sanford, S. D. (2006). Measurement of agency, communion, and emotional vulnerability with the personal

attributes questionnaire. *Journal of Personality Assessment*, *86*(2), 206–216. https://doi.org/10.1207/s15327752jpa8602_10

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, *12*, 129–140.

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the big five. *Frontiers in Psychology*, *2*(178). https://doi.org/10.3389/fpsyg.2011.00178

Weiss, S. (2015, December 18). *7 Sneaky Ways Internalized Misogyny Manifests In Our Everyday Lives*. Bustle. https://www.bustle.com/articles/130737-7-sneaky-ways-internalized-misogyny-manifests-in-our-everyday-lives

Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 U.S. presidential election. *Journal of Cognitive Neuroscience*, *18*(11), 1947–1958. https://doi.org/10.1162/jocn.2006.18.11.1947

Whatley, M. A. (2008). The dimensionality of the 15 item attitudes towards women scale. *Race, Gender & Class*, *15*, 265–273.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, *127*(6), 797–826. https://doi.org/10.1037/0033-2909.127.6.797

Wiggins, J. S. (2003). *Paradigms of personality assessment*. Guilford Press.

Williams, J. E., & Best, D. L. (1990). *Sex and psyche: Gender and self-viewed cross-culturally*. Sage Publications, Inc.

Wilson, A. (1996). How we find ourselves: identity development and two spirit people. *Harvard Educational Review*, *66*(2), 303–318. https://doi.org/10.17763/haer.66.2.n551658577h927h4

Wilson, F. R., & Cook, E. P. (1984). Concurrent validity of four androgyny instruments. *Sex Roles*, *11*, 813–837.

WISE. (2017). *Women in STEM workforce 2017 - Welcome to the WISE Campaign*. Welcome to the WISE Campaign. https://www.wisecampaign.org.uk/statistics/women-in-stem-workforce-2017/

Wojciszke, B. (1994). Multiple meanings of behavior: construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, *67*(2), 222–232. https://doi.org/10.1037/0022-3514.67.2.222

Wolf, S., & Montgomery, D. A. (1977). effects of inadmissible evidence and level of judicial admonishment to disregard on the judgments of mock jurors. *Journal of Applied Social Psychology*, *7*(3), 205–219. https://doi.org/10.1111/j.1559-1816.1977.tb00746.x

Wong, N. C. H., & Harrison, K. J. (2014). Nuances in inoculation: protecting positive attitudes toward the HPV vaccine and the practice of vaccinating children. *J. Women's Health Issues Care*, *3*(6). https://doi.org/10.4172/2325-9795.1000170

Wood, M. L. M. (2007). Rethinking the inoculation analogy: effects on subjects with differing preexisting attitudes. *Human Communication Research*, *33*(3), 357–378. https://doi.org/10.1111/j.1468-2958.2007.00303.x

Wood, T., & Porter, E. (2018). The elusive backfire effect: mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychological Bulletin*, *128*(5), 699–727. https://doi.org/10.1037/0033-2909.128.5.699

Wood, W., & Eagly, A. H. (2010). Gender. *Handbook of Social Psychology*. https://doi.org/10.1002/9780470561119.socpsy001017

Worchel, S., & Brehm, J. W. (1970). Effect of threats to attitudinal freedom as a function of agreement with the communicator. *Journal of Personality and Social Psychology*, *14*(1), 18–22. https://doi.org/10.1037/h0028620

Ysseldyk, R., Greenaway, K. H., Hassinger, E., Zutrauen, S., Lintz, J., Bhatia, M. P., Frye, M., Starkenburg, E., & Tai, V. (2019). A leak in the academic pipeline: identity and

health among postdoctoral women. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01297

Zawadzki, M. J., Danube, C. L., & Shields, S. A. (2012). How to talk about gender inequity in the workplace: Using WAGES as an experiential learning tool to reduce reactance and promote self-efficacy. *Sex Roles, 67*(11-12), 605–616. https://doi.org/10.1007/s11199-012-0181-z

Zucker, A. N., & Bay-Cheng, L. Y. (2010). Minding the gap between feminist identity and attitudes: the behavioral and ideological divide between feminists and non-labelers. *Journal of Personality*, *78*(6), 1895–1924. https://doi.org/10.1111/j.1467-6494.2010.00673.x

APPENDIX

Appendix A. Chapter Two

I) Consent Form

Pilot Study for Understanding Social Decision-making

Before you decide to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. If anything is unclear or you wish to ask questions, please contact the lead researcher listed below. If you understand the information provided and freely consent to participate, please click on the "I agree" button at the bottom of this page.

What is the purpose of this research?

You are invited to participate in a research study investigating factors that affect the way people interpret and evaluate information about other people. To participate, you must confirm you are (1) 18 years of age or older, (2) fluent in English, and (3) able to use audio during this survey (headphones or speakers).

What procedures does this study involve?

This study is an online survey and will take you approximately 7-10 minutes to complete. In the first phase of the study, you will be randomly assigned to watch a video clip. Afterwards, you will be asked a series of questions about what you have watched. Lastly, you will complete a set of questions on your opinions and demographic background. **Please note that as part of this study, you may be exposed to information that is either partially or entirely false or inaccurate.**

What are the potential risks or discomforts?

There are no known risks associated with this research study and there are no right or wrong answers. However, if you feel too uncomfortable to complete the survey, you can withdraw from the study at any time without penalty or loss. This study has been approved by the Psychology Research Ethics Committee of the University of Cambridge.

Are there benefits to participating in this research?

There are no direct benefits to participating in this research.

What about privacy and confidentiality?

All responses are confidential and anonymous. We will not ask you for your name or identifiable information - you will only be identified by a research number. All data will be published in aggregate form only.

Will I be paid for my participation in this research?

You will receive £0.75 for your participation.

Who can I contact if I have questions?

If you have further questions about this study or your rights, or if you wish to lodge a complaint or concern, you may contact the researcher, [name redacted] at [email address redacted].

By clicking “I consent” I confirm that I understand the information above and I consent to participate in this study. If you wish to participate in this study, please click “I consent” to continue. Otherwise, please click “I **do not** consent.”

II) Debriefing Form

Thank you for taking part in our study.

This study examined strategies to reduce the impact of inaccurate gender stereotypes in STEM. Specifically, we investigated the effect of different videos on reducing gender-stereotypic beliefs. One group of participants watched a video with inaccurate information about why few women are in STEM, another group watched a video with accurate information about the effects of gender stereotypes in STEM, and a third group watched both videos. A final group of participants watched a science video unrelated to gender stereotypes.

The scientific consensus is that there are no inborn gender differences in STEM aptitude. If this is inconsistent with your previous or current belief, you may have felt uncomfortable during the study. It is normal to feel threatened when one's beliefs are challenged, and it is cognitively challenging to adopt a new belief, especially about something as salient as gender. While stereotype development is normal, it is important to note that it results from basic information processing and associations. Therefore, people can bring awareness to their inaccurate stereotypes and dismiss them. It is important to do this to help reduce the negative effects they can have on beliefs, attitudes, and behaviours.

Thank you again for participating in this study. Please do not share your knowledge about the study with potential participants, as this might distort the results of our research. This research has important implications for improving gender equality in STEM positions.

Your results are confidential and anonymous. Your responses do not include your name or anything else that could identify you. You will not be contacted by us. You have the right to withdraw your data if you so wish.

If you have any concerns or questions regarding this research project, please contact the researcher: [name redacted] at [email address redacted].

Please click the next arrow below to be redirected to Prolific Academic so that your completion can be recorded.

III) Materials

Intervention Videos

1. Inoculation video: https://youtu.be/5vCU9i6fP6U, 4:07 min

I designed a 4-minute 'antibiotic' video with the aim to 'treat' the inaccurate stereotype that men are innately more competent at STEM fields than women. For this purpose, I combined video clips from:

1. A previously tested and validated expert video (Pietri et al, 2017) (https://youtu.be/5vCU9i6fP6U)
2. A video obtained from YouTube channel AsapScience, which is dedicated to producing animated videos to educate on a range of scientific topics ('Are Boys Smarter than Girls?', YouTube https://youtu.be/HKJFESU7hqE)
3. A video from the YouTube channel of the CERCA Institute (Research Centers of Catalonia), which addresses the issue of recruitment bias in research institutes ('Questioning Competencies – Recruitment Bias in research institutes', YouTube, https://youtu.be/3nMzdxAcTWM?list=PLIRaTLZqki-CuOpGbTSp_CugJvyBM_Btj)
4. A video from Life at Google channel, which was used as part of a workshop aimed at Google workers to mitigate the potentially negative influence of unconscious bias in the workplace ('Unconscious Bias at Work- Making the Unconscious Conscious', YouTube, https://youtu.be/NW5s_-Nl3JE)

Following inoculation theory design, the video begins with an expert presenting the scientific consensus ("it's not a lack of ability causing women to not pursue careers in [STEM] fields") and debunking the myth that women perform worse than men in math and science. This is followed by a warning that some people use misleading tactics to try to convince others that men are innately better at STEM. The video continues with clips of animation, expert interviews, narrative scenes, and text to specifically explain the flaws in two misleading tactics with scientific facts:

> Misleading tactic 1: Exaggerate the size of biological differences between men and women and incorrectly link them to STEM performance

Scientific fact 1: There are very few, specific mental tasks that show small or negligible gender differences, but these have not been found to relate to STEM

Misleading tactic 2: Argue STEM ability is inherent and somehow connected to one's gender

Scientific fact 2: The brain changes in response to experiences. When people do STEM, their brains change to make them better at STEM. There is no reliable scientific evidence that men are naturally better at STEM.

The video concludes with expert analysts explaining that our subtle biases can have effects on the people we hire and promote, and that we have a responsibility to understand the assumptions and errors we make.

2. Misinformation video: https://youtu.be/UC0OlTXXrZs, 0:30 seconds

The video clip is cut from an online video published by a popular blogger, Lauren Chen ('Are Men & Women Different? | Gender Disparities & STEM', https://youtu.be/Niv9fDJ-T0s). It promotes and defends inaccurate information about innate gender differences in STEM ability.

3. Control video: https://youtu.be/Wf0kwUq64Bc, 3:39 min

I designed a control video that narrates an animated STEM video about protein synthesis and enzymes. The video was obtained from a popular YouTube channel known as FuseSchool, a global Open Education charity aimed at educating the public on a range of topics through the use of animation videos. The control video was created using two videos from this source, one on "Protein synthesis" and one on "Enzymes" and included both male and female scientists to ensure there is no confounding effect of expert gender.

IV) Measures

Scales without references are face-valid items constructed by me.

State Logical Thinking (Pietri et al., 2017; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. It made me think logically about the evidence it presented
2. It made its point clearly with evidence and logic
3. It made me think critically about the evidence it presented
4. It presented its evidence in a clear and logical manner

State Perspective Taking (Pietri et al., 2017)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. It made me think about how I would feel being wrongfully stereotyped
2. It made me think about what it would feel like to be negatively stereotyped

State Empathic Concern

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. It made me feel concern for women in STEM.
2. It made me feel concerned about gender stereotypes in STEM.

Perceived Increase in Knowledge of Gender Stereotypes in STEM (Pietri et al., 2017; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements.

1. This video increased my knowledge of the facts and evidence about gender stereotypes in STEM
2. After watching this video, I feel like I know more about research on gender stereotypes in STEM
3. This video increased my understanding of gender stereotypes in STEM

4. After watching this video, I feel like I know more about the impact of gender stereotypes in STEM

Counterarguing (Miller et al. 2013)

Items were on a 1 (I agreed with the arguments) to 7 (I thought of arguments against) Likert-type scale.

On the following scale, please indicate your response to the messages you heard in the video.

	I agreed with the arguments	.	.	I thought of arguments both for and against	.	.	I thought of arguments against
Response	○	○	○	○	○	○	○

Credibility (Miller et al., 2013; adapted)

Five semantic differential items anchored by bipolar adjectives on a 7-point scale.

On the following scale, please rate the video you watched.

Not scientific	○ ○ ○ ○ ○ ○ ○	Scientific
Unpersuasive	○ ○ ○ ○ ○ ○ ○	Persuasive
Inaccurate	○ ○ ○ ○ ○ ○ ○	Accurate
Not trustworthy	○ ○ ○ ○ ○ ○ ○	Trustworthy
Not credible	○ ○ ○ ○ ○ ○ ○	Credible

Video Comprehension

Face-valid semantic differential items anchored by bipolar adjectives on a 7-point scale.

On the following scale, please rate the video you watched.

Not too slow	○ ○ ○ ○ ○ ○ ○	Too slow
Not too fast	○ ○ ○ ○ ○ ○ ○	Too fast
Not too much information	○ ○ ○ ○ ○ ○ ○	Too much information
Easy to follow	○ ○ ○ ○ ○ ○ ○	Confusing
Understandable wording	○ ○ ○ ○ ○ ○ ○	Complicated wording

Manipulation Check

The video I watched presented evidence about

a) Racism in film
b) Organic farming in England
c) Protein synthesis in humans
d) Gender stereotypes in STEM

Active inoculation

In 1-2 sentences, what do you think you will remember most about the video you watched?

Discernment of Gender Stereotypes in STEM

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements.

1. The scientific consensus is that there are no inborn gender differences in STEM ability.
2. More females are not in STEM because they are naturally interested in other things. [R]
3. On average, men's and women's brains function differently. [R]
4. The scientific consensus is that male brains are naturally better at STEM. [R]
5. Some people use misleading tactics to claim there are differences between men's and women's STEM ability.
6. Gender stereotypes decrease women's chances of being hired in STEM, even when they are equally qualified.

Awareness of Bias Against Women in STEM (Pietri et al., 2017; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements.

1. In my opinion, women in STEM fields do not face discrimination based on their gender
2. In my opinion, women in STEM fields are not taken as seriously as their male colleagues.
3. In my opinion, women in STEM fields often have to work harder than their male colleagues to show they are equally competent as the men.

Gender Stereotype Index (Rudman & Glick, 2001)

Agentic items were on a –3 (*more true of women*) to +3 (*more true of men*) 7-point scale.

Please indicate to what extent you think each of the following traits are more true of women or of men.

1. Individualistic
2. Competitive
3. Independent
4. Self-sufficient

Demographics

[Age, collected as a continuous variable]

What is your age? (For example: if you are 30 years old, please enter: 30)

[Education]

What is the highest level of education you have completed?

1. Primary school
2. GCSEs or equivalent
3. A-levels or equivalent
4. University undergraduate degree
5. University post-graduate degree
6. University master's degree
7. Doctorate degree

[Employment status]

What is your current employment status?

1. Full-time employment
2. Part-time employment
3. Self-employed
4. Student
5. Unemployed, looking for work
6. Unemployed, not looking for work
7. Unable to work
8. Retired

[Ethnicity]

Which of the following best describes your ethnicity/race?

1. White
2. Mixed
3. Black or Black British
4. Asian or Asian British
5. Chinese
6. Other

[Gender]

Which of the following best describes your gender identity?

1. Female
2. Male
3. Transgender male
4. Transgender female
5. Non-conforming/queer
6. Other
7. Prefer not to say

[Sexual orientation]

Which of the following best describes your sexual orientation?

1. Heterosexual (straight)
2. Homosexual (gay)
3. Bisexual
4. Pansexual
5. Asexual
6. Other
7. Prefer not to say

[Political ideology]

On a scale from very conservative (1) to very liberal (7), how would you describe your political views?

1. Very conservative
2. Mostly conservative
3. Somewhat conservative
4. Neither conservative or liberal
5. Somewhat liberal
6. Mostly liberal
7. Very liberal

[Religion]

Which of the following best describes your religion?

1. Christian (all denominations)
2. Buddhist
3. Sikh
4. Hindu
5. Jewish
6. Muslim
7. No religion
8. Other
9. Prefer not to say

Appendix B. Chapter Three

I) Consent Form

Understanding Social Decision-making
Before you decide to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. If anything is unclear or you wish to ask questions, please contact the lead researcher listed below. If you understand the information provided and freely consent to participate, please click on the "I agree" button at the bottom of this page.

What is the purpose of this research?
You are invited to participate in a research study investigating factors that affect the way people interpret and evaluate information about other people. To participate, you must confirm you are (1) 18 years of age or older, (2) fluent in English, and (3) able to use audio during this survey (headphones or speakers).

What procedures does this study involve?
This study is an online survey and will take you approximately 15 minutes to complete.
In the first phase of the study, you will be randomly assigned to watch a video clip. Afterwards, you will be asked a series of questions about what you have watched. Lastly, you will complete a set of questions on your opinions and demographic background. **Please note that as part of this study, you may be exposed to information that is either partially or entirely false or inaccurate.**

What are the potential risks or discomforts?
There are no known risks associated with this research study and there are no right or wrong answers. However, if you feel too uncomfortable to complete the survey, you can withdraw from the study at any time without penalty or loss. This study has been approved by the Psychology Research Ethics Committee of the University of Cambridge.

Are there benefits to participating in this research?
There are no direct benefits to participating in this research.

What about privacy and confidentiality?
All responses are confidential and anonymous. We will not ask you for your name or identifiable information - you will only be identified by a research number. All data will be published in aggregate form only.

Will I be paid for my participation in this research?
You will receive £1.25 for your participation.

Who can I contact if I have questions?
If you have further questions about this study or your rights, you may contact the researcher, [name redacted] at [email address redacted].

By clicking "I consent" I confirm that I understand the information above and I consent to participate in this study.

If you wish to participate in this study, please click “I consent” to continue. Otherwise, please click “I **do not** consent.”

II) Debriefing Form

Thank you for taking part in our study.

*****YOU MAY HAVE BEEN EXPOSED TO FALSE INFORMATION IN THIS STUDY. IF YOU VIEWED A VIDEO CLAIMING MEN AND WOMEN HAVE INNATE DIFFERENCES IN STEM CAPABILITIES AND INTERESTS, THIS WAS FALSE INFORMATION.*****

This study examined strategies to reduce the impact of inaccurate gender stereotypes in STEM. Specifically, we investigated the effect of different videos on reducing gender-stereotypic beliefs. One group of participants watched a video with inaccurate information about why few women are in STEM, another group watched a video with accurate information about the effects of gender stereotypes in STEM, and a third group watched both videos. A final group of participants watched a science video unrelated to gender stereotypes.

The scientific consensus is that there are no biologically determined gender differences in STEM aptitude. If this is inconsistent with your previous or current belief, you may have felt uncomfortable during the study. It is normal to feel threatened when one's beliefs are challenged, and it is cognitively challenging to adopt a new belief, especially about something as salient as gender. While stereotype development is normal, it is important to note that it results from basic information processing and associations. Therefore, people can bring awareness to their inaccurate stereotypes and choose to dismiss them. It is important to do this to help reduce the negative effects they can have on beliefs, attitudes, and behaviours.

Thank you again for participating in this study. Please do not share your knowledge about the study with potential participants, as this might distort the results of our research. This research has important implications for improving gender diversity in STEM positions.

Your results are confidential and anonymous. Your responses do not include your name or anything else that could identify you. You will not be contacted by us. You have the right to withdraw your data if you so wish.

If you have any concerns or questions regarding this research project, please contact the researcher: [name redacted] at [email address redacted].

Please click the next arrow below to be redirected to Prolific Academic so that your completion can be recorded.

III) Materials

Intervention Videos

1. **Inoculation video: https://youtu.be/5vCU9i6fP6U (4:07 min)**

I designed a 4-minute 'antibiotic' video with the aim to 'treat' the inaccurate stereotype that men are innately more competent at STEM fields than women. For this purpose, I combined video clips from:

5. A previously tested and validated expert video (Pietri et al, 2017) (https://youtu.be/5vCU9i6fP6U)
6. A video obtained from YouTube channel AsapScience, which is dedicated to producing animated videos to educate on a range of scientific topics ('Are Boys Smarter than Girls?', YouTube https://youtu.be/HKJFESU7hqE)
7. A video from the YouTube channel of the CERCA Institute (Research Centers of Catalonia), which addresses the issue of recruitment bias in research institutes ('Questioning Competencies – Recruitment Bias in research institutes', YouTube, https://youtu.be/3nMzdxAcTWM?list=PLIRaTLZqki-CuOpGbTSp_CugJvyBM_Btj)
8. A video from Life at Google channel, which was used as part of a workshop aimed at Google workers to mitigate the potentially negative influence of unconscious bias in the workplace ('Unconscious Bias at Work- Making the Unconscious Conscious', YouTube, https://youtu.be/NW5s_-Nl3JE)

Following inoculation theory design, the video begins with an expert presenting the scientific consensus ("it's not a lack of ability causing women to not pursue careers in [STEM] fields") and debunking the myth that women perform worse than men in math and science. This is followed by a warning that some people use misleading tactics to try to convince others that men are innately better at STEM. The video continues with clips of animation, expert interviews, narrative scenes, and text to specifically explain the flaws in two misleading tactics with scientific facts:

Misleading tactic 1: Exaggerate the size of biological differences between men and women and incorrectly link them to STEM performance

Scientific fact 1: There are very few, specific mental tasks that show small or negligible gender differences, but these have not been found to relate to STEM

Misleading tactic 2: Argue STEM ability is inherent and somehow connected to one's gender

Scientific fact 2: The brain changes in response to experiences. When people do STEM, their brains change to make them better at STEM. There is no reliable scientific evidence that men are naturally better at STEM.

The video concludes with expert analysts explaining that our subtle biases can have effects on the people we hire and promote, and that we have a responsibility to understand the assumptions and errors we make.

2. **Misinformation video: https://youtu.be/UC0OlTXXrZs (1:00 min)**

The video clip is cut from an online video published by a popular blogger, Lauren Chen ('Are Men & Women Different? | Gender Disparities & STEM', https://youtu.be/Niv9fDJ-T0s). It promotes and defends inaccurate information about innate gender differences in STEM ability.

3. **Control video: https://youtu.be/Wf0kwUq64Bc (3:39 min)**

I designed a control video that narrates an animated STEM video about protein synthesis and enzymes. The video was obtained from a popular YouTube channel known as FuseSchool, a global Open Education charity aimed at educating the public on a range of topics through the use of animation videos. The control video was created using two videos from this source, one on "Protein synthesis" and one on "Enzymes" and included both male and female scientists to ensure there is no confounding effect of expert gender.

Job Description

UNIVERSITY OF CAMBRIDGE

Research Assistant/Associate in Concrete Structures (Fixed Term)

A position is open for a Research Associate in the Department of Engineering, University of Cambridge, to undertake research on the applications of new types of reinforced concrete. The project is funded by the EPSRC in collaboration with International Industrial and Academic Partners.

The successful candidate will join a vibrant research team dedicated to advancing sustainable concrete solutions. The key responsibilities and duties are planning, directing and undertaking world-class technical research including: leading-edge experimental and analytical research; advanced analyses and predictions of concrete properties; studies on the complex behaviour of reinforced concrete structures; compiling and analysing data; preparing high quality journal articles and reports; working in collaboration with others; presenting scientific results to diverse audiences.

The successful applicant must have an MEng in Civil Engineering, Structural Engineering or Engineering Materials; experience in the analysis, design and manufacture of reinforced concrete structures and concrete materials; experience with associated research methods and techniques to deliver successful research programmes; excellent communication skills; previous experience of working in teams and with collaborators; experience of managing own workload.

Salary Range: £31,604 - £38,833

Fixed-term: The funds for this post are available for 36 months in the first instance.

Male Job Application

UNIVERSITY OF CAMBRIDGE

Job Application Form

Personal Details

Name	[redacted]
Date of birth	03/02/1992
Gender identity	☒ Male ☐ Transgender male ☐ Non-conforming ☐ Female ☐ Transgender female ☐ Prefer not to say
Ethnic identity	☒ White ☐ Black/African/Caribbean/Black British ☐ Asian/Asian British ☐ Mixed/ multiple ethnic groups ☐ Other Ethnic Group

Qualifications

2012	A-Levels	A*	Mathematics
		A	Further Mathematics
		A	Physics
		A*	Chemistry
2016	Degree	MEng in Civil Engineering (Hons) - University of [redacted]	

Please provide details of all relevant work experience below.

July - August 2014	Summer placement at AECOM: Worked in railway team, helping to design a railway bridge for a high profile bid.
July - August 2015	Summer placement at Arup: Worked in bridge team, helping to assess stability of existing structures and design innovative upgrades to those falling below acceptable standards.
July 2016 - Present	Research assisstant at Arup: Creating computational models to test concrete structures to failure, using this to suggest design improvements.

Please tell us a bit more about yourself below, in no more than five to six sentences.

During my time at university, I was part of the competitive swim team. Outside of work and swimming, I enjoy travelling, reading, and volunteering as a science tutor. In my current position, I have gained valuable research skills, time-management skills, and co-authored two journal articles. I thrive both in teams and independently, and I believe my leadership skills and interpersonal skills will serve me well in this position. I am focused and driven, and I am enthusiastic about this opportunity to work in a dynamic research team at a leading institution.

Female Job Application

UNIVERSITY OF CAMBRIDGE

Job Application Form

Personal Details

Name	[redacted]
Date of birth	03/02/1992

Gender identity

- ☐ Male
- ☐ Transgender male
- ☐ Non-conforming
- ☒ Female
- ☐ Transgender female
- ☐ Prefer not to say

Ethnic identity

- ☒ White
- ☐ Black/African/Caribbean/Black British
- ☐ Asian/Asian British
- ☐ Mixed/ multiple ethnic groups
- ☐ Other Ethnic Group

Qualifications

2012	A-Levels	A*	Mathematics
		A	Further Mathematics
		A	Physics
		A*	Chemistry
2016	Degree	MEng in Civil Engineering (Hons) - University of [redacted]	

Please provide details of all relevant work experience below.

July - August 2014	Summer placement at AECOM: Worked in railway team, helping to design a railway bridge for a high profile bid.
July - August 2015	Summer placement at Arup: Worked in bridge team, helping to assess stability of existing structures and design innovative upgrades to those falling below acceptable standards.
July 2016 - Present	Research assisstant at Arup: Creating computational models to test concrete structures to failure, using this to suggest design improvements.

Please tell us a bit more about yourself below, in no more than five to six sentences.

During my time at university, I was part of the competitive swim team. Outside of work and swimming, I enjoy travelling, reading, and volunteering as a science tutor. In my current position, I have gained valuable research skills, time-management skills, and co-authored two journal articles. I thrive both in teams and independently, and I believe my leadership skills and interpersonal skills will serve me well in this position. I am focused and driven, and I am enthusiastic about this opportunity to work in a dynamic research team at a leading institution.

IV) Measures

Scales without references are face-valid items constructed by me.

Manipulation Check

The video I watched presented evidence about

a) Racism in film
b) Organic farming in North America
c) Protein synthesis in humans
d) Gender stereotypes in STEM

Salzburger State Reactance Scale (Brehm, 1966)

Items were on a 1 (not at all) to 7 (very much) Likert-type scale.

Please indicate your response to the following.

1. Did you perceive the message in the video as a restriction of freedom?
2. Are you frustrated by the message in the video?
3. How much did the message in the video annoy you?
4. To what extent are you offended/disturbed by the message in the video?

Counterarguing (Miller et al. 2013)

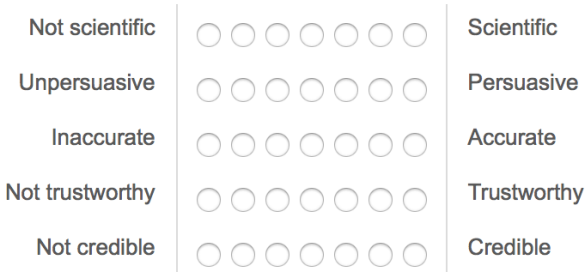
On the following scale, please indicate your response to the messages you heard in the video.

	I agreed with the arguments	.	.	I thought of arguments both for and against	.	.	I thought of arguments against
Response	○	○	○	○	○	○	○

Credibility (Miller et al., 2013; adapted)

Five semantic differential items anchored by bipolar adjectives on a 7-point scale.

On the following scale, please rate the video you watched.



State Logical Thinking (Pietri et al., 2017; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. It made me think logically about the evidence it presented
2. It made me think critically about the evidence it presented

State Perspective Taking (Pietri et al., 2017)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. It made me think about how I would feel to be wrongfully stereotyped
2. It made me think about what it would feel like to be negatively stereotyped

State Empathic Concern

Face-valid items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. It made me feel concerned about gender stereotypes in STEM.
2. It made me feel empathy for women in STEM.

Perceived Increase in knowledge of gender stereotypes in STEM (Pietri et al., 2017, adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. This video increased my knowledge on the topic of gender stereotypes in STEM
2. After watching this video, I feel like I know more about the impact of gender stereotypes in STEM

Male-Agentic Stereotype (Masculine Subscale of the Personality Attributes Questionnaire; Spence et al., 1973)

Items were on a scale from –3 (more true of women) to +3 (more true of men)

Please indicate to what extent the following traits are more true of women or of men.

1. Independent
2. Active
3. Competitive
4. Can make decisions easily
5. Never gives up easily
6. Self-confident
7. Feels superior
8. Stands up well under pressure

Discernment of Gender Stereotypes in STEM

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate your present opinion on the following statements.

1. The scientific consensus is that there are no inborn gender differences in STEM ability.
2. More females are not in STEM because they are naturally interested in other things. [R]
3. On average, men's and women's brains function differently. [R]
4. The scientific consensus is that male brains are naturally better at STEM. [R]
5. Some people use misleading tactics to claim there are differences between men's and women's STEM ability.
6. Gender stereotypes decrease women's chances of being hired in STEM, even when they are equally qualified.

Modern Sexism Scale (Swim et al., 1995)

Please rate how much you currently agree with the following statements.

1. Discrimination against women is no longer a problem in our society.
2. Women often miss out on good jobs due to sexual discrimination. [R]
3. It is rare to see a woman treated in a sexist manner on television.
4. On average, people in our society treat husbands and wives equally.
5. Society has reached a point where women and men have equal opportunities for achievement.
6. It is easy to understand the anger of women's groups in our society. [R]
7. It is easy to understand why women's groups are still concerned about societal limitations of women's opportunities. [R]

8. The news and media show more concern about the treatment of women than is warranted by women's actual experiences.

Affirmative Action (Affirmative Action Subscale of the Neosexism Scale; Tougas et al.,1999; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

*Please state the extent to which you agree or disagree with the following statements. * Note: affirmative action is the policy to increase representation of individuals belonging to groups known to have been discriminated against previously.*

1. If there are no affirmative action programs helping women in employment, they will continue to be unfairly treated.
2. After years of discrimination, it is only fair to set up special programs to make sure women are given a fair and equitable treatment.
3. I favour the implementation of affirmative action programs for women in industries.

Awareness of Bias in STEM (Pietri et al., 2017; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

Please rate how much you currently agree with the following statements.

1. Women in STEM no longer face discrimination based on their gender. [R]
2. Women in STEM are not taken as seriously as their male colleagues.
3. Women in STEM have to work harder than male colleagues to show they are equally competent.

Hostile Sexism (Hostile Sexism Subscale of the Ambivalent Sexism Scale; Glick & Fiske; 1997)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

Please rate how much you currently agree with the following statements.

1. Women exaggerate problems they have at work
2. Women are too easily offended.
3. Most women interpret innocent remarks as being sexist.
4. Many women are actually seeking special favours, under the guise of asking for "equality".

Evaluation of Job Applicant (Moss-Racusin et al., 2012; adapted)

[Hireability]
On a scale from 1 (strongly disagree) to 7 (strongly), please state your response to the following statements:

I think this applicant:

1. Should be interviewed

2. Should be hired
3. Would likely be hired

[Competence]
On a scale from 1 (strongly disagree) to 7 (strongly), please state your response to the following statements:

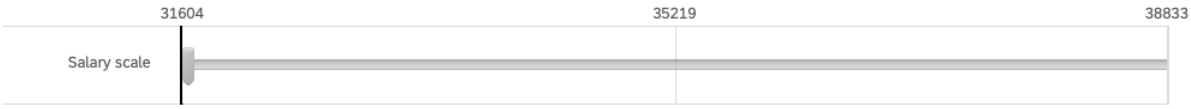
I think this applicant:
1. Is qualified
2. Has the necessary skills for the job
3. Is competent

[Likeable]
On a scale from 1 (strongly disagree) to 7 (strongly), please state your response to the following statements:

I think this applicant:
1. Would make a likeable colleague if hired

[Salary]

Using a sliding scale from £31,604 to £38,833 what salary should the applicant be offered if hired?



Perceived Agency of Job Applicant (Masculine Subscale of the Personality Attributes Questionnaire; Spence et al., 1973; scale adapted)

Items were on a 5-point bipolar semantic scale.

Please rate the applicant as best as you can on the following traits.

9. Independent
10. Active
11. Competitive
12. Can make decisions easily
13. Never gives up easily
14. Self-confident
15. Feels superior
16. Stands up well under pressure

Cognitive Rigidity (Oreg, 2003)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements.

1. I often change my mind. [R]
2. Once I've come to a conclusion, I'm not likely to change my mind.
3. I do not change my mind easily.
4. My views are consistent over time.

Trait Perspective Taking (Perspective Taking subscale of the Interpersonal Reactivity Index, Davis, 1980; Davis, 1983)

Items were on a 5-point scale from 0 (does not describe me well) to 5 (describes me very well)

On a scale from 1 (does not describe me well) to 5 (describes me very well), please rate the following statements about your thoughts and feelings in a variety of situations.

1. I sometimes find it difficult to see things from the "other person's" point of view. [R]
2. I try to look at everybody's side of a disagreement before I make a decision.
3. I sometimes try to understand my friends better by imagining how things look from their perspective.
4. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.[R]
5. I believe that there are two sides to every question and try to look at them both.
6. When I'm upset at someone, I usually try to "put myself in their shoes" for a while.
7. Before criticizing somebody, I try to imagine how I would feel if I were in their place.

Trait Empathic Concern (Empathic concern subscale of the Interpersonal Reactivity Index (IRI); Davis, 1980; Davis, 1983)

Items were on a 5-point scale from 1 (does not describe me well) to 5 (describes me well) scale.

The following statements inquire about your thoughts and feelings in a variety of situations.

1. I often have tender, concerned feelings for people less fortunate than me.
2. Sometimes I don't feel very sorry for other people when they are having problems. [R]
3. When I see someone being taken advantage of, I feel kind of protective towards them.
4. Other people's misfortunes do not usually disturb me a great deal. [R]
5. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. [R]
6. I am often quite touched by things that I see happen.
7. I would describe myself as a pretty soft-hearted person.

Demographics

[Age, collected as a continuous variable]

What is your age? (For example: if you are 30 years old, please enter: 30)

[Education]

What is the highest level of education you have completed?

1. No formal qualification
2. Secondary School
3. College
4. Undergraduate degree (BA/BSc/other)
5. Graduate degree (MA/MSc/MPhil/other)
6. Doctorate degree (PhD/MD/other)

[STEM occupation]

Do you study or work in STEM?

1. Yes
2. No

[Hiring experience]

Do you have any experience in making hiring decisions (i.e. have you been responsible for hiring job candidates)?

1. Yes
2. No

[Ethnicity]

Which of the following best describes your ethnicity/race?

7. White
8. Mixed
9. Black
10. Asian
11. Chinese
12. Latino/Hispanic
13. Other

[Gender]

Which of the following best describes your gender identity?

8. Female
9. Male
10. Transgender male
11. Transgender female
12. Non-conforming/queer
13. Other
14. Prefer not to say

[Religion]

Which of the following best describes your religion?

10. Christian (all denominations)
11. Buddhist
12. Sikh
13. Hindu
14. Jewish
15. Muslim
16. No religion
17. Other
18. Prefer not to say

[Political ideology]

On a scale from very conservative (1) to very liberal (7), how would you describe your political views?

8. Very conservative
9. Mostly conservative
10. Somewhat conservative
11. Neither conservative or liberal
12. Somewhat liberal
13. Mostly liberal
14. Very liberal

Manipulation check

[Applicant gender]

What was the gender of the applicant you reviewed in this study?

1. Female
2. Male
3. Gender minority
4. I don't remember

[Further information]

Is there anything else you would like to tell us or share with us?

This item was an open-response.

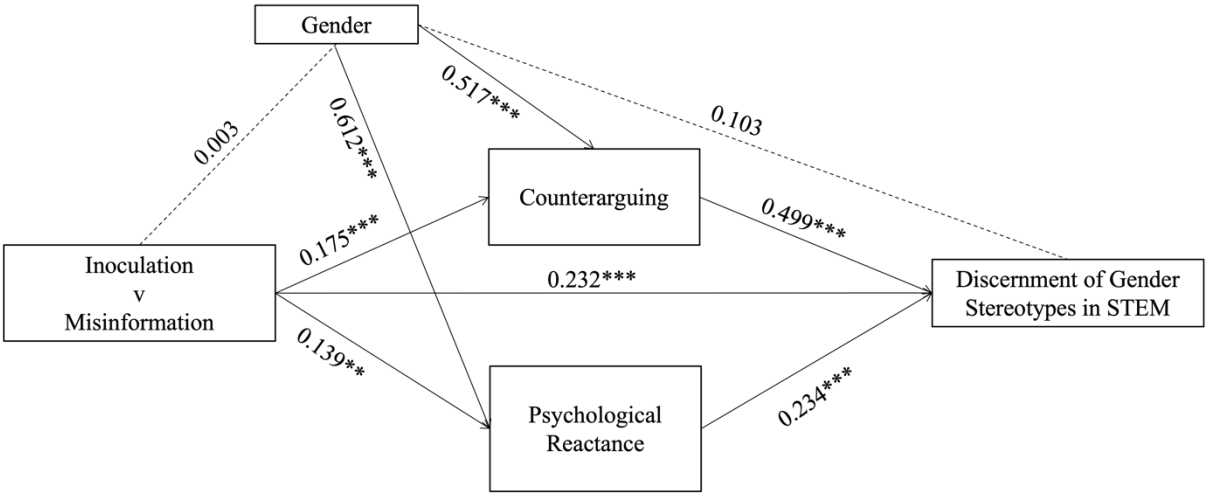
Misinformation warning

Please be aware ***YOU MAY HAVE BEEN EXPOSED TO FALSE INFORMATION IN THIS STUDY. IF YOU VIEWED A VIDEO CLAIMING MEN AND WOMEN HAVE INNATE DIFFERENCES IN STEM CAPABILITIES AND INTERESTS, THIS WAS FALSE INFORMATION.***

V) Supplementary Figures

Figure S3.1

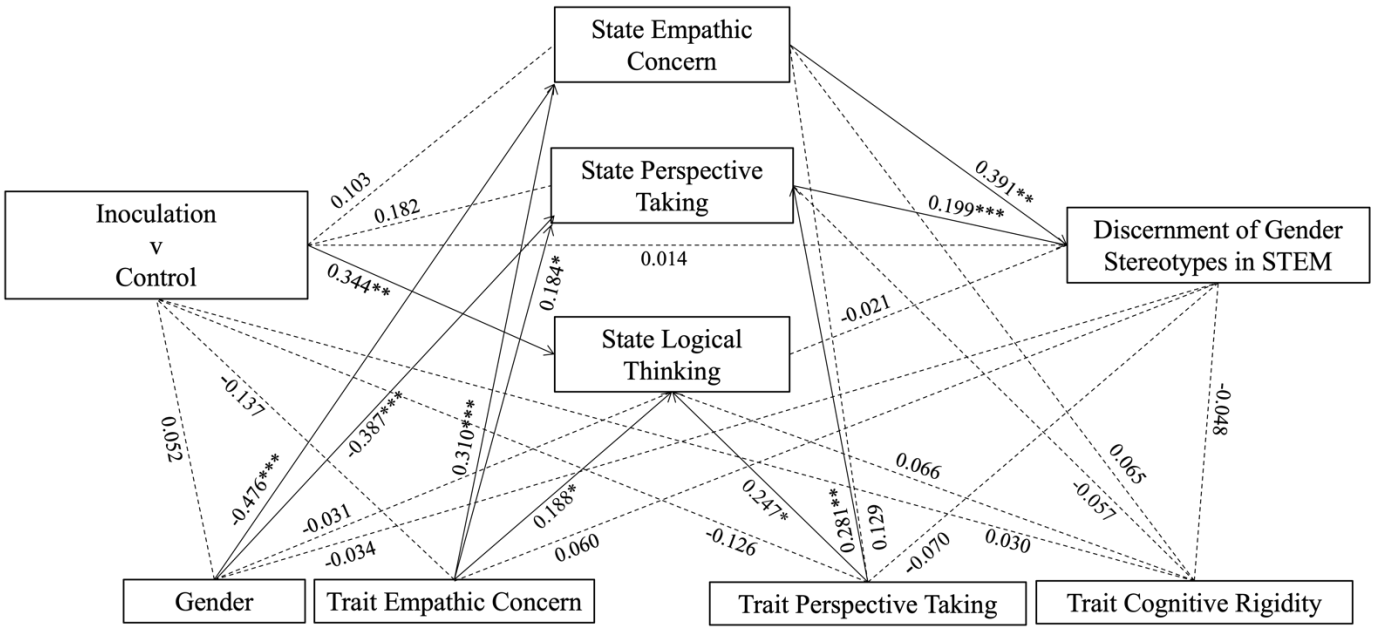
Model testing the indirect effect of the inoculation intervention on gender-STEM stereotype discernment through counterarguing and psychological reactance controlling for gender



Note. Coefficients are standardized. * $p < .05$; **$p < .01$; *** $p < .001$.

Figure S3.2

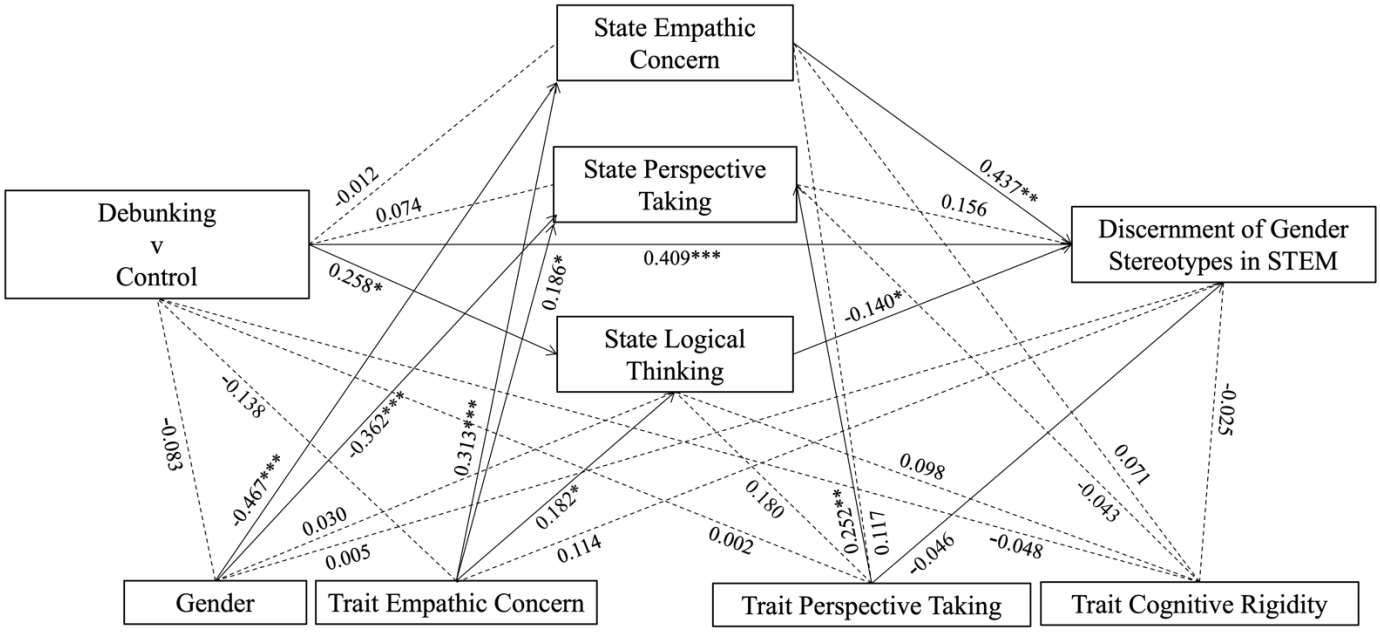
Model testing the indirect effect of the inoculation intervention on gender-STEM stereotype discernment through state empathic concern, state perspective taking, and state logical thinking controlling for gender and related trait variables



Note. Coefficients are standardized. * p < .05; **p < .01; *** p < .001.

Figure S3.3

Model testing the indirect effect of the debunking intervention on gender-STEM stereotype discernment through state empathic concern, state perspective taking, and state logical thinking controlling for gender and related trait variables



Note. Coefficients are standardized. * p < .05; **p < .01; *** p < .001.

VI) Supplementary Tables

Table S3.1

Model estimates for indirect effect of the inoculation intervention on gender-STEM stereotype discernment through counterarguing and psychological reactance controlling for gender

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Inoculation v misinformation → DOGSS	0.232	0.037	6.275	<.001	0.159	0.304
Indirect effects						
Inoculation v misinformation → Counterarguing → DOGSS	0.087	0.028	3.088	0.002	0.032	0.142
Inoculation v misinformation → psychological reactance → DOGSS	0.033	0.015	2.218	0.027	0.004	0.061
Total effects						
Inoculation v misinformation → DOGSS	0.351	0.050	7.089	<.001	0.254	0.448
Total indirect effects						
Inoculation v misinformation → DOGSS	0.120	0.037	3.237	0.001	0.047	0.192

Table S3.2

Model estimates for indirect effect of the inoculation intervention on gender-STEM stereotype discernment through state empathic concern, state perspective taking, and state logical thinking controlling for related trait variables and gender

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Inoculation v control → DOGSS	0.014	0.095	0.148	0.883	-0.172	0.200
Indirect effects						
Inoculation v control → State Empathic Concern → DOGSS	0.040	0.039	1.033	0.302	-0.036	0.116
Inoculation v control → State Perspective Taking → DOGSS	0.036	0.024	1.513	0.130	-0.011	0.083
Inoculation v control → State Logical Thinking → DOGSS	-0.007	0.021	-0.352	0.725	-0.048	0.033
Total effects						
Inoculation v control → DOGSS	0.083	0.077	1.081	0.279	-0.067	0.234
Total indirect effects						
Inoculation v control → DOGSS	0.069	0.062	1.121	0.262	-0.052	0.190

Table S3.3

Model estimates for indirect effect of the debunking intervention on gender-STEM stereotype discernment through state empathic concern, state perspective taking, and state logical thinking controlling for related trait variables and gender.

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Debunking v control → DOGSS	0.409	0.059	6.898	<.001	0.292	0.525
Indirect effects						
Debunking v control → State Empathic Concern → DOGSS	-0.005	0.039	-0.135	0.892	-0.081	0.071
Debunking v control → State Perspective Taking → DOGSS	0.012	0.013	0.866	0.386	-0.015	0.038
Debunking v control → State Logical Thinking → DOGSS	-0.036	0.033	-1.110	0.267	-0.100	0.028
Total effects						
Debunking v control → DOGSS	0.379	0.066	5.760	<.001	0.250	0.508
Total indirect effects						
Debunking v control → DOGSS	-0.030	0.053	-0.568	0.570	-0.133	0.073

Appendix C. Chapter Four

I) Consent Form

Understanding Stereotyping Consent Form
Before you decide to take part in this study, it is important for you to understand why the research is being done and what it will involve. Please read the following and sign if you are in agreement with all of the statements.

What is the purpose of this research?
You are invited to participate in a research study investigating factors that affect the way people interpret and evaluate information about other people. To participate, you must confirm you are (1) 18 years of age or older and (2) fluent in English.

What procedures does this study involve?
This study is in-person and will take approximately 45 minutes to complete. In the first phase of the study, you will receive a stereotype-reduction training. Afterwards, you will be asked a set of questions on your opinions, beliefs, and demographic background.

What are the potential risks or discomforts?
There are no known risks associated with this research study and there are no right or wrong answers. However, you can withdraw from the study at any time. This study has been approved by the Psychology Research Ethics Committee of the University of Cambridge.

Are there benefits to participating in this research?
There are no direct benefits to participating in this research.

What about privacy and confidentiality?
All responses are confidential and anonymous. We will not ask you for your name or identifiable information - you will only be identified by a research number. All data will be published in aggregate form only.

Will I be paid for my participation in this research?
No.

General Data Protection Regulations (GDPR)
We will be using any personal information you give us in order to undertake this study and the University of Cambridge will act as the data controller for this purpose. The legal basis for using your personal information is to carry out academic research in the public interest. We will keep identifiable information about you for as long as necessary for the study. Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible. For further general information about the University of Cambridge's use of your personal data as a participant in a research study, please see https://www.information-compliance.admin.cam.ac.uk/data-protection/research-participant-data.

Who can I contact if I have questions?
If you have further questions about this study or your rights, you may contact the researcher, [name redacted] at [email address redacted].

If you agree to be contacted in the future, please provide your email address below.

Email _____

Signature _____ Date_____

Printed name _____

II) Debriefing form

Thank you for taking part in our study. This study examined strategies to reduce the endorsement and therefore impact of inaccurate gender stereotypes. While stereotype development is normal, it is important to note that it results from basic information processing and associations and can have negative effects on decision-making. Thus, it is important to investigate ways to reduce the effects of gender stereotyping. This research forms part of a PhD dissertation in Psychology at the University of Cambridge.

Your responses are confidential and anonymous. They do not include your name or anything else that could identify you. You have the right to withdraw your data if you so wish.

If you have any concerns or questions regarding this research project, please contact the researcher: [name redacted] at [email address redacted].

III) Materials

Training Presentations

1. The presentation used for the Gender-STEM stereotype-reduction training can be viewed here:

Engineering Gender-STEM Stereotype Reduction Training

2. The presentation used for the unconscious bias training can be viewed here:

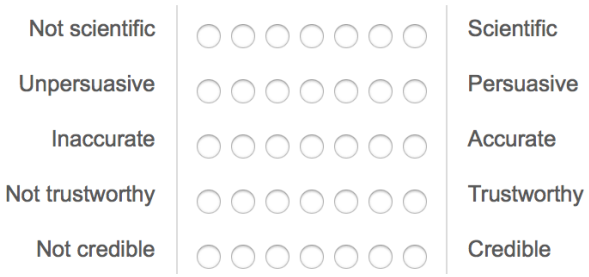
Engineering TA Unconscious Bias Training

IV) Measures

Credibility (Miller et al., 2013; adapted)

Items were on a 7-point bipolar semantic scale.

Please rate the training you received



Salzburger State Reactance Scale (Brehm, 1966)

Items were on a 1 (not at all) to 7 (very much) Likert-type scale

Please indicate your response to the following.

1. Did you perceive the training as a restriction of freedom?
2. Are you frustrated by the training?
3. How much did the training annoy you?
4. To what extent are you offended/disturbed by the training?

State Perspective Taking (Pietri et al., 2017)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

3. It made me think about how I would feel being wrongfully stereotyped
4. It made me think about what it would feel like to be negatively stereotyped

State Empathic Concern

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

3. It made me feel concern for women in STEM.
4. It made me feel concerned about gender stereotypes in STEM.

State Logical Thinking (Pietri et al., 2017; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

1. It made me think logically about the evidence it presented
2. It made me think critically about the evidence it presented

Perceived Increase in knowledge of gender stereotypes in STEM (Pietri et al., 2017, adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements about the video you watched.

3. This training increased my knowledge on gender stereotypes
4. After this training, I feel like I know more about the impact of gender stereotypes

Male-Agentic Stereotype (Masculine Subscale of the Personality Attributes Questionnaire; Spence et al., 1973)

Items were on a scale from –3 (more true of women) to +3 (more true of men)

Please indicate to what extent the following traits are more true of women or of men.

17. Independent
18. Active
19. Competitive
20. Can make decisions easily
21. Never gives up easily
22. Self-confident
23. Feels superior
24. Stands up well under pressure

Discernment of Gender Stereotypes in STEM

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

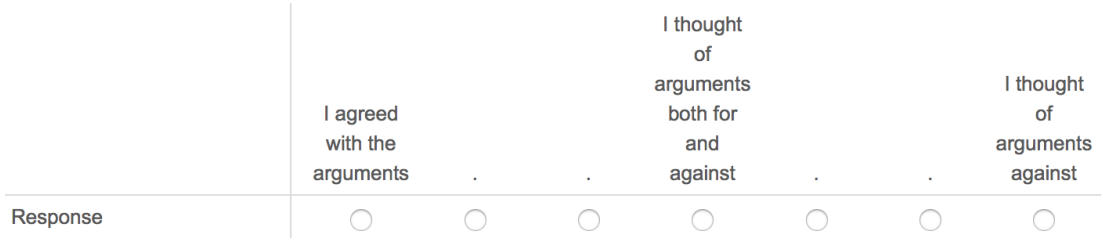
Please rate your opinion on the following statements.

1. The scientific consensus is that there are no inborn gender differences in STEM ability.
2. More females are not in STEM because they are naturally interested in other things. [R]
3. On average, men's and women's brains function differently. [R]
4. The scientific consensus is that male brains are naturally better at STEM. [R]
5. Some people use misleading tactics to claim there are differences between men's and women's STEM ability.
6. Gender stereotypes decrease women's chances of being hired in STEM, even when they are equally qualified.

Counterarguing (Miller et al. 2013)

Items were on a 1 (I agreed with the arguments) to 7 (I thought of argument against) Likert-type scale.

On the following scale, please indicate your response to the messages you heard in the training..



Modern Sexism Scale (Swim et al., 1995)

Please rate how much you currently agree with the following statements.

1. Discrimination against women is no longer a problem in our society.
2. Women often miss out on good jobs due to sexual discrimination. [R]
3. It is rare to see a woman treated in a sexist manner on television.
4. On average, people in our society treat husbands and wives equally.
5. Society has reached a point where women and men have equal opportunities for achievement.
6. It is easy to understand the anger of women's groups in our society. [R]
7. It is easy to understand why women's groups are still concerned about societal limitations of women's opportunities. [R]

Affirmative Action (Affirmative Action Subscale of the Neosexism Scale; Tougas et al.,1999; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

*Please state the extent to which you agree or disagree with the following statements. * Note: affirmative action is the policy to increase representation of individuals belonging to groups known to have been discriminated against previously.*

1. If there are no affirmative action programs helping women in employment, they will continue to be unfairly treated.
2. After years of discrimination, it is only fair to set up special programs to make sure women are given a fair and equitable treatment.
3. I favour the implementation of affirmative action programs for women in industries.

Awareness of Bias in STEM (Pietri et al., 2017; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

Please rate how much you currently agree with the following statements.

1. Women in STEM no longer face discrimination based on their gender. [R]
2. Women in STEM are not taken as seriously as their male colleagues.
3. Women in STEM have to work harder than male colleagues to show they are equally competent.

Hostile Sexism (Hostile Sexism Subscale of the Ambivalent Sexism Scale; Glick & Fiske; 1997)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

Please rate how much you currently agree with the following statements.

1. Women exaggerate problems they have at work
2. Women are too easily offended.
3. Most women interpret innocent remarks as being sexist.
4. Many women are actually seeking special favours, under the guise of asking for "equality".

Trait Perspective Taking (Perspective Taking subscale of the Interpersonal Reactivity Index, Davis, 1980; Davis, 1983)

Items were on a 5-point scale from 0 (does not describe me well) to 5 (describes me very well)

On a scale from 1 (does not describe me well) to 5 (describes me very well), please rate the following statements about your thoughts and feelings in a variety of situations.

1. I sometimes find it difficult to see things from the "other person's" point of view. [R]
2. I try to look at everybody's side of a disagreement before I make a decision.
3. I sometimes try to understand my friends better by imagining how things look from their perspective.

4. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.[R]
5. I believe that there are two sides to every question and try to look at them both.
6. When I'm upset at someone, I usually try to "put myself in their shoes" for a while.
7. Before criticizing somebody, I try to imagine how I would feel if I were in their place.

Trait Empathic Concern (Interpersonal Reactivity Index, Davis, 1980; Davis, 1983)

Items were on a 5-point scale from 1 (does not describe me well) to 5 (describes me well) scale.

The following statements inquire about your thoughts and feelings in a variety of situations.

8. I often have tender, concerned feelings for people less fortunate than me.
9. Sometimes I don't feel very sorry for other people when they are having problems. [R]
10. When I see someone being taken advantage of, I feel kind of protective towards them.
11. Other people's misfortunes do not usually disturb me a great deal. [R]
12. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. [R]
13. I am often quite touched by things that I see happen.
14. I would describe myself as a pretty soft-hearted person.

Cognitive Rigidity (Oreg, 2003)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

On a scale from 1 (strongly disagree) to 7 (strongly agree), please rate how much you agree with the following statements.

1. I often change my mind. [R]
2. Once I've come to a conclusion, I'm not likely to change my mind.
3. I do not change my mind easily.
4. My views are consistent over time.

Gender Identity Centrality (Importance to Identity; Luhtanen & Crocker, 1992; adapted)

Items were on a 1 (strongly disagree) to 7 (strongly agree) Likert scale.

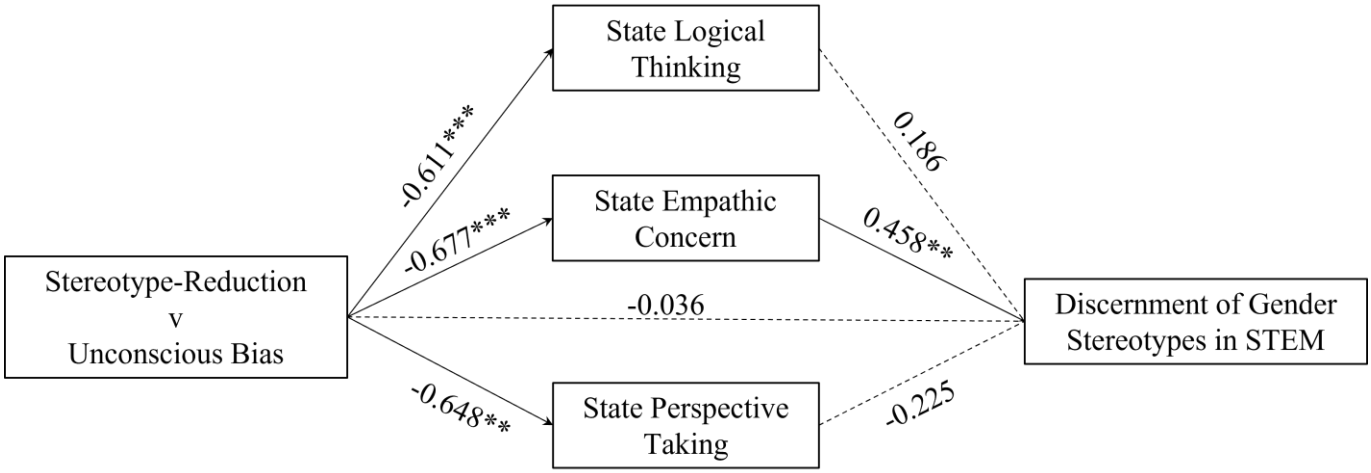
Please rate the statements below.

1. My gender has very little to do with how I feel about myself. [R]
2. My gender is an important reflection of who I am.
3. My gender is unimportant to my sense of what kind of a person I am. [R]
4. In general, my gender is an important part of my self-image.

V) Supplementary Figures

Figure S4.1

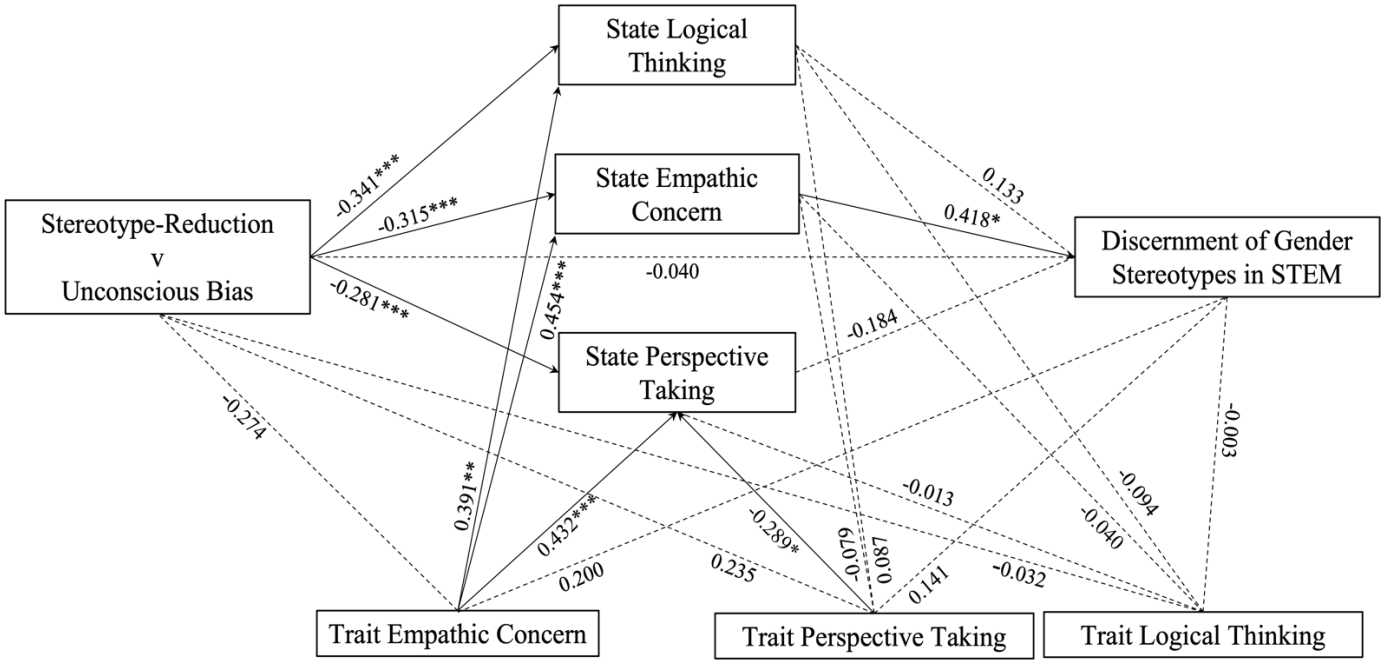
Indirect effect of intervention condition on discernment of gender-STEM stereotypes through state empathic concern, state perspective taking, and state logical thinking



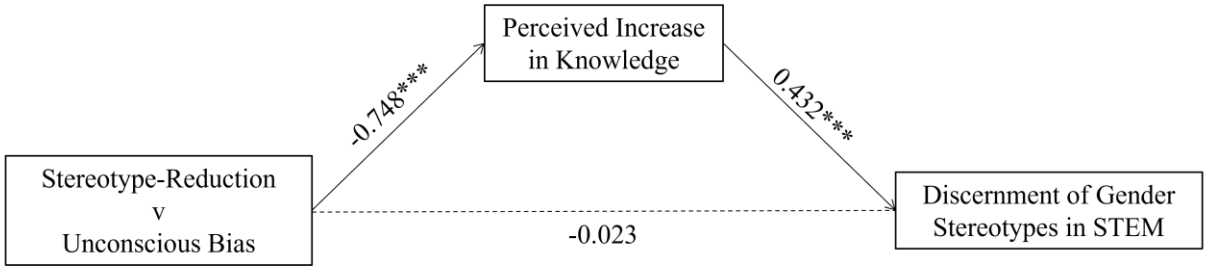
Note. Coefficients are standardized. ** $p < .01$; *** $p < .001$

Figure S4.2

Indirect effect of intervention condition on discernment of gender-STEM stereotypes through state empathic concern, state perspective taking, and state logical thinking controlling for trait empathy, trait perspective taking, and trait logical thinking



Note. Coefficients are standardized. * $p < .05$; **$p < .01$; *** $p < .001$.
Figure S4.3



Note. Coefficients are standardized. * p < .05; **p < .01; *** p < .001.

Supplementary Tables

Table S4.1

Indirect effect of intervention condition on discernment of gender-STEM stereotypes through state empathic concern, state perspective taking, and state logical thinking

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.036	0.159	-0.229	.819	-0.349	0.276
Indirect effects						
Stereotype-Reduction v Unconscious Bias→ State Logical Thinking → DOGSS	-0.113	0.073	-1.554	0.120	-0.257	0.030
Stereotype- Reduction v Unconscious Bias→ State Empathic Concern → DOGSS	-0.310	0.137	-2.268	0.023	-0.578	-0.042
Stereotype-Reduction v Unconscious Bias → State Perspective Taking →DOGSS	0.146	0.097	1.504	0.133	-0.044	0.335
Total effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.314	0.163	-1.925	.054	-0.635	0.006
Total indirect effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.278	0.092	-3.030	0.002	-0.458	-0.098

Table S4.2

Indirect effect of intervention condition on discernment of gender-STEM stereotypes through state empathic concern, state perspective taking, and state logical thinking controlling for trait empathy, trait perspective taking, and trait logical thinking

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.040	0.082	-0.483	.629	-0.201	0.121
Indirect effects						
Stereotype-Reduction v Unconscious Bias→ State Logical Thinking → DOGSS	-0.046	0.043	-1.063	0.288	-0.129	0.038
Stereotype-Reduction v Unconscious Bias→ State Empathic Concern → DOGSS	-0.132	0.068	-1.930	0.054	-0.266	0.002
Stereotype-Reduction v Unconscious Bias → State Perspective Taking →DOGSS	0.052	0.045	1.163	0.245	-0.036	0.139
Total effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.165	0.082	-2.018	.044	-0.326	-0.005
Total indirect effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.126	0.044	-2.868	0.004	-0.211	-0.040

Table S4.3

Indirect effect of intervention condition on discernment of gender-STEM stereotypes through state empathic concern, state perspective taking, and state logical thinking

Effect	*β*	SE	*z*	*p*	95% CI	
					LL	*UL*
Direct effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.023	0.160	-0.141	.888	-0.337	0.291
Indirect effects						
Stereotype-Reduction v Unconscious Bias→ Perceived Increase in Knowledge→ DOGSS	-0.323	0.088	-3.676	<.001	-0.496	-0.151
Total effects						
Stereotype-Reduction v Unconscious Bias → DOGSS	-0.346	0.163	-2.118	.034	-0.666	0.026

Appendix D. Chapter Five

I) Consent Form

STUDY INFORMATION AND INFORMED CONSENT
Traits & Objects: How do we perceive them?

You are being invited to take part in a research study. Before you decide whether to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

What is the purpose of this study?
You are invited to participate in a research study investigating the way people perceive other people and objects that children play with. To participate, you must be (1) 18 years of age or older and (2) fluent in English. **You cannot complete this survey on a mobile device.**

Do I have to take part?
Taking part is entirely voluntary, and refusal or withdrawal will involve no penalty or loss, now or in the future. You are free to withdraw from the experiment at any time without explaining why.

What will the study involve?
This study is an online survey and will take you approximately 20 minutes to complete. You will be asked questions about your perceptions of traits in others. Then, you will view a series of objects that children play with and be asked questions about your opinions of these items. Lastly, you will complete a set of questions on your demographic background. After completing the study, you can choose to enter your email address into a draw for a £20 Amazon gift card.

Are there any possible disadvantages or risks in taking part?
There are no known risks associated with participating in this study.

Confidentiality – who will have access to my personal data?
Personally identifiable data will not be made available outside of the study team. We will be using any personal information you give us in order to undertake this study and the University of Cambridge will act as the data controller for this purpose. The legal basis for using your personal information is to carry out a task (i.e. academic research) in the public interest. We will keep identifiable information about you for as long as necessary for the study. Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will do our best to only use the minimum personally-identifiable information possible.
For further general information about the University of Cambridge's use of your personal data as a participant in a research study, please see https://www.information-compliance.admin.cam.ac.uk/data-protection/research-participant-data.

What will happen to the study results?

They will be kept for a minimum of 10 years securely in the Cambridge Department of Psychology in accordance with good research practice. Results from groups of individuals, without any means of identifying the individuals involved, may be presented at conferences and written up in journals.

Who is organising the research?
The study is organised in the Department of Psychology, University of Cambridge, by [name redacted] and [name redacted] Questions or concerns can be directed to them at [email address redacted] or [email address redacted].

Ethics Committee Approval
The project has been reviewed by the Psychology Research Ethics Committee of the University of Cambridge.

II) Debriefing form

Thank you for participating in this study. We will now explain a bit more about the background and aims of the study.

People perceive the world through a prism of social groups. Different groups are usually assigned different characteristics to help us differentiate between them. Cognitively, it makes our social worlds more organised and simplified, which makes them "easier" to navigate. In this study, we investigate which traits and objects people associate with males and/or females to advance our knowledge about how people use the prism of gender to navigate the world.

If you have any questions about the study, feel free to contact the researchers, [name redacted] and [name redacted] at [email address redacted] or [email address redacted].

Thank you once again for your help with this research. If you would like to be entered into the draw to win a £20 Amazon gift card, please enter your email address below. The winner will be contacted no later than 30.12.2020.

III) Measures

Gender-Trait Stereotypes

Items were on a –3 (*more true of women*) to +3 (*more true of men*) 7-point scale.

Please indicate to what extent you think each of the following traits are, in general, more true of women or of men.

1. Not at all independent – Very independent
2. Very passive – Very active
3. Not at all competitive – Very competitive
4. Likely to give up very easily – Likely to never give up easily
5. Not at all self-confident – Very self-confident
6. Likely to feel very inferior – Likely to feel very superior
7. Not at all able to devote self to others - Able to devote self completely to others
8. Not at all helpful to others – Very helpful to others
9. Not at all kind – Very kind
10. Not at all aware of feelings of others – Very aware of feelings of others
11. Not at all understanding of others – Very understanding of others
12. Very cold in relations with others – Very warm in relations with others
13. Not at all logical – Very logical
14. Not at all analytical – Very analytical
15. Not at all ambitious – Very ambitious
16. Not at all decisive – Very decisive
17. Not at all able to create or invent new things – Able to create or invent new things
18. Not at all creative – Very creative
19. Not at all innovative – Very innovative
20. Not at all intelligent – Very intelligent
21. Not at all level-headed – Very level-headed
22. Not at all organized – Very organized
23. Not at all smart – Very smart
24. Not thorough in handling details – Very thorough in handling details
25. Not willing to accept new ideas – Very willing to accept new ideas
26. Not at all likeable – Very likeable
27. Not at all mathematical – Very mathematical
28. Not at all honest – Very honest
29. Not able to communicate effectively – Able to communicate effectively
30. Not at all driven – Very driven
31. Not at all charismatic – Very charismatic
32. Not at all collaborative – Very collaborative
33. Not at all curious – Very curious
34. Not at all able to inspire others – Able to inspire others
35. Not at all able to convince others – Able to convince others
36. Not at all able to assess others' strengths and weaknesses - Able to assess others' strengths and weaknesses
37. Not at all strategic – Very strategic
38. Not at all able to encourage others – Able to encourage others
39. Not at all empathic – Very empathic
40. Not at all able to persuade others – Able to persuade others

Male-Trait Stereotypes

Semantic differential items anchored by bipolar adjectives on a 7-point scale.

In general, men are…

1. independent
2. active
3. competitive
4. never give up easily
5. self-confident
6. feel superior
7. able to devote self completely to others
8. helpful to others
9. kind
10. aware of feelings of others
11. understanding of others
12. warm in relations with others
13. logical
14. analytical
15. ambitious
16. decisive
17. able to create or invent new things
18. creative
19. innovative
20. intelligent
21. level-headed
22. organized
23. smart
24. thorough in handling details
25. willing to accept new ideas
26. likeable
27. mathematical
28. honest
29. able to communicate effectively
30. driven
31. charismatic
32. collaborative
33. curious
34. able to inspire others
35. able to convince others
36. able to assess others' strengths and weaknesses
37. strategic
38. able to encourage others
39. empathic
40. able to persuade others

Female-Trait Stereotypes

Semantic differential items anchored by bipolar adjectives on a 7-point scale.

In general, women are…

1. independent
2. active
3. competitive
4. never give up easily
5. self-confident
6. feel superior
7. able to devote self completely to others
8. helpful to others
9. kind
10. aware of feelings of others
11. understanding of others
12. warm in relations with others
13. logical
14. analytical
15. ambitious
16. decisive
17. able to create or invent new things
18. creative
19. innovative
20. intelligent
21. level-headed
22. organized
23. smart
24. thorough in handling details
25. willing to accept new ideas
26. likeable
27. mathematical
28. honest
29. able to communicate effectively
30. driven
31. charismatic
32. collaborative
33. curious
34. able to inspire others
35. able to convince others
36. able to assess others' strengths and weaknesses
37. strategic
38. able to encourage others
39. empathic
40. able to persuade others

IV) Supplementary Results

Ward et al. (2006) Model

Bipolar Scale

First, the Ward et al. (2006) revised PAQ model was tested with the bipolar-scale responses. Ward and colleagues (2006) re-specified the original PAQ factor structure, eliminating the following items from the masculine (M) and feminine (F) subscales: "can make decisions easily" (M), "gentle" (F), and "emotional" (F). In an earlier study (McCreary & Steinberg, 1992), "can make decisions easily" and "emotional" were also found to be misspecifiedl. Our CFA revealed an unacceptable model fit, consistent with findings from Curran and Warber (2011), X^2 (53, N = 121) = 118.608, p <.001 (CFI = .896, TLI = .871, RMSEA = .102, SRMR = 0.117). The factor correlation was -.343, p < .001. The internal reliability of the communal subscale was excellent (α = 0.915), while the internal reliability of the agentic subscale was acceptable at best (α = 0.679). Ward et al. (2006) found an α of 0.73 for the agentic scale, and concluded the agentic subscale of the PAQ may not be an adequate measure of agency due to correlations with other indicators and factors (factorial invariance).

Examination of the factor loadings indicated that "independent" did not load onto a factor with statistical significance and "never gives up easily" loaded onto the agentic factor to a lesser extent than the other items. The residual correlations between variables suggested substantial issues. As the RMSEA is based on the residual correlations, these deviations are driving the unacceptable RMSEA value in this model. Twenty-four residuals were equal to or exceeded 0.10 in the standardised residual covariance matrix, suggesting misspecificity and issues of construct validity (see Appendix D; Supplementary Figures).

Additionally, the modification indices indicated that four items loaded on both factors, indicating multidimensionality of the factor structure. Specifically, "feels superior", "active", and "independent" cross-loaded onto the communal factor, and "able to devote self completely to others" cross-loaded onto the agentic factor.

Further, there were several covariances between items on the same latent factor, which means they share additional relationships on top of their loadings on the agentic or communal factors. On the communal subscale, there were six: "helpful" and "warm", "able to devote self completely to others" and "aware of feelings of others", "helpful" and "aware of feelings of others", "aware of feelings of others" and "understanding of others", and "helpful to others" and "kind". On the agentic subscale, there was one: "active" and "never

gives up easily". These covariances may be explained by the items being more closely related to each other than the other items on their specified factor.

Lastly, the modification index for residual covariances showed additional covariance between items that do not share the same latent factor: "feels superior" and "warm", "active" and "kind", "self-confident" and "helpful", and "independent" and "helpful". This further demonstrates the multidimensionality of the factors and indicators.

The modification indices indicate many parameters (cross-loadings and residual covariances) that would significantly improve the model fit if freed in the model. However, these values are empirically driven, based solely on the small sample data, and are therefore not advised to include unless for very good theoretical reasoning. Taken together, and to avoid 'overfitting', an EFA may be more appropriate.

Difference-Score Scale

Next, a CFA on the Ward et al. (2006) model using the difference-score data was examined. Compared to the Ward et al. (2006) bipolar scale, the Ward et al. (2006) difference-score scale demonstrated a slightly improved, but still unacceptable, model fit: X^2 (53, N = 121) = 107.750, $p < .001$ (CFI = .895, TLI = .870, RMSEA = .096, SRMR = .093). The factor correlation was -.506, $p < .001$. The internal reliability of the communal subscale was excellent (α = 0.88), while the internal reliability of the agentic subscale was acceptable at best (α = 0.72).

Consistent with the bipolar scale, the factor loadings indicated that "never gives up easily" did not load onto a factor with statistical significance. In contrast to the bipolar scale, "independent" loaded significantly, but to a lesser extent than the other variables. The standardised residual covariance matrix indicated several issues. Twenty residuals, four fewer than the bipolar scale, were equal to or exceeded 0.10 in the standardised residual covariance matrix, suggesting misspecificity and issues of construct validity (see Appendix D; Supplementary Figures).

The modification indices indicated that three items loaded on both factors, indicating multidimensionality of the factor structure. Specifically, "feels superior" and "independent" cross-loaded onto the communal factor, and "able to devote self completely to others" cross-loaded onto the agentic factor. Again, a slight improvement from the bipolar scale.

Further, there were several covariances between items on the same latent factor, suggesting they shared additional relationships on top of their factor loadings. On the communal subscale, there were three problematic covariances: "able to devote self

completely to others" and "warm", "warm" and "aware of feelings of others", and "warm" and "aware of feelings of others". On the agentic subscale, there were also three: "independent" and "active", "competitive" and "never gives up easily", "active" and "feels superior". These covariances may be explained by the items being more closely related to each other than the other items on their specified factor.

Lastly, the residual covariances between items not on the same latent factor were difference than those found with the bipolar scale: "self-confident" and "able to devote self completely to others", self-confident" and "understanding of others", "independent" and "aware of feelings of others", and "competitive" and "kind".

While the modification indices indicate the modifications that would significantly improve the model fit if we incorporated the parameters into the model, it also needs to make theoretical sense to free the parameters. Further, these indices are subject to sampling bias, so, again, should only be included with very good theoretical reasoning. Therefore, due to the issue of multidimensionality, and to avoid 'overfitting', results suggest an EFA may be more appropriate.

V) Supplementary Figures

Figure S5.1

Standardised Residual Covariance Matrix of Agentic and Communal Items based on the Ward et al. (2006) Model

	1	2	3	4	5	6	7	8	9	10	11	12
1 - Independent		0.15	0.00	0.08	0.06	0.05	0.24	0.37	0.35	0.26	0.19	0.23
2 - Active	0.23		0.05	0.18	0.04	0.01	0.08	0.16	0.28	0.16	0.06	0.11
3 - Competitive	0.09	0.10		0.04	0.02	0.02	0.13	0.01	0.05	0.02	0.05	0.07
4 - Never gives up	0.14	0.04	0.24		0.05	0.04	0.07	0.13	0.20	0.13	0.12	0.09
5 - Self-confident	0.01	0.01	0.04	0.03		0.00	0.07	0.16	0.10	0.00	0.00	0.11
6 - Feels superior	0.05	0.09	0.02	0.13	0.01		0.23	0.08	0.07	0.15	0.17	0.20
7 - Devote	0.14	0.02	0.06	0.14	0.13	0.09		0.00	0.03	0.08	0.03	0.03
8 - Helpful	0.20	0.08	0.03	0.14	0.04	0.02	0.04		0.05	0.06	0.04	0.06
9 - Kind	0.16	0.16	0.16	0.16	0.03	0.02	0.00	0.00		0.05	0.03	0.02
10 - Aware of feelings of others	0.28	0.00	0.08	0.17	0.02	0.05	0.05	0.03	0.01		0.05	0.02
11 - Understanding of others	0.13	0.04	0.08	0.11	0.06	0.08	0.03	0.00	0.00	0.01		0.01
12 - Warm	0.14	0.08	0.02	0.07	0.02	0.17	0.09	0.03	0.05	0.08	0.05	

Note. Values on the top right of the diagonal correspond to the bipolar measure and values on the bottom left of the diagonal correspond to the difference-score measure.

VI) Supplementary Tables

Table S5.1

EFA Total Factor Loadings of 40 Items

	Factor 1	Factor 2	Factor 3	Uniqueness
Helpful	**0.837**	-0.016	0.113	0.282
Understanding of others	**0.811**	-0.064	0.001	0.342
Warm	**0.804**	-0.021	-0.006	0.356
Kind	**0.772**	0.090	0.086	0.367
Empathic	**0.762**	-0.186	-0.109	0.386
Aware of others' feelings	**0.758**	0.038	-0.088	0.421
Devote	**0.756**	-0.019	-0.165	0.412
Collaborative	**0.712**	0.059	0.038	0.478
Organised	**0.677**	0.030	0.096	0.521
Honest	**0.575**	0.284	-0.018	0.570
Encouraging of others	**0.557**	0.077	0.094	0.659
Accepts new ideas	**0.551**	0.081	-0.012	0.685
Able to assess others	**0.538**	0.190	-0.002	0.660
Good communicator	**0.521**	0.038	0.201	0.669
Details	**0.506**	-0.043	0.192	0.705
Feels superior	-0.494	0.321	0.358	0.489
Able to inspire others	0.478	0.365	0.152	0.549
Creative	0.469	-0.135	0.369	0.651
Curious	0.468	-0.070	0.397	0.624
Likeable	0.412	0.242	-0.141	0.765
Competitive	-0.359	**0.607**	0.216	0.409
Smart	0.314	0.459	0.207	0.561
Independent	0.301	0.290	0.058	0.798
Intelligent	0.294	0.453	0.280	0.523
Level-headed	0.284	**0.584**	-0.344	0.575
Decisive	-0.267	0.321	0.312	0.683
Self-confident	-0.262	0.432	0.390	0.509
Strategic	-0.211	0.495	0.175	0.641
Driven	0.176	0.250	0.207	0.821
Innovative	0.176	0.365	0.393	0.575
Never gives up easily	0.112	0.127	0.466	0.710
Logical	0.101	**0.696**	-0.151	0.542
Able to convince others	0.098	-0.063	**0.692**	0.531
Able to create/invent new things	0.085	0.661	0.226	0.399
Able to persuade others	0.057	-0.039	**0.692**	0.531
Mathematical	-0.042	**0.814**	-0.151	0.397
Ambitious	0.041	**0.565**	0.171	0.583
Active	-0.022	0.356	0.298	0.718
Charismatic	-0.018	0.292	0.383	0.697
Analytical	-0.017	0.712	-0.052	0.516

Note. Applied rotation method is oblique oblimin. Factor loadings arranged by factor loading size on Factor 1. Bold items indicate significant factor loading above 0.5. Items measured on scale from -3 (*more true of women)* to +3 (*more true of men)*

Table S5.2

EFA Factor Loadings Using Difference-Score Scale

	Factor 1	Factor 2	Factor 3	Uniqueness
Understanding of others	**0.859**			0.239
Empathic	**0.779**			0.373
Helpful	**0.741**			0.436
Aware of others' feelings	**0.741**			0.346
Devote	**0.685**			0.456
Good communicator	**0.655**			0.565
Kind	**0.642**			0.584
Warm	**0.642**			0.584
Organised	**0.631**			0.548
Collaborative	**0.614**			0.613
Encouraging of others	**0.598**			0.562
Accepts new ideas	**0.593**			0.591
Creative	**0.566**			0.653
Details	**0.551**			0.620
Intelligent	**0.547**			0.537
Able to inspire others	**0.541**			0.599
Analytical		**0.779**		0.350
Logical		**0.696**		0.503
Mathematical		**0.674**		0.473
Strategic		**0.577**		0.661
Able to create/invent new things		**0.560**		0.594
Innovative		**0.545**		0.656
Self-confident			**0.673**	0.315
Persuade others			**0.544**	0.709
Able to convince others			**0.516**	0.737
Decisive			**0.514**	0.567
Active			**0.506**	0.718
Independent				0.705
Competitive				0.572
Never gives up				0.784
Feels superior				0.525
Ambitious				0.709
Level-headed				0.798
Smart				0.653
Likeable				0.623
Honest				0.718
Driven				0.713
Charismatic				0.778
Curious				0.767
Able to assess others				0.764

Note. Applied rotation method is oblimin.