Gender in language and gender in the social mind

Hualin Xiao

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure

Gender in Language and Gender in the Social Mind

Soutenue par
Hualin XIAO
Le 15 décembre 2021

Ecole doctorale n° 540
Lettres, Arts, Sciences humaines et sociales

Spécialité
Sciences cognitives

Composition du jury :

Heather, BURNETT DR, Université de Paris	*Présidente* *Examinatrice*
Judit, GERVAIN Professeure, Università di Padova	*Rapportrice*
Pascal, GYGAX Professeur, Université de Fribourg	*Rapporteur*
Salvador, MASCARENHAS Enseignant-chercheur contractuel, ENS	*Examinateur*
Sharon, PEPERKAMP DR, École normale supérieure	*Co-directrice de thèse*
Brent, STRICKLAND CR, École normale supérieure	*Co-directeur de thèse*

ENS | PSL

Acknowledgements

Four years ago, I arrived in Paris alone, carrying the largest suitcase I ever had. Before that, I knew very little about the feeling of carrying a Chinese suitcase in a foreign land, but I did know what I came here for. Now it's about time to pack again. I still have the suitcase I brought from my homeland, only that the items to fill in have changed.

I'm able to complete the four-year PhD thanks to my advisors Brent Strickland and Sharon Peperkamp. They not only initiated me into Psychology, but more importantly helped me develop another way of thinking that is rational and scientific. This may sound surprising, but it is a big challenge to someone with only a background in English language and literature to acquire this kind of reasoning as it is contrary to intuition or imagination-oriented thinking. Their knowledge, skills, attitudes towards science, and ways of doing things have broadened my horizon. As a supervisor, Sharon may give the impression that she goes too much into details (or just perfectionist), but after working four years with her I finally understand that it is a quality I should have to compensate for my impatient and clumsy personality. Now I feel jealous that she was born perfectionist. Brent, on the other hand, always has his eyes on the bigger picture. He is the one who told me on our first meeting to think about my future career and research agenda. I'm impressed by how efficiently these two minds cooperate to foster a PhD. They have also helped me financially. I am grateful to Brent for finding the research assistant job at the School of Collective Intelligence, UM6P (Morroco), and to Sharon for paying me with her ANR funding, which together supported the last year of my PhD. They have supported me in many other ways that I can't specify here but I

keep deeply in my heart (not just in my mind, as a cognitive scientist may prefer).

I managed to live in Paris without getting totally lost thanks to my beloved life partner Timothée. We met on a weekend hiking, and since then we've been exploring the Alps together. He makes bad jokes, he cooks standard sandwiches, his singing is beyond description, but he has a gentle heart and a curious mind that I'm lucky enough to able to appreciate. I thank him for being caring and supportive during my PhD, and for keeping me on the right track in life, work, and in the mountains. I would also like to thank his family for always being nice and considerate to me.

To my parents, I'm a "rebel". I don't have a stable job, living somewhere 8000 km far from home, and I don't have plans for settling down and starting a family. They sacrificed a lot for the education of my brother, my sister and myself. They certainly hoped that we could all graduate from college and live what they would depict as life. I'm not sure if they would see completing a PhD as an achievement, but I'm thankful that they didn't tell me to stop doing what I do. I feel sorry that I haven't been able to see them due to the Covid crisis and that I constantly let them worry about me.

Last but not least, my gratitude goes to my lab IJN and LSCP where I've had the chance to meet with the smartest and kindest colleagues. I'm especially grateful to Cathal O'Madagain (currently Scientific Director of the SCI, UM6P) for generously offering me the research assistant job which financially helped me through the fourth year of my PhD. I'm also thankful for him being so helpful and kind whenever I seek advice from him. I'd like to thank Léonard Guillou for answering a lot of my questions and always in a nice and pedagogical way. I'd like to express my gratitude to my

collaborators Antoine Marie and Alexandre Cremers for their constructive suggestions. I thank all the helpful colleagues and friends: Claire Kabdebon, Camille Straboni, Guido Löhr, Mauricio Martins, Andreas Falck, Takuya Niikawa, Gerda Melnik, and especially Yiyun Liao who I met in college and has inspired and supported me in numerous ways. My final acknowledgements go to my jury members: Judit Gervain, Pascal Gygax, Heather Burnett, and Salvador Mascarenhas for reading my thesis and sharing with me their thoughts.

Table of Contents

Chapter 1. Introduction

Gender has multiple facets. When used to refer to biological sex, it is a term for distinguishing a male from a female person with regard to their anatomical and hormonal differences. Put in the context of society, gender is a product of social norms and interactions embodied by the different roles, qualities, and behaviors associated with men and women (Eagly & Wood, 2016). Gender is also an important concept for human languages. Many of the world's languages have a grammatical gender system in which nouns are categorized into different gender classes based on features like sex, animacy, shape and size etc., and following this categorization, words appearing with the nouns, such as adjectives, articles, verbs and pronouns change their forms accordingly (Aikhenvald, 2016; Corbett, 1991; Gygax et al., 2019). The meanings and implications of gender vary depending on which facet of the notion one refers to. In this dissertation, I focus on the three aspects of gender and present three empirical studies on how each aspect as well as interactions between them shape various cultural landscapes.

The first study (Chapter 2) investigates the relationship between language and thought, in particular the influence of grammatical gender on the conceptualization of objects. The Neo-Whorfian hypothesis postulating that the structure of one's native language impacts the way one thinks about the world (Whorf, 1956) has earned empirical support among cognitive scientists from various perspectives such as the influence of color terms on color perception (e.g. Davies & Corbett, 1997; Gilbert et al., 2006; Thierry et al., 2009), and the impact of space and time framing in language on speakers' spatial reasoning and conceptualization of time (e.g. Boroditsky, 2001; Haun et al., 2006; Levinson, 2003). However, research remains controversial on whether the grammatical gender of nouns denoting genderless objects prompts language users to associate different gender properties with the objects. The study described in Chapter 2 is an

attempt to answer this question. To ensure that any positive/null results obtained are not due to unreliable research methods, I adopt an innovative approach by crowdsourcing the materials for Experiment 1, and I followed the open science initiative by submitting the study as a registered report[1] for peer review before we actually run the experiments. As this report is currently under review, I present in Chapter 2 in detail the design of two psycholinguistic experiments (without results), in addition to two pilot experiments (with results).

The second study (Chapter 3) examines the relationship between grammatical gender and mental representations of human referents. It focuses on the controversial use of masculine generics in French and its consequences for the underrepresentation of women in the mind of language users. Through two experiments, I compare the masculine form (e.g. *les musiciens* 'the musicians$_{masc}$') to two gender-fair forms (double-gender: *les musiciens et musiciennes* 'the musicians$_{masc}$ and musicians$_{fem}$'; and middot: *musicien·ne·s*) across professions that have balanced and biased gender distributions. I ask three specific questions: 1) whether the two gender-fair forms differ from each other with regard to their effects on mental representations of gender distributions; 2) whether the effect of language form is moderated by the gender stereotype of a professional group; 3) and finally whether or not the representations induced by these language forms are *consistent* perceptions of real-world gender ratios.

The last study (Chapter 4) is about people's moral attitudes on gender equality between men and women and how it relates to their trust in scientific evidence of gender discrimination in academia. Moral attitudes toward a controversial issue have been shown to affect individuals' processing of new information, such that they tend to selectively assimilate or reject evidence

[1] Registered reports are defined by the Royal Society as follows: "Registered reports are a format of empirical article where a study proposal is reviewed before the research is undertaken. Pre-registered proposals that meet high scientific standards are then provisionally accepted before the outcomes are known, independently of the results." (see https://royalsociety.org/blog/2016/11/registered-reports-what-are-they-and-why-are-they-important/).

depending on whether it is consistent with their pre-existing attitudes or not (Lord et al., 1979). In Chapter 4, I present six experiments exploring the effects of individuals' moral commitment to gender equality on their evaluations of research summaries on sex-based hiring bias in STEM fields.

In this introductory chapter, I first provide an overview of research on language and thought with respect to object conceptualization, as the empirical evidence pertaining to this topic is less conclusive. Then I introduce the investigations on the role of language in the mental representations of persons, and finally I cover the study of the relationship between moral attitudes and trust in science.

Gender in Language and Its Influences on Mental Representations of Gender

More than a core characteristic of human beings, gender is also a productive feature of human languages. Depending on how gender is encoded in the language structure, the world's languages can be roughly divided into a few categories: grammatical gender languages (e.g. French, German), natural gender languages (e.g. English, Swedish), and genderless languages (e.g. Chinese, Finnish) (Gygax et al., 2019; Stahlberg et al., 2007). In grammatical gender languages, every noun, no matter if it refers to persons, objects or abstract concepts, is assigned to a gender class. The number of gender class varies across languages. For example, among the 256 languages documented by the World Atlas of Language Structures online (https://wals.info/) based on the work of Corbett (1991), 144 do not have grammatical gender, 50 have a two-gender system (e.g. French: masculine and feminine), 26 a three-gender system (e.g. German: masculine, feminine and neuter), 12 a four-gender system (e.g. Dyirbal: gender I , II, III, and IV) and 24 languages have five and more genders (e.g. Swahili: gender I - VII). The grammatical gender of nouns determines the form of other lexical categories appearing in the same sentence including

verbs, adjectives, articles, and pronouns. For instance, in French, the word *student* has two forms *étudiant* and *étudiante* respectively for a male and a female student. In many cases, a French speaker cannot avoid indicating the sex of human referents in their phrases (except when they mean to do so) as the gender information of the referent is marked on the words they use. If the author writes *Elle est une bonne étudiante* 'She is a good student', the reader would not doubt that it is a female student the author is describing. Similarly, nouns denoting inanimate entities have gender markings and determine the form of other words appearing with them, as in *une petite table* 'a_{fem} $small_{fem}$ table' and *un petit bureau* 'a_{masc} $small_{masc}$ desk'. Here, *a* and *small* were shown in their feminine and masculine forms, respectively, modifying the feminine noun *table* and the masculine noun *desk*. The gender assignment of person nouns in grammatical gender languages mostly corresponds to the sex of the referent, while that of object nouns seems to be arbitrary, if not based on shape, size or other semantic features (Corbett, 1991).

Natural gender languages do not group nouns into gender classes, except that personal pronouns (e.g. *he* and *she* in English) and some person nouns (e.g. *waiter/waitress*, *actor/actress*) distinguish between the male and female forms. In genderless languages, finally, the sex information of referents is conveyed through lexical means, hence the absence of grammatical gender marking for nouns and pronouns (Gygax et al., 2019; Stahlberg et al., 2007).

Following the Neo-Whorfian hypothesis that the structure of a language influences the way representations of the world are constructed in the minds of the speakers, one would argue that the grouping of nouns into masculine or feminine gender class should lead language users to associate male or female gender properties with the referents of the nouns. Research on this topic can be divided into two lines. One line investigates person nouns whose grammatical gender has a sex-based semantic underpinning as it overlaps the natural gender of the referents. The other

line of research focuses on inanimate nouns (e.g. objects, concepts) whose grammatical gender does not overlap natural gender and has no semantic underpinnings related to sex. Overall, research on person nouns has provided converging evidence of grammatical gender influencing the mental representations of person referents, while results from the second line of research seem to be contradictory.

Does Grammatical Gender Affect Our Conceptualization of Objects?

Previous research on the effects of *linguistic relativity* has provided some evidence supporting the weak version of the Whorfian hypothesis postulating that language influences thought (the Neo-Whorfianism), but not the strong version that language determines thought (see Wolff & Holmes, 2011; Zlatev & Blomberg, 2015). The hypothesized language effect has been found from many perspectives such as color terms and color perception (Davies & Corbett, 1997; Gilbert et al., 2006; Thierry et al., 2009; Winawer et al., 2007), linguistic labels and conceptual category learning (Boutonnet & Lupyan, 2015; Lupyan et al., 2007), and the impact of space and time framing on speakers' spatial orientation and conceptualization of time (Boroditsky, 2001; Haun et al., 2006; Levinson, 2003; Levinson et al., 2002; Li et al., 2011; Loewenstein & Gentner, 2005; Majid et al., 2004; Munnich et al., 2001). However, regarding grammatical gender and its influences on the perceived properties of objects, the existing empirical evidence remains rather controversial as suggested by the inconsistent findings in the literature (see Bassetti, 2007; Beller et al., 2015; Bender et al., 2011, 2016a; Boroditsky et al., 2003; Boutonnet et al., 2012; Cubelli et al., 2011; Haertlé, 2017; Imai et al., 2014; Kousta et al., 2008; Mickan et al., 2014; Saalbach et al., 2012, 2012; Sato & Athanasopoulos, 2018; Sera et al., 2002).

As mentioned before, grammatical gender refers to the classification of nouns into different classes followed by a grammatical rule of agreement according to which the forms of

other lexical categories change, including that of adjectives, articles and pronouns. Even though the grammatical gender of nouns denoting objects does not have a sex-based semantic underpinning, one can ask whether the classification of masculine and feminine nouns, as in French, would prompt speakers to make different gender associations with the objects. To my knowledge, this question was first experimentally investigated by the cognitive scientists Guiora and Sagi (1978). Using a semantic differential test, Guiora and Sagi (1978) showed Hebrew speaking Israeli kindergarteners and adults a list of object nouns and asked them to judge to what extent the words could be related to masculine or feminine characteristics. In addition to grammatical gender, they varied the cultural gender connotations of the stimulus words: male-related (e.g. aircraft, tank), female-related (e.g. doll, skirt), and gender-neutral (e.g. clock, book). Counter the predictions of the hypothesis, they found that both child and adult participants categorized the words according to their gender connotations rather than grammatical gender. This study actually replicated results of their previous cross-cultural study comparing American and Israeli adults (cited in Guiora & Sagi 1978) where they found English and Hebrew speakers showed a similar pattern of responses – the associations were made based on gender connotations. The authors thus concluded that grammatical gender in Hebrew did not influence native speakers' perceptions of objects. Being the first researchers to look at the relationship between grammatical gender and object perception, they provided the initial counter evidence of the hypothesized gender effect.

Later, Clarke et al. (1981) replicated the study with Arabic (a grammatical gender language) and English (for comparison). This time, contrary to the findings of Guiora & Sagi (1978), a gender effect was detected. The Arabic participants assigned gender qualities to words according to their grammatical gender, while the English participants relied on the words' gender

connotations. After that, continuous attempts were made to replicate the studies with other languages like German, Spanish, French, Italian and Polish, on age groups from 5-year-olds to adults, employing various paradigms ranging from voice assignment, word and picture grouping, inference tasks, to word error induction tasks (Bassetti, 2007; Bender et al., 2011, 2016b; Boutonnet et al., 2012; Cubelli et al., 2011; Flaherty, 2001; Imai et al., 2014; Konishi, 1993; Kousta et al., 2008; Kurinski & Sera, 2011; Maciuszek et al., 2019; Mills, 1986; Ramos & Roberson, 2011; Sato & Athanasopoulos, 2018; Sera et al., 1994, 2002; Vigliocco et al., 2005). Although most of the studies reported positive results, the picture is more of a complex one. For example, the gender effect was more robust in languages with a sex-based two-gender system (e.g. French, Italian) than in those having three grammatical gender classes (e.g. German) (Sera et al., 2002b; Vigliocco et al., 2005; but see Bender et al., 2018); the influence of grammatical gender was not detected if participants' lexical access was blocked by articulatory suppression when performing the target task (Cubelli et al., 2011); the effect was found with monolinguals but not with bilinguals who spoke two grammatical gender languages (Bassetti, 2007), or that the language effect in bilinguals was dependent on the test language (Kousta et al., 2008); the language effect was limited among adult language learners (Kurinski & Sera, 2011); the effect was stronger with explicit measures (e.g. biological sex assignment) and linguistic stimuli than with implicit measures (e.g. Extrinsic Affective Simon Task) and visual stimuli (Bender et al., 2016b; Ramos & Roberson, 2011), and so on and so forth. Another complexity is that some positive effects could be attributed to task demands, or participants' employment of response strategies in completing the task that may have biased the results in favor of the hypothesis. For example, when asked to assign a male or female voice to objects as in Sera et al. (2002),

participants could simply rely on the cue of grammatical gender to complete the somewhat strange task.

That said, some research did adopt measures that were less subject to task demands (e.g. word or image priming), but again mixed results were reported. Using a semantic categorization task, Boutonnet et al. (2012) presented Spanish-English bilinguals with three pictures of objects one by one and asked them to decide whether the third picture in a series belonged to the same semantic category as the first two while measuring the event-related brain potentials (ERP). The researchers manipulated the semantic relatedness and grammatical gender of the objects. The participants' explicit judgments revealed no gender effect, but their ERPs did, suggesting that participants retrieved grammatical gender information even when it was irrelevant to the task (Boutonnet et al., 2012). Later, Sato & Athanasopoulos (2018) tested native English and French-English bilinguals with a facial image categorization task and found that English speakers were influenced by the gender connotations of object primes when asked to categorize male and female facial images, while French-English bilinguals relied on the grammatical gender of objects in performing the task. Other studies that employed different implicit methods also reported positive effects of grammatical gender (Bender et al., 2011, 2018). It is worth noting that Bender et al. (2018) reported equally strong effects for neuter and gender-marked nouns in German. This result, however, should be interpreted with a caveat: it may have been confounded with gender connotations since neuter nouns should not exhibit such an effect except if they were stereotypically associated with males or females.

Among the previous studies, the one reported by Boroditsky, Schmidt, & Phillips (2003) is particularly worth mentioning here. Employing a word association method, Boroditsky et al. (2003) asked German and Spanish participants to produce adjectives for a list of 24 object nouns

that had opposite grammatical gender in the two languages (i.e. masculine in German and feminine in Spanish, feminine in German and masculine in Spanish). To minimize the influence of participants' native language, the task was administered in English. Then to test if the adjectives generated for masculine and feminine nouns differed in their gender associations, the researchers asked a group of English speakers to rate the adjectives on the extent to which they were associated with masculine or feminine properties. According to the summary of results provided by the authors (Boroditsky et al., 2003), German and Spanish participants generated adjectives that were rated as "masculine" or "feminine" for grammatically masculine or feminine nouns in their native language. For example, the word "bridge", which is grammatically feminine in German ("Brücke"), elicited female-typed adjectives from German speakers (e.g. *beautiful*, *elegant*, *fragile*, *peaceful*, *pretty*, and *slender*), while it is grammatically masculine in Spanish ("puente") and thus induced male-typed adjectives from Spanish speakers (e.g. *big*, *dangerous*, *long*, *strong*, *sturdy*, and *towering*). The study has garnered much attention from experts and laypeople alike as evinced by the more than 800 citations on Google scholar and the over seven million views of Boroditsky's YouTube video in which these findings were described (https://www.youtube.com/watch?v=RKK7wGAYP6k). Despite the wide attention it has drawn, the important aspects of the study remain unknown, including the experimental materials, procedure and results, except being briefly described in the book chapter (Boroditsky et al., 2003).

Recently, Mickan, Schiefke and Stefanowitsch (2014) made an attempt to replicate Boroditsky et al. (2003)'s study, but this time the German and Spanish speakers were tested in their native language instead of a third language. Contrary to the findings that were originally reported, results of the replication study showed no grammatical gender effect (Mickan et al.,

2014). The study of Mickan et al., (2014), however, had its own limitations with regard to the selection of items and its sample size. The authors tested a small number of items (N = 10) with two of them denoting animals ("whale" and "mouse") and eight, objects (e.g. "pumpkin", "clock"). The mixture of animate and inanimate nouns could bias the results in opposite directions, as previous studies suggested differential effects of grammatical gender for the two types of items (see Bender et al., 2011, 2016a; Maciuszek et al., 2019). In addition, the nouns were not controlled for gender connotations and natural/artificial classification, two factors that could confound the grammatical gender effect (see Bender et al., 2016a, 2018; Mullen, 1990; Sera et al., 1994). With respect to the sample size, within each language group, there were only 15 participants for the adjective generation task and 10 for the adjective rating task. One would argue that results from such a small sample size are hardly generalizable to a greater population.

Overall, more reliable research methods are needed to better answer the question of whether grammatical gender affects people's conceptualization of objects. Thus, building on the work of Boroditsky et al. (2003) and Mickan et al. (2014), Chapter 2 of the dissertation presents two psycholinguistic experiments, using a similar word association method, to investigate French and German speakers mental representations of objects. Here, we adopt a "stack the deck" approach in which we diverge from previous studies by stacking the deck *in favor* of the original hypothesis, in such a way that any null result(s) obtained (suggested by extensive piloting) would provide strong weight against it. Specifically, unlike Boroditsky et al. (2003) who tested participants in English as a way to reduce the influence of their native language, we will test participants directly in their native language by asking one group of participants to produce adjectives to gender marked masculine and feminine nouns, and another group of native participants to rate these adjectives as representing typically male or female qualities. All things

being equal, we would expect that testing in a speaker's gendered native language (as opposed to testing them in English) while using gender marked test items would only serve to enhance any existing Neo-Whorfian effects. In addition, to address concerns over the experimenter bias (see Strickland & Suben, 2012 for a discussion), we crowdsourced the materials for Experiment 1 by asking participants to create semantically related noun pairs in French, and we standardized our item selection process to better control for the potential confounding factors such as the number of syllables and gender connotations.

As the underlying hypothesis/results which we are addressing is one which has captured the imagination of millions of people, including both scientists and the public, but there are very real doubts about the veracity or robustness of the findings, we decided to adopt an open science-based approach in assessing whether it is actually true that grammatical gender deeply influences how we think about inanimate objects. Specifically, considering that a potential publication bias – a greater likelihood of positive results that support a hypothesis being published on academic journals - may exist in science practices, any null effects that we observe may have a lesser chance of being accepted for publication. To combat such a publication bias, and at the same time, to ensure the reliability of our research design, we decided to follow the open science initiative by submitting the described work (Chapter 2) as a registered report. In doing so, we can have our experimental design peer reviewed before we commence data collection. The registered report is currently under review. In Chapter 2, I present the manuscript of the registered report as submitted, including detailed descriptions of two experiments (for which data collection will start after we receive an in-principle acceptance of the report by a journal) followed by the descriptions and results of two pilot studies.

Linguistic Gender Inequality and Language Reform

In languages with a sex-based grammatical gender system, i.e. nouns referring to male persons have masculine gender and those for female persons have feminine gender, the roles of masculine and feminine genders are often asymmetrical. The masculine gender is typically assigned the role of a generic and can be used to refer to a group of women and men, or to persons whose sex is unknown or irrelevant (Corbett, 1991; Gygax et al., 2019). On the contrary, the feminine gender has a specific female meaning and can only be used to denote female persons. The two French sentences (1a) and (1b) illustrate such a difference.

(1) a. *Les sportifs français ont gagné 10 médailles d'or.*

'The French$_{masc}$ athletes$_{masc}$ have gained 10 gold medals'

b. *Les sportives françaises ont gagné 10 médailles d'or.*

'The French$_{fem}$ athletes$_{fem}$ have gained 10 gold medals'

Written in the masculine form, sentence (1a) can be interpreted in three ways: only the male French athletes have gained 10 gold medals, the male and female French athletes together gained 10 gold medals, and the French athletes whose sex is unknown or irrelevant here earned 10 medals. However, the feminine form in (1b) exclusively refers to female athletes.

Another comparable example is English: the masculine third person pronoun *he* is designated as the generic pronoun while the feminine counterpart *she* has a female specific meaning. Similar use of masculine generic can be found in other grammatical gender languages such as German (e.g. *Lehrer* 'teachers$_{masc}$'), and natural gender languages like Swedish (i.e. *han* 'he').

Since 1970s, the world has seen heated debates on gender equality regarding the unequal treatment of masculine and feminine gender in languages, in particular, the use of masculine gender as generics. Language is both a reflection and a source of stereotyped gender beliefs. The unequal roles of masculine and feminine gender in a language has been argued to largely mirror the asymmetrical power and status relations between men and women in a speech community (Bodine, 1975; Menegatti & Rubini, 2018). Consistent with the converging evidence of the linguistic mapping of gender stereotype across languages (T. E. S. Charlesworth et al., 2021; Garg et al., 2018; Lewis & Lupyan, 2020; Tavits & Pérez, 2019), a previous cross-national research has shown that the more explicit gender information is encoded linguistically, as in grammatical gender languages, the more likely that gender stereotypes are made salient and thus the higher are the chances of reproducing sexist beliefs in speakers (Prewitt-Freilino et al., 2012). The use of masculine generics was considered to be problematic as it may represent a linguistic means of legitimizing the dominant status of the masculine gender, and in consequence turning sexist beliefs into a routine practice with most language users being unaware of it (Ng, 2007).

Over the last five decades, people who were aware of linguistic inequalities in their languages have expressed concern over the role of masculine generics in contributing to the underrepresentation of women in many fields. As a remedy to the male-oriented language structure, alternative forms that are considered more gender-fair have been proposed to replace masculine generics. For instance, the French community has introduced several gender-fair forms: double-gender (e.g. *étudiants et étudiantes* 'students$_{masc}$ and students$_{fem}$'), and contracted forms using a slash (e.g. *étudiant/es*), a dash (e.g. *étudiant-e-s*), brackets (e.g. *étudiant(e)s*), and a middot (e.g. *étudiant·e·s*) (Abbou, 2011). In a similar vein, German has seen the existence of contracted forms with an asterisk (e.g. *Reporter*in* 'reporter$_{masc}$*$_{fem}$') (Kruppa et al., 2021), word

pair forms (e.g. *Politikerinnen und Politiker* 'politicians$_{fem}$ and politicians$_{masc}$'), capital I (e.g. PolitikerInnen), and nominalized form (e.g. *die Studierenden* 'the students' derived from the verb *studieren* 'to study') (Sato et al., 2016). Likewise, in English, alternatives such as pair pronouns *he or she* (also *he/she*), singular *they*, and contracted forms *s/he* and *(s)he* have been proposed as replacements of the male generic *he* (Bodine, 1975; Gastil, 1990; Hyde, 1984; see Stahlberg et al., 2007, for a review), and in Swedish, a gender-neutral third person pronoun *hen* was invented as a substitute for *han* (Gustafsson Sendén et al., 2015).

The course of gender-fair language never runs smoothly. Initiatives for language reform were received with mixed responses by the public (Blaubergs, 1980; Parks & Roberton, 1998; see Stahlberg et al., 2007 for a review). On the one hand, proponents of gender-fair language believe that the use of masculine generics in reference to a mixed-sex group leads to biased mental representations favoring males, and in result, leaves women invisible or ignored; masculine formulations put women at a disadvantage as they are less likely to be considered for roles presented in masculine forms, and women themselves also find it hard to identify with such roles and positions (Stahlberg et al., 2007). On the other hand, opponents of language reform responded with a list of reasons for not using gender-fair language forms: there is no causal relation between language structure and gender inequality, changing sexist language is rather frivolous as a matter compared to other forms of injustice in society, people cannot be coerced to use a nonsexist language, language is not sexist by itself but instead it is the hearer/reader who interprets it in a sexist way, masculine generics are not male-biased, changing the language will estrange the current and future generations from historical heritage, and in extreme cases, people even admit and support sexism and linguistic patriarchy (Blaubergs, 1980; Parks & Roberton, 1998; Stahlberg et al., 2007). In French, additional skepticism was pointed to the possible

reading and learning difficulties[2] created by the innovative, contracted gender-fair language forms (see Gygax & Gesto, 2007).

When the proponents and critics of gender-fair language were tossing arguments around at each other, a growing body of literature from the last five decades has documented converging evidence of male-biased interpretations of masculine generics compared to the gender inclusive alternatives (see Stahlberg et al., 2007 for a review). Beginning in the 1970s, researchers investigated the difference between generic *he* and gender-fair alternatives in English. Results showed that relative to *they* and *he or she*, the male generic *he* was disproportionately associated with men (Gastil, 1990; Hamilton, 1988; Hyde, 1984; MacKay, 1980; Martyna, 1978, 1980; Moulton et al., 1978). In a similar vein, empirical studies revealed male-oriented representations when masculine generics was used in German and French, while gender-fair alternative forms improved women's visibility in mental representations (Braun et al., 2005; Gabriel et al., 2008; Gygax et al., 2008, 2012; Gygax & Gabriel, 2008; Hansen et al., 2016; Sato et al., 2016; Stahlberg et al., 2001; Stahlberg & Sczesny, 2001). These studies thus justified the concerns over sexist language forms across cultures by showing that masculine generics indeed evoke representations favoring males, leaving the roles of women largely forgotten.

Although proposals for language change encountered strong resistance at the beginning, over time, gender-inclusive language forms seemed to be accepted and applied more commonly in public communication. For instance, in Germany, German-speaking Switzerland, and Austria, using gender-fair generic form in official language has become a well-established norm (Bußmann & Hellinger, 2003; M. M. Formanowicz et al., 2015; Mucchi-Faina, 2005; Sarrasin et al., 2012); the use of generic *he* in English declined over time and the appearance of gender-

[2] For a discussion, see https://madame.lefigaro.fr/societe/lecriture-inclusive-peut-elle-vraiment-changer-la-place-des-femmes-dans-la-societe-040621-196755

inclusive forms grew significantly in public discourse (Rubin et al., 1994), and now one would say it is common to encounter gender-inclusive formulations in both oral and written English (Sarrasin et al., 2012); in Sweden, people's attitudes towards the gender-neutral pronoun *hen* became positive over time, and accordingly they were more willing to use it in everyday communication (Gustafsson Sendén et al., 2015); additionally, even though gender-fair language, especially the middot form, remains a controversial topic in France, one can find gender-fair language forms more and more often in official language (e.g. middot form is adopted by the City Hall of Paris https://www.paris.fr/municipalite and Lyon https://www.lyon.fr/solidarite).

Controversial Impact of Gender-Fair Language

Previous studies have provided consistent evidence of gender-inclusive language reducing male bias in people's mental representations (thus increasing women's presence in the minds of language users). Now, one question to ask is what social impacts does gender-fair language bring about. This may involve the perceived competence and social status of women, the inclusion of females in the workplace, and how a gender-fair language user is perceived, to name a few. Answers to these questions are by far mixed. Some research results revealed positive effects of gender-fair language as demonstrated by the findings that both women and men were more willing to apply for an opposite-sex job when the job advertisement was written in gender-inclusive language (e.g. *man or woman*) than when the language was gender-specific (Bem & Bem, 1973). Similarly, the French double-gender form (e.g. *infirmier/infirmière*, 'nurse$_{masc}$/nurse$_{fem}$') and contracted form (e.g. *infirmier (ère)*, 'nurse$_{masc(fem)}$') augmented young adolescents' perceptions of professional self-efficacy, compared to the masculine form (Chatard et al., 2005). Furthermore, speakers of gender-inclusive language were perceived as less sexist and were evaluated more positively than those who used gender-exclusive language (Greene &

Rubin, 1991), and applicants for the job of a spokesperson for UNICEF were rated as less sexist, warmer, more competent, and were more likely to be hired when they used gender-fair language than when they chose masculine generics (Vervecken & Hannover, 2012). Nevertheless, Horvath et al., (2016) showed that gender-fair forms in German and Italian increased women's visibility across male- and female-stereotyped professions, but the perceived social status and competence of people working in those professions were not affected by linguistic forms. A similar null effect of language form on the evaluations of professions was reported for the French language (Gygax & Gesto, 2007).

Inclusive language can help create an inclusive work environment. Using a mock interview method, Stout & Dasgupta (2011) found that participants' perceived sexism in the workplace, their feel of belonging, motivation to pursue a job, and identification with the job were influenced by the language form (i.e. masculine generic *he* vs. gender-fair *he or she* vs. gender-neutral *one*) used in the job description as well as by the interviewer. Their results demonstrated that job seekers, in particular women, perceived lower degree of sexism exhibited by the interviewer, felt less ostracized, more motivated to pursue the job and more identified with the job when gender-fair and gender-neutral language forms were used than when gender-exclusive *he* was adopted (Stout & Dasgupta, 2011). Consistent with this finding, Horvath & Sczesny (2016) reported that word pair forms improved the perceived fitting of women applicants for a high-status position compared to masculine form.

On the negative side, however, some evidence showed that gender-fair language could fail its good intentions by activating gender stereotypes that are disadvantageous to women. A good example is that in Italian, a female professor was seen as less persuasive and reliable when presented with a feminine professional title (e.g. *professoressa*) than with a masculine title (e.g.

professore) (Mucchi-Faina, 2005). The negative evaluations invoked by the feminine title may be attributed to the derogatory associations with the suffix *–essa* in Italian (Merkel et al., 2012) and the controversial state of gender-fair language reform in Italy the time when the study was conducted. By comparing the traditional masculine form with two feminine forms (i.e. suffix *–essa*, and neologisms *–a* and *-e*) in Italian, Merkel et al., (2012) did prove that feminized occupational titles with *-essa* led to a status loss for women relative to the masculine form and neologisms *-a* and *-e*. Similarly in Polish, feminine forms (e.g. suffix *-ka*) often derived from masculine terms and had derogatory connotations (e.g. referring to the "wife of" or "possessions of") (Koniuszaniec & Blaszkowa, 2003). Accordingly, in a CV evaluation study, women applicants of a position were rated less favorably when introduced with a feminized title (e.g. *nanotechnolożka* 'nanotechnologist$_{fem}$') than with a masculine one (e.g. *nanotechnolog* 'nanotechnologist$_{masc}$') (M. Formanowicz et al., 2013); in a similar vein, female applicants for a fictional job were perceived as less competent by both women and men, but only less warm by men when a feminine title (e.g. *aborolożka*) was used (Budziszewska et al., 2014). Even though these deleterious effects of gender-fair language documented could be attributed to the negative associations specific to some forms but not the practice of gender-fair language in general, one would anticipate some backlash effects of language reform in the short run.

Gender-fair language does not influence everyone to the same degree just as everyone in a society is not unanimously pro or against language reform. The effects of gender-fair language could be moderated by factors such as the status of language reform in a community and language users' attitudes on the issue. One would expect more positive influences of gender-inclusive language in a community where the usage of gender-fair language is well established than a society for which language reform is novel and controversial. For instance, Formanowicz

et al., (2015) observed backlash effects of gender-fair language with Polish speakers who were not used to the presence of such forms, while positive effects were found with Austrian participants to whom using gender-fair language was already a well-established norm. Specifically, the authors presented a fictitious initiative to participants and asked for their evaluations. With Polish participants, the authors found when the initiative was related to gender equality, the gender-fair language framing evoked more negative evaluations and less support, especially among male participants; when the initiative was unrelated to gender, language form did not show any effect on the evaluations. Conversely, the Austrian results showed that the gender-equality initiative was evaluated more favorably when presented in gender-fair form than in masculine form. In Swedish, another language that adopted a gender-fair form, Tavits & Pérez (2019) observed similar positive effects – women politicians were more likely to be acknowledged when the gender-inclusive pronoun *hen* was used.

With regard to the moderating role of people's attitudes and ideology, Sarrasin et al. (2012) found that individuals holding sexist beliefs were more inclined to express negative attitudes toward gender-related language reform. Furthermore, Formanowicz et al. (2013) revealed that conservatives were more likely to devalue a female applicant presented with a feminine title than liberals since the former tended to maintain traditional gender role beliefs, and were more resistant to social changes and feminist reforms. That said, the current adherence to gender-fair language in the U.K., Austria, Sweden and other countries seems to suggest that a converging attitude toward language reform and in result, positive impacts of gender-fair language on the societal level is not unattainable.

Gender-Fair Forms: Which One to Choose?

Difficulties in finding a proper gender-fair alternative to masculine generics vary from one language to another. It may be relatively more complicated for grammatical gender languages than for natural gender and genderless languages, given that changes of nominal forms in the former have consequences for the grammatical agreement in adjectives, articles and other lexical categories. Constrained by the language structure, the advancement of language reform in grammatical gender languages often gets pushed back due to the unconventional looking of newly invented forms that speakers find hard to accept. For example, in France, debates on sexist language dated back to the late 1990s (Mucchi-Faina, 2005), while gender-inclusive language, especially the middot form, remains a highly controversial subject in the year of 2021. Although various gender-fair forms (e.g. *étudiants et étudiantes*, *étudiant/es*, and *étudiant·e·s*) are frequently seen in the public space such as in subway stations, and online in official language, there is never a normative criterion as to which gender-fair form should be adopted. Oftentimes one finds multiple forms appearing in a single piece of writing[3]. The chaotic state of gender-fair language use can be attributed to the arguable drawbacks with regard to each individual form[4] that put speakers off the idea: the double-gender form (e.g. *étudiants et étudiantes*) is long and repetitive; the splitting form with a slash (e.g. *étudiant/es*) places feminine gender in a secondary position, the middot form (e.g. *étudiant·e·s*) creates reading and learning problems.

Much of the criticism concerns the unease that speakers might experience when using gender-fair language. However, despite being intuitive, these claims are not empirically validated. Gygax & Gesto, (2007) provided suggestive evidence that gender-fair forms did not make

[3] See https://www.cidj.com/metiers/infirmiere-infirmier
[4] See https://madame.lefigaro.fr/societe/lecriture-inclusive-peut-elle-vraiment-changer-la-place-des-femmes-dans-la-societe-040621-196755 for a discussion

reading more difficult than it normally is. By asking participants to read texts containing different language forms (i.e. masculine: *avocats* 'lawyers$_{masc}$'; double-gender: *avocats et avocates* 'lawyers$_{masc}$ and lawyers$_{fem}$'; and contracted form with a dash: *avocat-e-s* 'lawyer$_{masc}$-$_{fem}$-s'), the authors showed that the contracted form slowed down reading when it was shown the first time, compared to masculine and double-gender forms, but the difference in reading speed disappeared when the contracted form was encountered the second and third time in the texts. Furthermore, readers of the contracted form did not report any experience of reading difficulties. The message from this study is that French readers and hearers may need to take a short break from their task and process a bit more the contracted forms when they are encountered for the first time, but the powerful human brain will not be defeated by this tiny little puzzle.

To answer the question of which gender-fair form should be used in France, one needs to consider the costs and benefits of language reform on the societal level. Currently, empirical data that can allow for such an analysis are still much needed. For instance, we need to assess the influences of language forms on the mental representations of gender groups, the effects of gender-fair forms on language learning and processing, the social impacts of language reform and other aspects important to the decision making. To provide data bearing on this analysis, Chapter 3 presents two empirical studies on the consistency of mental representations induced by gender-fair forms, especially, the most controversial middot form. The experiments compared two candidate alternative generic forms (i.e. double-gender and middot) with the masculine form across professions of differing gender stereotypicality. Participants were asked to read a short text about the taking place of a professional gathering and provide their estimates of proportions of women and men present at the gathering. Consistent with existing evidence in English and German (e.g. Braun et al., 2005; Gastil, 1990; Hyde, 1984; Moulton et al., 1978; Stahlberg et al.,

2001), results of the two experiments showed that both gender-fair forms increased the presence of women in the minds of language users. We also compared the participants' estimates with data of a previous norming study (Misersky et al., 2013) to examine the consistency of those estimates and establish whether some language form led to biased representations of gender ratios. The results suggested that depending on whether a profession is gender-balanced, male- or female-dominated, the two forms seemed to work differently, leading to various effects regarding the reduction of male bias and the consistency of representations.

From Gender Stereotyping to Gender Inequality

Stereotyping is "the attribution of general psychological characteristics to large human groups" (Tajfel, 1969). Holding stereotypes makes people neglect individual differences and believe that all members of a group are similar (Hogg & Abrams, 1988), and once stereotypes are formed, they are spontaneously applied to members of a group (Devine, 1989).

Characteristics such as gender and ethnicity are common sources of stereotypes. Take gender for example, men are often associated with agentic qualities (e.g. *independent*, *competitive*, *arrogant* and *boastful*) while women are thought of as having communal properties (e.g. *warm*, *emotional*, *gullible* and *whiny*) (Bem, 1974; Deaux & Major, 1987; Eagly & Mladinic, 1989; Garg et al., 2018). In accordance with these stereotypes, men and women are assigned to different roles in society. Generally speaking, women are significantly underrepresented in senior, leadership positions (Bertrand & Hallock, 2001; Soarea et al., 2013). For example, a recent word embedding study revealed a century long (from 1910 to 1990) gender segregation in occupations, i.e., males dominated high-status professions like *architect*, *soldier*, *engineer*, and *judge* while females took on roles of caregiver or entertainer such as *nurse*, *housekeeper*, *midwife*, and *dancer* (Garg et al., 2018). Similar patterns of gender distributions

have been shown in norming studies and census data (Gabriel et al., 2008; Garnham et al., 2015; Kennison & Trofe, 2003; Misersky et al., 2013).

The interaction between social role assignment and stereotypical beliefs about the two sexes perpetuates the gender disparities across professional domains. According to Social Role Theory (Eagly et al., 2000; Eagly & Wood, 2016), we often associate people with the roles they play in the society where they live, and at the same time, we perceive the qualities people exhibit when they are in roles as their personal properties. Since women and men typically occupy different roles in a community, they are thought of possessing different properties specific to the gender. In return, the stereotyped gender qualities feed normative role beliefs that the two sexes should behave differently and take on different roles (Eagly & Wood, 2016). For example, the observation that women occupy the role of caregiver leads us to think that they possess caregiving properties – being warm, emotionally expressive, and mindful of other people's feelings – and that they should play the role of caregiver. However, women are considered as less suitable for a leadership role, and are less often seen as assertive, dominant and aggressive, qualities that characterize a leader and decision maker, due to the fact that females are a rarity in positions of authority.

There are two mechanisms by which gender stereotypes may affect the distributions of men and women: externally through societal expectations and internally by individuals' identification with the gender roles. On the one hand, stereotypes may blind us to differences between members of a group, and in consequence bias our perceptions and judgments about individual group members. External bias, manifested in the discriminatory practices against females in hiring and promotion processes, has been argued to contribute to the underrepresentation of women in traditional male-dominated professions such as the STEM

(Science, Technology, Engineering and Mathematics) fields (Charlesworth & Banaji, 2019; Koch et al., 2015). For instance, when applying for a lab manager position at a research intensive university, a female candidate was evaluated less positively than a male candidate even though they had exactly the same resume (Moss-Racusin et al., 2012a). Similarly, the same performance review of a veterinarian was evaluated less favorably and in result lower salaries were proposed by employers when the purported gender on the document was female rather than male (Begeny et al., 2020), and, compared to men, women were less likely to be employed to perform mathematical tasks (Reuben et al., 2014). Furthermore, even when women did succeed in male-typed fields, they were penalized for violating the stereotype that women are communal but not competitive, as shown in less favorable evaluations of female leaders than male ones (Eagly & Karau, 2002; Heilman, 2001; Heilman et al., 1995). These studies suggest that gender stereotyped beliefs are able to induce biased perceptions and evaluations of women's qualities and competence, which may contribute to the sustaining underrepresentation of women in professions of high social status and economic reward.

On the other hand, internalized gender stereotypes and social norms may bind individuals' behaviors. Once identified with a social group, individuals think and behave in conformity with the social norms that define the group (Hogg & Abrams, 1988). The normative belief that the two sexes possess different qualities and should align with different social roles can constrain the life and career choices individuals make (see Knight & Brinton, 2017). For example, the dearth of women in STEM fields has also been arguably attributed to females' lower interest in these fields and higher personal preferences for less math-intensive occupations (Breda & Napp, 2019; Ceci et al., 2009; Ceci & Williams, 2011). According to this view, women's higher interest in *persons* relative to *objects* leads them to opt out of math-intensive fields even though they are

equally competent as men (see Ceci et al., 2009 for a review). However, women's career choices may not reflect their vocational aspirations but instead be constrained by the traditional gender stereotypes such as 'math is not for girls' and 'women are family-centered' (Breda et al., 2020; Charles & Bradley, 2009; Knight & Brinton, 2017). Thus, individuals themselves are not free from the influence of gender norms when they decide what to do in their life.

Attitudes on Gender Equality Impacts Evaluations of Research on Gender Bias

The fact is indisputable that the STEM fields are male-dominated (Shen, 2013). As illustrated before, the persistent underrepresentation of women has been attributed to multiple interacting factors, ranging from stereotype-based external biases (explicit and implicit), inherent individual differences in math abilities, to free or constrained life and career choices (see Ceci et al., 2009; Charlesworth & Banaji, 2019). Among these factors, previous research has provided evidence that discriminatory hiring and promotion practices disfavoring women partially accounts for the current gender disparities (Begeny et al., 2020; Moss-Racusin et al., 2012b; Régner et al., 2019a; Reuben et al., 2014). For example, in the study of Moss-Racusin et al. (2012), the experimenters asked professors at research-intensive universities to evaluate the CV of an undergraduate student for a lab manager position. All professors received the exact same CV except in one experimental group, the applicant was given a male name and in the other group, a female name. Results revealed a significant bias in favor of the male applicant, who was rated as more competent, hirable and was offered more mentoring and a higher starting salary than the female applicant with the same qualifications.

Despite research consistently showing that women are being discriminated against in STEM fields, people's reactions to this evidence are varied (Danbold & Huo, 2017; Handley et al., 2015; Moss-Racusin et al., 2015). For instance, by asking participants to rate the quality of

the research by Moss-Racusin et al (2012), Handley et al. (2015) showed that males rated the research less favorably than female participants, and that this sex difference in evaluations was more pronounced among STEM experts than the general public. A similar sex difference was documented by Moss-Racusin et al. (2015) who conducted a content analysis of the comments posted by readers of three press articles reporting on the research of Moss-Racusin et al (2012). The authors found more positive reactions (e.g., calls for social change) among female readers, and more negative reactions (e.g. justifications of gender bias) among males. The reported sex difference in reactions was ascribed to *ingroup bias*, a mechanism by which men defend their dominant identity in STEM (Danbold & Huo, 2017; Handley et al., 2015). However, previous research also revealed that both men and women can act as defenders of the male-dominated status quo (Charles & Bradley, 2009; Glick et al., 2000; Glick & Fiske, 2001; Jost & Kay, 2005; Napier et al., 2010).

Finding the explanation for the differing responses to research on gender bias is critical to the advance of gender equality. Individuals skeptical about the existence of COVID-19 and its ability to kill an infected person are more likely to violate social distancing rules, just as people who do not believe the existence and severity of gender inequality tend to make biased judgments and decisions against women. There is some evidence that individuals who think gender inequality is not any more a problem are more inclined to discriminatory views and practices. For example, Begeny et al, (2020) asked a group of managers in the profession of veterinary medicine (men and women have equal representations in this profession in the United States) to evaluate the performance review of a vet. The experimenters randomly assigned a male or female name to the vet while keeping every other content in the review identical. The authors found that managers endorsing the belief that gender inequality is a problem of the past (given

the balanced representations of men and women) were more susceptible to under-evaluations of women's competence and worth, and as a result, assigning fewer career opportunities and lower salaries to female professionals relative to their male colleagues who possessed the exact same qualifications (Begeny et al., 2020). Consistent with this finding, scientific evaluation committees promoted fewer women to elite research positions, especially those who rejected the belief that external barriers (e.g. discrimination) rather than internal abilities constrained women's success in academia and caused their underrepresentation in STEM fields (Régner et al., 2019b). These findings suggest that people who are unaware of or having doubts about the extant gender disparities are more often than not the ones who act on stereotypes, have biased perceptions of women and make discriminatory decisions that contribute to the persistence of gender inequality.

Here, I propose that in addition to ingroup bias, a person's moral attitudes on gender equity may play a crucial role in their reception of new information related to gender bias. In particular, the moralization of gender equality - the tendency of seeing gender equality as a moral imperative that is central to one's personal identity (Skitka, 2010; Skitka et al., 2005) - can have significant implications for their factual beliefs and judgments about research touching on gender discrimination.

People tend to confirm what they believe to be true by selectively assimilating or rejecting new evidence, and more often than not, this tendency is moderated by their attitudes on the relevant issue. When faced with pro-attitudinal information, people are prompted to accept the information at face value as a way to validate their initial views, while for counter-attitudinal information, they would be skeptical about its relevance and reliability. This type of attitude-

based (mis)trust in new information is a good demonstration of "motivated confirmation bias" (Nickerson, 1998) or "motivated thinking" (Kunda, 1987, 1990).

Motivated confirmation bias has been documented in many areas, such as biased processing of health messages (Ditto & Lopez, 1992; Kunda, 1987; Liberman & Chaiken, 1992), and differential evaluations of arguments and evidence pertaining to controversial social issues like anthropogenic climate change, nuclear power and vaccine (Campbell & Kay, 2014; Edwards & Smith, 1996; Lewandowsky & Oberauer, 2021; Lord et al., 1979; Nisbet et al., 2015; Pennycook et al., 2021; Rutjens et al., 2018; Taber & Lodge, 2006; Washburn & Skitka, 2018). For example, when confronted with new evidence pertaining to the crime deterrent effects of the death penalty, both the supporters and opponents of this practice were found to evaluate the information compatible with their prior beliefs as more convincing and reliable (Edwards & Smith, 1996; Lord et al., 1979).

Following this line of research, it seems plausible that individuals respond to research showing discriminatory practices against women in academia based on their pre-existing moral attitudes, such that those who have moral convictions about gender equality find the evidence of gender bias against females more convincing than those who are less morally concerned. Having a strong moral commitment to gender equality, however, may also prompt individuals to make systematic errors in judgment, such as perceiving imprecise causal relations from correlations.

To investigate the origin of individuals' differing reactions to evidence of gender bias, and in particular, the effects of people's moral attitudes towards gender equality on their trust in science, Chapter 4 presents six experiments examining whether individuals' self-reported moral commitment to gender equality predicts their evaluations of research summary demonstrating gender bias *against* vs. *favoring* women, which may be consistent or inconsistent with their prior

beliefs. Additionally, we also examined if individuals are more inclined to make inadequate inferences when faced with a palatable conclusion.

Chapter 2. Does grammatical gender influence how we conceive of objects?

Hualin Xiao[1,2,3], Alexandre Cremers[5], Brent Strickland[2,3,4], and Sharon Peperkamp[1]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS–PSL, EHESS, CNRS), Paris, France

[2] Institut Jean Nicod (ENS–PSL, EHESS, CNRS), Paris, France

[3] School of Collective Intelligence, UM6P, Rabat, Morocco

[4]Africa Business School and School of Collective Intelligence, UM6P, Rabat, Morocco

[5] Filologijos Fakultetas, Vilniaus Universitetas, Vilnius, Lithuania

This section presents the manuscript "Does grammatical gender influence how we conceive of objects?" that has been submitted as a registered report, currently under review at *Royal Society Open Science*. The structure of this chapter replicates that of the submitted manuscript. Following the journal's requirements, the experiments not to be registered are presented as pilot studies.

Abstract

The Whorfian hypothesis, or *Linguistic Relativity*, postulates that linguistic categories influence the way we conceptualize the world. While strong, deterministic, forms of it have been refuted, weaker "Neo-Whorfian" versions are still debated. A specific question that has attracted considerable attention is whether speakers of languages that assign grammatical gender to nouns denoting inanimate objects think of these objects as "more masculine" or "more feminine" depending on the noun's grammatical gender. Such effects have been reported, but are put into question by recent replication failures. Here, we use word association tasks, asking one group of participants to associate adjectives to gender-marked nouns, presented in their native language, and another group to rate these adjectives as representing typically male or female qualities. Experiment 1 tests semantically related French nouns that have opposite grammatical genders. Experiment 2 compares French and German translation-equivalents that have opposite grammatical gender in these languages. Importantly, various aspects of our methodology stack the deck *in favor* of the Neo-Whorfian view, such that any null results provide strong weight against it. Bayesian analyses allow us to evaluate the Neo-Whorfian hypothesis, which predicts that masculine nouns more likely evoke adjectives denoting male qualities than feminine nouns, and vice versa.

Keywords:

Neo-Whorfian hypothesis, Linguistic relativity, Grammatical gender, Object conceptualization, Word association

Introduction

The relationship between language and thought, and in particular the question of whether the structure of one's language has a deep influence on how we think about and perceive the world around us, has long intrigued philosophers, linguists, anthropologists, and psychologists (Boroditsky, 2001; Levinson et al., 2002; Pinker, 2007; Whorf, 1956). Strong views of Whorfianism, i.e. the theory that our native language strictly determines the representational structures and processes underlying thought, have been largely discredited (see Wolff & Holmes, 2011; Reines & Prinz, 2009, for a review). For example, if linguistic differences radically determine thought, then we would not be able to create new words or accurate translation from one language to another should not be so commonplace (Pinker, 2007).

There is, however, a body of more recent empirical work claiming to show that weaker versions of the Whorfian thesis may hold true: While language does not strictly determine thought, it may influence it in subtle and non-obvious ways (see Wolff & Holmes, 2011; Reines & Prinz, 2009; Casanto, 2008, for a review). This topic has been examined from a number of angles, including looking at the relationship between color terms and color perception (Davies & Corbett, 1997; Gilbert et al., 2006; Thierry et al., 2009; Winawer et al., 2007), the influence of linguistic labels on conceptual category learning (Boutonnet & Lupyan, 2015; Lupyan et al., 2007), and the influence of language on spatial and temporal reasoning (Boroditsky, 2001; Haun et al., 2006; Levinson, 2003; Levinson et al., 2002; Li et al., 2011; Loewenstein & Gentner, 2005; Majid et al., 2004; Munnich et al., 2001). Critics of this view have argued that any observed effects are small enough so as to lack much ecological importance (Bloom & Keil, 2001; Pinker, 2007), can be explained through non-linguistic differences, such as differences in culture (Björk, 2008), or can be explained through relatively uninteresting task demands (Mickan et al., 2014;

Cubelli et al., 2011).

Here we focus on one of the most popular examples of linguistic structure (putatively) influencing thought: The idea that the grammatical gender for nouns referring to objects causes speakers to conceive of those objects as having more masculine or feminine characteristics (Bassetti, 2007; Beller et al., 2015; Bender et al., 2011, 2016a, 2016b; Boroditsky et al., 2003; Boutonnet et al., 2012; Cubelli et al., 2011; Haertlé, 2017; Imai et al., 2014; Kousta et al., 2008; Mickan et al., 2014; Saalbach et al., 2012; Sera et al., 1994, 2002). In order to understand what is at stake, we suggest a few theoretical distinctions here.

One can distinguish natural (i.e. biological) sex, conceptual gender, and grammatical gender. Biological sex refers to the physical characteristics of an individual's reproduction system which determine whether a person or an animal is male, female, or intersex. By definition this is not applicable to objects. Conceptual gender relates to mental representations of and associations with biological sex, and can thus go beyond the limits of biological traits. It can pick out social roles typically related to members of each sex or can refer to a form of personal identification based on an internal awareness. Thus, certain objects can also evoke strong gender associations. For example, "lipstick" is often associated with females because women and not men typically use lipstick in modern western societies. Finally, grammatical gender refers to a linguistic way of categorizing nouns into classes which triggers agreement in other words, such as articles, adjectives, or verbs (Aikhenvald, 2016; Corbett, 1991). A large proportion of the world's languages have a grammatical gender system (Aikhenvald, 2016; Corbett, 1991). These systems can be roughly divided between those that are sex-based and those based on other semantic features such as animacy or size. In sex-based grammatical gender systems, terms which clearly refer to males vs. females predictably fall, respectively, into grammatical

masculine and feminine categories. As to objects, they are assigned grammatical gender in a mostly arbitrary fashion in such languages (Corbett, 1991). For example, “lipstick” in French (*rouge à lèvres*) is grammatically masculine despite its female gender association. Similarly, French nouns such as *table* (“table”) and *bureau* (“desk”) differ in their grammatical gender despite obvious semantic similarity (with the former being feminine in French and the latter being masculine).

The central question that has sparked so much scientific and public interest is whether the grammatical gender of nouns that refer to objects fundamentally changes how speakers of the language conceive of these objects. When an object has feminine gender, do we think of it as possessing more feminine characteristics (and vice-versa for objects that have masculine gender)? While this question has been investigated from a variety of angles (Bassetti, 2007; Beller et al., 2015; Bender et al., 2011, 2016a, 2016b; Boroditsky et al., 2003; Boutonnet et al., 2012; Cubelli et al., 2011; Haertlé, 2017; Imai et al., 2014; Kousta et al., 2008; Mickan et al., 2014; Saalbach et al., 2012; Sera et al., 1994, 2002), it was brought to wider attention by a study summarized by Boroditsky, Schmidt, & Phillips (2003).

This study was conducted entirely in English on bilinguals whose native language was either German or Spanish. Participants were asked to associate adjectives with nouns, whose translations were either grammatically feminine in German and masculine in Spanish or vice-versa. These adjectives were then rated as describing masculine or feminine properties by a group of English speakers. The key finding was that people produced qualitatively different adjectives depending on the grammatical gender of the relevant nouns in their native language. For example, the word “bridge” when translated into German is grammatically feminine (*brücke*) and when translated into Spanish is masculine (*puente*). For “bridge”, German speakers were

reported to have produced adjectives that were rated as more feminine (e.g. *beautiful*, *elegant*, *fragile*, *peaceful*, *pretty*, and *slender*), while Spanish speakers produced adjectives that were rated as more masculine (e.g. *big*, *dangerous*, *long*, *strong*, *sturdy*, and *towering*).

These results have seemingly entered into the public's understanding of how language influences thought. For example, Borodisksy's TED talk describing this work (https://www.youtube.com/watch?v=RKK7wGAYP6k) has been viewed more than 5 million times on Youtube alone. Boroditsky et al., (2003) is also highly cited and continues to be cited in scientific work as evidence for an effect of grammatical gender on how we conceive of objects despite the crucial experiment summarized in it having never been published. Given the apparent widespread belief in the underlying theoretical claims, it is thus worth assessing just how robust the empirical evidence is.

The experimental paradigm described in Boroditsky et al., (2003) is particularly well-suited to address the issue of a conceptual effect of grammatical gender, as it does not explicitly ask participants to rate how masculine or feminine they perceive a given noun to be and therefore does not suffer from obvious task demands that could potentially serve as alternative explanations. This separates the paradigm from those which ask participants to explicitly rate how masculine or feminine they find specific nouns referring to objects to be (e.g. Clarke et al., 1981), to classify pictured objects as either masculine or feminine (e.g. Sera et al., 1994), or to assign a male or female voice to objects or animals (e.g. Haertlé, 2017; Sera et al., 2002). In all of these cases, one can reasonably worry about any seeming gender effect merely reflecting how participants explicitly reason through or strategize about the task at hand. In contrast, empirically robust results from the paradigm used by Boroditsky et al. (2003) would clearly strengthen the theoretical claim that grammatical gender does in fact influence object cognition. So far, one

failed attempt at a conceptual replication (Mickan et al., 2014) has shed doubt on the replicability of the results. To the best of our knowledge, the use of different paradigms has similarly failed to generalize the results (e.g. Degani, 2007; Kousta, et al., 2008), thus enhancing the need for empirical clarity.

Our approach

Here we offer two innovative conceptual replications of Boroditsky et al. (2003) using a methodology broadly based on their original approach. The novelty of the experiments here is that we "stack the deck" *in favor* of the original hypothesis. We choose to do this based on extensive piloting work (see supplementary materials) in addition to the published non-replication (Mickan et al., 2014), suggesting that any influence of grammar on conceptual object representations is either weak or non-existent. Given the question marks we had about these effects, we decided on the "stack the deck" approach such that if there were a true underlying influence of grammatical gender on object cognition, our methodology would provide the best possible chance of detecting it.

In particular, in contrast to previous studies, participants will be tested in their native language as opposed to in a second language that has no grammatical gender, such as English. The choice to originally test bilinguals in English was meant to counter any effects that might bias the results in favor of the hypothesis (Boroditsky et al., 2003). All else being equal, one would thus expect that testing directly in a language that contains grammatical gender (such as French or German) would only serve to enhance gender effects on thought. Moreover, related experimental work using other paradigms has shown that experiments conducted in the participants' native language can produce significant effects of grammatical gender (Bassetti, 2007; Kousta et al., 2008) and that English-French and English-Spanish bilinguals show weaker

sex-stereotype thinking when tested in English than when tested in gendered Romance languages (Wasserman & Weseley, 2009).

A second way in which we increase the likelihood of detecting an underlying effect (if there is one to be found) is by including gender marked determiners on the relevant noun items. One would presume that, if anything, making grammatical gender information more salient in this way would again only serve to enhance any underlying effects if they are truly present.

The advantage of our approach is that if (as suggested by our piloting results) we find null effects of grammatical gender on the masculinity/femininity of adjectives people associate with masculine vs. feminine nouns, this will allow for strong inference against the original hypothesis. This inferential strength from a potential null result comes at the cost of inferential strength in the case of a positive result. In other words, if positive effects of grammatical gender on adjective choice are found, this could be due to the underlying hypothesis being true, or it could be due to possible confounds which were intentionally introduced to increase the odds of finding any effect. For example, perhaps mere statistical associations between grammatically gendered determiners and certain adjectives, which are independent of conceptualization, could drive a positive effect. We therefore also plan for "conditional" experiments: if positive effects are found when gendered determiners are visible (nouns presented in singular forms), we will run follow-up experiments in which the nouns will be presented in plural forms, thus the relevant determiners are no longer used (German group) or do not visibly carry gender information (French group). To ensure that any null effect we find is not due to a lack of statistical power, we will use Bayesian analyses. These methods will allow us to quantify our confidence in null results.

Experiment 1

In this experiment we will test the gender effect hypothesis for pairs of French nouns that are related in meaning but have opposite grammatical gender. The choice of studying French instead of Spanish or Italian is theoretically neutral in that putative effects of grammatical gender on object conceptualization is not intended to be language specific. Thus, this choice should not in principle affect the underlying hypothesis derived from the broader theory.

Choosing semantically related nouns as stimuli will allow us to reduce a possible bias due to word meaning. We will test nouns referring to objects, and add a control condition with nouns referring to persons. That is, as male and female persons are typically associated with different gender traits (Costa et al., 2001), we expect to observe a gender effect for person nouns.

Methods

We will manipulate two factors, noun type (i.e. referring to objects vs. persons) and grammatical gender (masculine vs. feminine).

Materials

Person nouns. As control items we selected 12 pairs of person nouns, including kinship terms (e.g. father, mother) and role names (e.g. king, queen) (see Appendix B1). Within each pair, the two nouns are semantically related, as one refers to a male person and the other to the female counterpart (e.g. *père* 'father' – *mère* 'mother'). Importantly, the grammatical gender of person nouns is consistent with the biological sex of their referents. For instance, *père* 'father' has masculine and *mère* 'mother' has feminine gender.

Object nouns. To avoid experimenter bias in selecting stimuli (see Strickland & Suben, 2012 for a discussion) we designed two tests to standardize our choices of object nouns. The first test asked a group of participants to generate semantically related noun pairs, and the second one

invited another group of participants to rate these pairs on a scale for semantic relatedness. Note that as the items are meant to be presented with a gendered definite article, all nouns should be consonant-initial; indeed, before vowel-initial nouns, the masculine and the feminine article, i.e. *le* and *la*, lose their vowel and hence become indistinct).

Noun pair generation: We recruited native French speakers (N = 155, 67 men, 86 women, and 2 other sex), aged between 18 and 61 years (M = 33, SD = 11.1), as participants on the platform Clickworker (https://www.clickworker.com/) for what was described as a "linguistic task". According to their report, none of them had learned a foreign language that had a grammatical gender system. They were paid 0.60 € for their participation.

The test was run online on Qualtrics (https://www.qualtrics.com). Participants were first shown instructions asking them to provide 12 pairs of semantically related nouns (e.g. *sable* "sand" – *plage* "beach") that obey several constraints: Each pair should consist of one masculine and one feminine noun; all nouns should start with a consonant, and this consonant should not be 'h' (as before many h-initial nouns the masculine and feminine articles become indistinct, just like before vowel-initial words); all words should refer to objects or concrete places and no word should refer to animals or persons.

When they finished reading the instructions, participants were then shown seven pairs that violated a given constraint, along with explanations of why those pairs were not acceptable. Next, they were shown seven different pairs with only one of them obeying all constraints, and they had to find the good one. Those who had found the right pair were told so, while those who had chosen an unacceptable pair were explained why their choice was wrong and they were shown the right pair. After receiving the feedback, all participants were able to move to the main task, which was to type in 12 word pairs. The above constraints remained visible during the test.

The test ended with questions about participants' sex, age and native language.

After removal of pairs containing compound words (e.g. *brosse à dents* 'toothbrush'), non-French words (e.g. *ring*), person nouns (e.g. *princesse* 'princess'), adjectives (e.g. *pauvre* 'poor') and abbreviations (e.g. *CD*), the list of responses comprised 883 noun pairs. We selected the 249 pairs that had been proposed by the largest number of participants (range: 2 – 58 times) to be used in the semantic relatedness rating test for final stimuli selection.

Semantic relatedness rating test: In order to assess the semantic relatedness of the pairs generated in the previously described phase, a norming test was run online using Labvanced (https://www.labvanced.com/).

In addition to the 249 target pairs, we included 83 filler pairs (i.e., 1/3 of the number of target pairs), created by pairing one masculine and one feminine noun taken from different pairs. We then randomly divided the 332 word pairs (249 target and 83 filler pairs) into five lists, with each list containing between 65 and 67 pairs (i.e. 49 or 50 targets and 16 or 17 fillers). Participants were randomly assigned to one of the five lists. The order in which the pairs of a relevant list were displayed was fixed across participants, with filler pairs being evenly distributed among target pairs.

We recruited native French speakers (N = 94, 59 men and 35 women), aged between 21 and 76 years (M = 43, SD = 11.9) on the crowdsourcing platform Foulefactory (https://www.foulefactory.com). We restricted recruitment to participants who had not learned a foreign language that also had a grammatical gender system as participants. They were paid 0.50€ for their participation.

Participants were presented with the word pairs one by one and asked to indicate on a 6-point scale to what extent they thought the words were semantically related (1 – Not at all related,

6 – Perfectly related). We specifically made clear that they should not reflect too much on the task, but instead respond based on intuitions. After finishing all trials, participants answered questions about their sex, age, native language, and second language learned before age 10.

As a manipulation check, we ran a t-test on the average ratings between target and filler pairs. Results showed that target noun pairs (M = 4.56, SE = .008) were rated as more semantically related than filler pairs (M = 1.27, SE = .004; $t(330) = 39.62$, $p < .0001$). Eighty test pairs had received a mean score between 5 and 5.9 (M = 5.41, SD = .24). From these, we removed certain pairs such that the final set contained no duplicated nouns, i.e. nouns appearing in more than one pair. For example, consider the pairs *toile* – *tableau* (5.76), *peinture* – *tableau* (5.62), and *peinture* – *pinceau* (5.47). Both *tableau* and *peinture* appeared twice. As *toile* – *tableau* had the highest rating among the three, we excluded *peinture* – *tableau* because of the duplicate noun *tableau* (as well as all other pairs containing either *toile* or *tableau*), after which *peinture* – *pinceau* no longer contained a duplicate noun and was hence kept. Then, from the unique pairs, we removed those containing words with ambiguous meanings (e.g., the word *mine* in French can refer to ore mining, a pencil core, a facial expression, or an explosive device). The final set consisted of 52 pairs that were matched in number of syllables ($M_{masc} = 1.81$, $M_{fem} = 1.60$, $t(50) = 1.75$, $p = 0.09$). See Appendix B2 for the complete list.

Procedure

The experiment will be conducted online using Labvanced software. It will consist of two parts, adjective generation and adjective rating, performed by different groups of participants.

Adjective generation. The first part of the experiment assesses which characteristics participants tend to associate with inanimate entities. Participants will be told that the experiment is about the associations between nouns (pre-tested and selected above) and adjectives, with no

mention of our interest in comparing masculine vs. feminine nouns. They will be shown four short sentences and asked to complete each sentence with three different adjectives. Nouns will be presented as subjects of sentences in singular forms, preceded by gender-marking articles *le* and *la* respectively for masculine and feminine words. The adverb *très* 'very' is used in the sentences to elicit gradable adjectives, which are more likely to represent people's subjective evaluations of non-arbitrary objects' characteristics (Wheeler, 1972). The presence of the adverb will also eliminate set phrases such as *courrier recommandé* ("registered mail").

Examples of an object and a person noun embedded in the carrier sentence are shown in (2).

(2) a. Object noun: *main* 'hand'

La main est très _____

'The hand is very_____'

b. Person noun: *mère* 'mother'

La mère est très _____

'The mother is very_____'

Participants will first be randomly assigned to either the object or person noun condition, and within each condition, randomly to a subset of four pairs. The goal is to assign random pairs to each participant while ensuring that each pair is seen by the same number of participants. To avoid making the comparison of grammatical gender too salient, each participant will see only either the masculine or the feminine word of a semantic pair. From the four pairs assigned to each participant, we will randomly select two masculine and two feminine nouns. We will also

ensure that the two nouns within each pair (masculine and feminine) are shown to the same number of participants. The test will end with demographic questions about participants' sex, age, native language, foreign language, and country of residence.

We will exclude all non-adjectival and non-French responses, and select adjectives that are used more than once for the following rating task. There is no minimum number of adjectives to be rated, but to limit the test to a manageable scale, we will test a maximum of 200 adjectives (in the adjective rating task described just below). If more than 200 adjectives are generated, we will order them by the number of times they were used and remove individual adjectives, starting from the least frequent one, until there are 200 left.

Adjective rating. The second part of the experiment tests if adjectives associated with feminine nouns are more likely to be related to female gender traits, and vice versa for masculine nouns. Participants will be randomly presented with adjectives and asked to indicate on a 7-point Likert scale to what extent they think the words represent a masculine or a feminine quality (1 'Very masculine' – 7 'Very feminine'). Adjectives that have gender inflection, i.e. varying forms for masculine and feminine gender agreement, will be presented in both forms (e.g. *petit/petite* '$small_{masc}$/$small_{fem}$'), while those having a single form for masculine and feminine gender will be shown in their unique form (e.g. *pratique* 'practical/convenient'). Participants will be told not to deliberate over the task but instead to respond based on intuitions. Each participant will be assigned to between 50 and 70 adjectives (the exact number will depend on how many items will be produced in the adjective generation test), presented one at a time in random order. As an attention check, they will be asked from time to time to recall the previous word they saw. There will be one such attention check every 10 trials over the test.

As before, the test will end with demographic questions about participants' sex, age,

native language, foreign language, and country of residence.

Sample sizes

For the adjective generation test, we will randomly generate 10 partitions of the object and person nouns into thirteen and three subsets of four pairs, respectively. We aim for a sample size of 160, i.e. 130 for the object noun condition and 30 for the person noun condition, thus with each noun being shown to 10 participants. As data will be collected in batches with the exclusion criteria applied after a batch is completed, we may end up with a slightly bigger sample size than 160.

For the adjective rating test, the exact sample size will depend on the number of adjectives to test. Each participant will rate between 50 and 70 adjectives. For each adjective, we aim for at least 25 ratings. Again, as data will be collected in batches with the exclusion criteria applied after a batch is completed, we may end up with a slightly bigger sample size than 25 per adjective.

Participants

Participants will be native French speakers, aged at least 18 years, who have not started learning a second language with a grammatical gender system before age 10, whose self-rated proficiency in any such language does not exceed 5 on a scale from 1 to 7, and who live in France. Participants in the adjective generation task will be paid 0.50€ and those in the adjective rating task 0.60€ for their time.

We will exclude participants who do not satisfy all our recruitment criteria. For both tests, participants who fail to complete the assigned task will be excluded from the analyses. If a participant took the study more than once, only the first set of responses will be kept. Finally, we will exclude data from participants who fail to answer all attention check questions correctly at

the rating task.

Statistical analysis

We will first extract the mean rating on the masculine-feminine scale for each adjective from the rating task to define its *female quality association* index (FQA). The data from the generating task will be analyzed with a Bayesian linear mixed-effects model fitted with Stan (Carpenter et al., 2017), using uninformative priors. The dependent variable will be the FQA of the generated adjectives and the predictors will be the gender of the noun for which the adjective was generated (sum-coded), the type of entity denoted by the noun (object or person, treatment-coded with object as the baseline), and their interaction. The model will include the maximal random effects structure for both participants and noun pairs (Barr et al., 2013). This includes random intercepts and random slopes for gender together with their correlation (gender is both within-participant and within-item), but no random slopes for object type, which is both between-participants and between-items. The Stan code for the model is provided in Appendix A1.

If grammatical gender does affect conceptualization, we expect the adjectives generated for an object noun with feminine grammatical gender to have a more feminine rating on average. This would translate as a positive main effect of gender in our model. For person nouns, we strongly expect a clear effect of gender, so the interaction between gender and object type is expected to be positive. This interaction is therefore of little theoretical interest, but can be used to estimate the sensitivity of our design.

We will report the 95% HDI Credible Interval for the main effect of noun gender and its interaction with object type, as well as the posterior probability $P(\beta>0)$ for each parameter. We will then compare the full model to a model without the main effect of gender, and report the

Bayes factor. If the Bayes factor is superior to 3 and/or $P(\beta>0) > .95$, we will run the follow-up experiment with plurals to control for the role of gendered articles.

Experiment 2

One limitation of Experiment 1 is that we are unable to compare exactly the same word in the masculine vs. feminine form. While focusing on semantically related pairs within a language is a reasonable proxy, looking across languages at words with matched meanings but which differ only in their grammatical gender is arguably a "purer" test of the original theory (despite the fact that translation equivalents are rarely 100% equivalent). Thus, the aim of Experiment 2 is to extend our test of the gender effect hypothesis from within a single language to across languages. This also brings us closer to the experiment summarized in Boroditsky et al. (2003).

Here we will compare across French and German to investigate if grammatical gender influences native speakers' conceptualization of objects. In particular, we ask whether translation equivalents in French and German that have opposite grammatical gender (e.g., *pont – Brücke* 'bridge') will be associated differentially with more male vs. female qualities by native speakers of these two languages, such that speakers of both languages associate grammatically masculine and feminine nouns with typically male and female qualities, respectively. In contrast to Boroditsky et al. (2003), instead of testing bilinguals in a non-gendered language like English, we instead test French and German speakers directly in their native language. We reason that if anything, doing this should increase the odds of finding an effect of grammatical gender on the femininity/masculinity of the adjectives produced, due to broad statistical associations that will likely have been built up between gender marked determiners and gender associated conceptual properties.

To address potential concerns about the influences of words' specific gender connotations, i.e. some words might be conceptually associated with males or females, we will run a control analysis with an added predictor for the gender association of English translations of the French/German pairs, measured independently.

Additionally, Experiment 2 will answer a secondary question of whether natural entities are more associated with female qualities and artifacts, with male qualities as suggested by previous studies (e.g. Mullen, 1990; Sera et al., 1994) by comparing words for natural items vs. artifacts.

Methods

We will manipulate two factors: language (French vs. German) and noun gender (masculine vs. feminine). Employing a similar method as for Experiment 1, we will test translation equivalent nouns that are assigned opposite grammatical gender in French and German (i.e. masculine in French and feminine in German, and vice versa). Unlike in Experiment 1, we only test nouns for inanimate entities.

Materials

We selected 59 pairs of translation equivalent nouns with opposite gender in French and German; 30 are masculine in French and feminine in German, and 29 are feminine in French and masculine in German (see Appendix B3). Among them, 23 nouns refer to natural entities (e.g., 'mountain'), 27 to artifacts (e.g., 'spoon'), and 9 cannot be easily categorized (e.g., 'milk'). Gender is balanced within each of these categories.

All nouns denote concrete objects, and none of them denotes an animal, body part, or clothing item. In addition, none has more than one, frequent, meaning in one of the languages (e.g. we excluded French *bureau*, which means both 'desk' and 'office'), and none has a

language-specific cultural connotation (e.g., we excluded the pair *masque* – *Maske* 'mask', as the COVID-19 crisis might have induced different attitudes towards face masks in France and Germany). All French nouns are consonant-initial (such that a preceding definite article is gender-marked).[5]

The French words contain between 1 and 3 syllables, the German ones between 1 and 4 (M_{French} = 1.66; M_{German} = 1.85; t(58) = -1.90; p = .06). Among the French words, there is no difference in number of syllables between masculine and feminine words (M_{masc} = 1.80; M_{fem} = 1.52; t(57) = 1.75, p = .09), while among the German words, the feminine ones are longer than the masculine ones (M_{masc} = 1.59; M_{fem} = 2.10; t(57) = -2.83, p < .006).

For the subset analysis that addresses the potential confound of gender association, we ran a pretest in a neutral language, English, that does not have a grammatical gender system. We asked a group of native English speakers (N = 107, 27 men and 80 women), aged between 18 and 65 years (M = 33, SD = 11.2), living in the UK, to indicate on a 7-point Likert scale to what extent they associated the English translation equivalents of our items with men or with women (1 "Men" – 7 "Women"). Words were presented in random order across participants.

Procedure

The procedure will remain identical to that of Experiment 1. Participants will be tested in their native language.

The nouns will be randomly divided into 15 subsets, with 14 of them containing 4 words

[5] In order to select the pairs, we first retrieved a list of French nouns from the electronic dictionary Lexique (http://www.lexique.org/), which we ordered from highest to lowest lemma frequency according to Lexique's database of movie subtitles. Starting from the most frequent one, and taking into account the constraints mentioned above, we used an online French-German dictionary (https://fr.langenscheidt.com/francais-allemand) to look up the German translation equivalents. Whenever a French word had more than one translation, we only considered the first, most frequent, one. All pairs in the final selection were checked by a native speaker of German who has learned French and has lived in France for nearly two decades.

(2 masculine and 2 feminine in French) and one subset containing 3 words (2 masculine and 1 feminine in French). Participants will be randomly assigned to one of the 15 subsets and within each subset, words will be presented in random order.

The relevant sentences for the word *bridge* in French and German are shown in (3).

(3) a. French

Le pont est très _____

b. German

*Die Brücke ist sehr*_____

'The bridge is very_____'

Sample sizes

For the adjective generation test, we aim for a sample size of 300 (150 for each language group), with each noun being presented to 10 participants. As data will be collected in batches with the exclusion criteria applied after a batch is completed, we may end up with a slightly bigger sample size than 300.

For the adjective rating test, the exact sample size will depend on the number of adjectives to test. Like in Experiment 1, each participant will rate between 50 and 70 adjectives. For each adjective, we aim for at least 25 ratings. Again, as data will be collected in batches with the exclusion criteria applied after a batch is completed, we may end up with a slightly bigger sample size than 25 per adjective.

Participants

Participants will be native French and German speakers who live in France and Germany,

respectively, aged at least 18 years, who have not started learning a second language with a grammatical gender system before age 10, whose self-rated proficiency in any such language does not exceed 5 on a scale from 1 to 7. Participants of the adjective generation task will be paid 0.50 € and those of the adjective rating task 0.60 € for their time. The exclusion criterion is identical to that for Experiment 1.

Statistical analysis

The data will be analyzed following the same procedure as Experiment 1. The predictors for this model will be noun gender (again, sum-coded), language (sum-coded), and their interaction. We will include the maximal random-effects structure, which this time includes random intercepts and gender slopes for participants, and random intercepts, gender and language slopes for noun pairs (but not their interaction), as well as all correlations. The Stan code for the model is provided in Appendix A2.

The effect of interest is the main effect of noun gender, for which we will report 95% HDI credible interval, posterior probability $P(\beta>0)$, and the Bayes factor in favor of a model without this main effect.

In addition to this main analysis, we will run the following three control analyses: First, as word length might influence perceived conceptual gender, we will control for the number of syllables by removing all pairs in which the German word has three or four syllables (N = 8). The remaining 51 word pairs (i.e. 23 masculine in French and feminine in German, 28 the reverse) show no difference in number of syllables between the French and German equivalents ($t(49) = 0.60$, $p = 0.55$), nor between masculine and feminine in either language (French: $t(49) = 0.89$, $p = 0.38$; German: $t(49) = 1.50$, $p = 0.14$).

Second, a strong association of certain items with either men or women (e.g., cigar, apron) could obfuscate a potential influence of grammatical gender. To control for this, we will run a model including gender association of the English translation of the word as a predictor, as well as its interaction with grammatical gender.

Finally, we will compare words for natural items vs. artifacts to test the hypothesis that natural entities are more associated with women and artifacts, with men (Mullen, 1990; Sera et al., 1994). Leaving aside the nine items which are neither typically natural nor artificial, we would run a model similar to the main analysis, but with an extra predictor for object type (artifact vs. natural, sum-coded).

If the Bayes factor is superior to 3 and/or $P(\beta > 0) > .95$ for the main analysis, and control analyses show that this effect is neither an artifact of word length nor gender association of the noun pairs, we will run the follow-up experiment with plurals to control for the role of gendered articles.

Pilot Studies

Pilot 1

This pilot is similar to Experiment 1 in design. The object and person noun conditions in this pilot were run separately at different times, with different participant groups, and slightly different experimental designs. To be consistent with descriptions of Experiment 1, and to help readers make sense of the results, we present the two conditions together as a single pilot.

Person noun condition

We used the same stimuli as described in Experiment 1, and similarly, it included two tasks: adjective generation and adjective rating.

Materials

The items were the same as those in Experiment 1, consisting of 12 pairs of person nouns in which the two nouns are semantically related, with one referring to a male person and the other to the female counterpart (e.g. *père* 'father', *mère* 'mother') (see Appendix B1).

Procedure

Adjective generation. The procedure of this test was similar to that of Experiment 1, with two exceptions. First, we used a single-trial design: participants were randomly presented with only one of the 24 person nouns. Thus, noun gender was a between-subject factor. Second, the words were presented in plural forms preceded by the gender-neutral determiner *les*, as illustrated in (4).

(4) Person noun: *père* 'father'

Les pères sont très___

'The fathers are very___'

We selected 23 adjectives that had been generated by at least two participants (range: 2 - 13) (a total of 168 adjectives had been generated). These 23 adjectives represented 20% all responses.

Adjective rating. The procedure was similar to that described in Experiment 1. Participants were presented with the 23 adjectives one by one in random order, but unlike in Experiment 1 where a Likert scale will be used, here they were asked to indicate on a response slider to what extent they thought the adjectives described a masculine or feminine characteristic. There were two versions of slider, depending on whether the left and right endpoints were labeled *Masculine* ('Masculine') and *Féminine* ('Feminine'), respectively, or the reverse. The initial position of the indicator was placed at the midpoint of the slider, labeled as *Neutre* ('Neutral'). The slider version was counterbalanced across participants. Like in Experiment 1, adjectives with gender inflection, i.e. adjectives with different forms for masculine and feminine gender agreement, were presented in both forms, while those having a single form for masculine and feminine gender were shown in their unique form. Instructions on how to use the slider were shown to participants at the beginning of the test. The test ended with demographic questions about participants' sex, age, native language, and foreign language experience.

Participants

Using Foulefactory, we recruited 144 native French speakers (73 women and 71 men), aged between 19 and 68 y (M = 41, SD = 12.3), for the adjective generation task, and 80 different native French speakers (44 women and 36 men), aged between 19 and 69 y (M = 41,

SD = 12), for the adjective rating task.

Object noun condition

The stimuli used here were different from that for Experiment 1. However, like in Experiment 1, this condition consisted of two tasks: adjective generation and adjective rating.

Materials

Twelve pairs of semantically related French nouns referring to objects were used as items (see Appendix B4). Within each pair, the nouns are semantically related, but have opposite grammatical gender (i.e. masculine vs. feminine). For example, *table* 'table' and *bureau* 'desk' are close in denotation, with the former being grammatically feminine and the latter masculine. Items were divided into two lists of 12, each list comprising 6 masculine and 6 feminine nouns, and the two nouns of each pair were assigned to separate lists.

Procedure

Adjective generation. The procedure was different from that of Experiment 1, as there was a pretest before the adjective generation task, and for this task, the noun items were not presented in a sentence. However, like in Experiment 1, the nouns were shown with a preceding gender-marking determiner.

Participants first went through a pretest on their knowledge about 'adjectives'. They read a sentence and answered how many adjectives were present in the sentence. They then received either positive or negative feedback depending on their response. As part of the negative feedback, the correct answer was given and the sentence was shown again to them with all adjectives in the sentence being highlighted. Regardless of their response, they were able to move to the main test.

After the pretest, participants were shown 12 nouns, one at a time, and asked to generate

three adjectives to describe the object denoted by it. They were randomly assigned to either list of 12 nouns; hence they saw only one noun of each pair. Thus, in this condition, noun gender was a within-subject factor. Within each list, the nouns were shown in random order one by one. All nouns were presented with a preceding gender-marking article *le* (for masculine nouns) or *la* (for feminine nouns). At the beginning of each trial, participants were told that answers other than adjectives would not be accepted. The test ended with two demographic questions about participants' age and sex.

We selected 19 adjectives as stimuli for the next rating test. They were the most common adjectives (range: 5 -13 times) provided for each noun (280 adjectives had been generated in total). Again, these 19 adjectives represented 38% of all responses.

Adjective rating. The procedure for this task was similar to the description of Experiment 1, except that a response slider, instead of Likert scale, was used.

Another group of participants were shown the 19 adjectives one at a time in random order, and asked to rate on a response slider to what extent the adjectives describe characteristics related to men or to women. As in the person noun condition, there were two versions of the response slider and the slider version was counterbalanced across participants. Again, adjectives that have gender inflection were presented in both forms while those having a single form for masculine and feminine gender were shown in their unique form. The test ended with two demographic questions about participants' sex and age.

Participants

Using the crowdsourcing platform Foulefactory, we recruited 37 native French speakers (10 men and 27 women), aged between 28 and 67 y (M = 47, SD = 12), for the adjective generation task, and 35 different native French speakers (12 men and 23 women), aged between

23 and 64 y (M = 44, SD = 11.3) for the adjective rating task.

Results

A plot of FQA as a function of grammatical gender is shown in Figure 1.

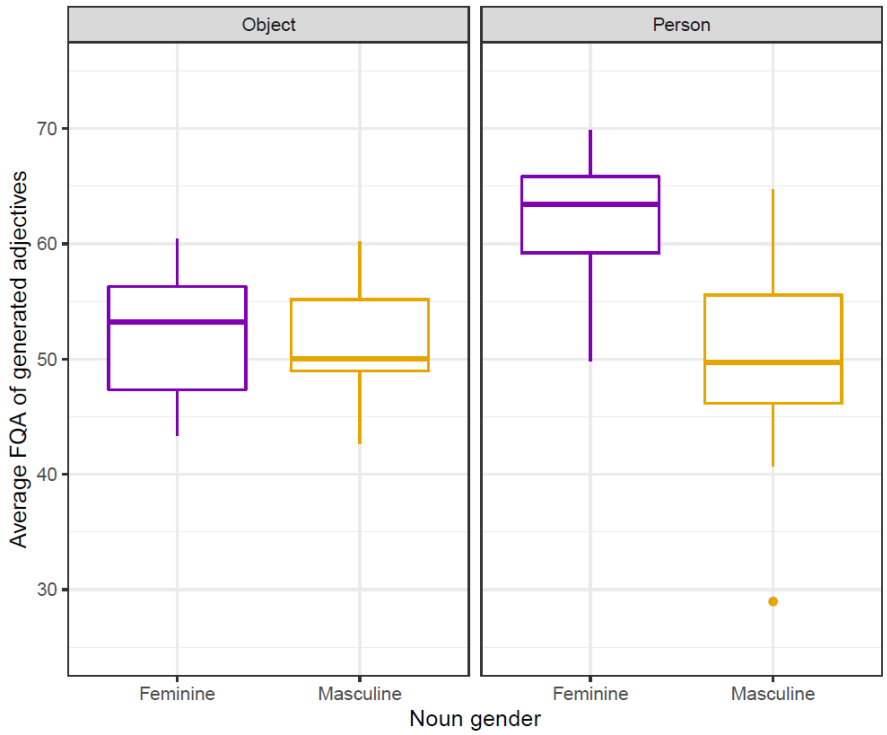


Figure 1. Female quality association (FQA) of the adjectives as a function of grammatical gender of the noun for which they were generated. Left panel: object nouns; Right panel: person nouns

We performed the same analyses as planned for Experiment 1, except that the model did not contain a by-participants random slopes for gender in the person condition (as participants only saw one noun). The mean posterior value for the effect of gender on object nouns was 0.07 (CI: [-0.54, 0.68], P(β>0) = 0.60). As a comparison, the mean posterior value for the interaction between noun gender and noun type was 1.11 (CI: [0.20, 2.0], P(β>0) = 0.990). Comparing

models with and without an effect of gender on object nouns, the Bayes factor in favor of the null hypothesis was 10.0. This value clearly is in favor of the null hypothesis, yet the discrepancies between the object and noun conditions and the small proportion of adjectives for which a rating was actually elicited cast doubt on the reliability of this result and call for a proper experiment with a cleaner design and higher sample size. Our Experiment 1 will address both problems (by testing more participants and introducing the intensifier très, which limits the class of possible adjectives, we should see less unique adjectives)

Pilot 2

This pilot is similar to Experiment 2 in terms of its cross-linguistic approach (French-German) and experimental procedure.

Methods

The stimuli used in this pilot were different from that for Experiment 2. However, similar to Experiment 2, it consisted in two tasks: adjective generation and adjective rating.

Materials

Twenty-four nouns spanning various categories, including *settings* (e.g. bridge, mountain), *objects* (e.g. key, arrow), *animals* whose biological sex is not obvious (e.g. mosquito, snake), and *abstract concepts* (e.g. thought, need) were selected as stimuli (see Appendix B5). Half of these nouns were grammatically masculine in French and feminine in German, and the other half were grammatically feminine in French and masculine in German. One word, "rain", was included in the French test but not in the German one since it does not have a plural form in German.

Procedure

Adjective generation. The procedure was similar to that for Experiment 2, except in the following two aspects. First, here each participant only saw one noun, and the nouns were presented in plural forms. Thus, French nouns were preceded by the gender-neutral determiner *les*, and German nouns were not preceded by any determiner. Next, the adverb "very" was not present in the carrier sentences for the two language groups. See the example for *bridge* below in French (5a) and German (5b).

(5) **a.** French

Les ponts sont _____

'The bridges are_____'

b. German

*Brücken sind*_____

'Bridges are_____'

For the following rating test, we selected 100 French adjectives that had been generated by at least two participants (range: 2 – 88) (except the word *diluvien* "diluvian" which had been generated only once), and 19 German adjectives adjectives that had been used by at least two participants (range: 2 – 7). (Three hundred French adjectives and 95 German adjectives had been generated in total). The 100 French and 19 German adjectives represented 73% and 51% of all responses of the two groups, respectively.

Adjective rating. The procedure was similar to that of Experiment 2, except that here a response slider was used instead of a Likert scale. For the French group, the set of 100 French adjectives was randomly divided into four lists of 25. French participants were randomly assigned to one list, the items of which were presented in random order, one at a time. For the German group, participants were presented with all 19 adjectives, one at a time, in random order. As described in Experiment 2, French adjectives with gender inflection, i.e. having different forms for masculine and feminine gender agreement, were presented in both forms, while those having a single form for masculine and feminine gender were shown in their unique form. Unlike in French, German adjectives do not have gender inflected suffixes, and thus all items in German were displayed in their unique form. Participants were asked to indicate on a response slider to

what extent they thought the adjectives described a masculine or feminine characteristic. There were two versions of slider, depending on whether the left and right endpoints were labeled 'Masculine' (French: *Masculine*; German: *maskuline*) and 'Feminine' (French: *Féminine*; German: *feminine*), respectively, or the reverse. The initial position of the indicator was placed at the midpoint of the slider, labeled as 'Neutral' (French: *Neutre*; German: *neutrale*). The slider version was counterbalanced across participants for each language group. The test ended with demographic questions about participants' sex, age, native language, and foreign language experience.

Participants

Using Foulefactory, we recruited 347 native French speakers (147 men and 200 women), aged between 21 and 69 y (M = 40, SD = 11), for the French adjective generation task, and 128 native French speakers (58 men and 70 women), aged between 19 and 68 y (M= 42, SD = 11.5), for the corresponding adjective rating task. None of them had participated in Pilot 1.

Using Prolific, we recruited 63 native German speakers (32 men and 31 women), aged between 18 and 60 y (M = 30, SD = 10.3), for the German adjective generation task, and 49 native German speakers (25 men and 24 women), aged between 20 and 58 y (M = 31, SD = 10.1), for the corresponding adjective rating task.

Results

A plot of FQA as a function of grammatical gender is shown in Figure 2.

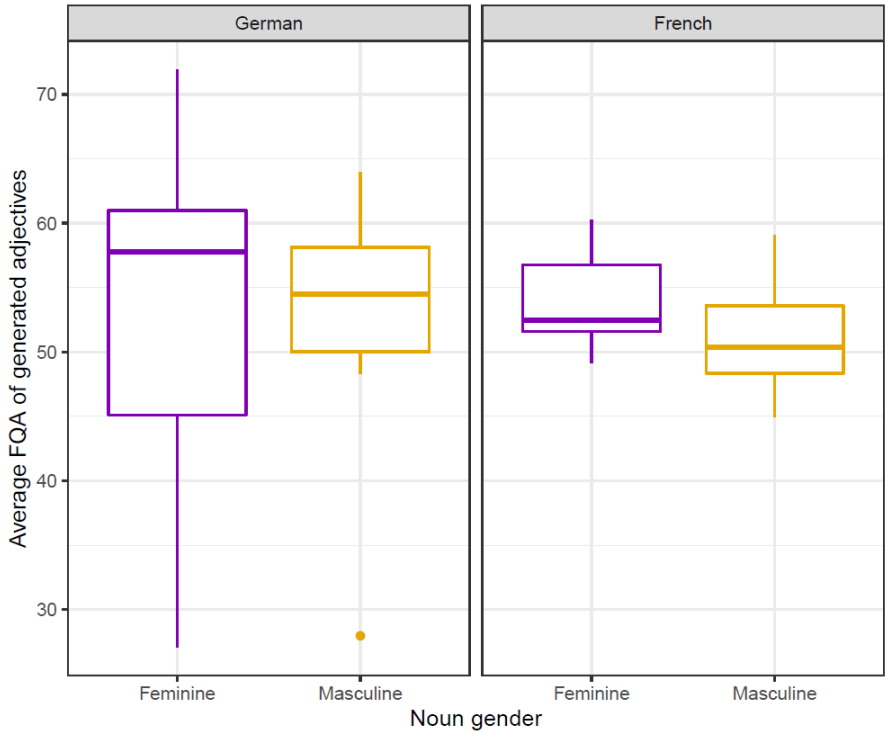


Figure 2. Female quality association (FQA) of the adjectives as a function of grammatical gender of the nouns they were generated for, by German and French speakers.

We ran the same analysis as planned for Experiment 2, except that there were no by-participant random effects (since each participant only saw one noun). We found a mean main effect of gender of 0.30 (CI: [-0.32, 0.93], $P(\beta>0) = 0.84$). The Bayes factor in favor of the null hypothesis was 5.6. Here again, the Bayes factor is in favor of the null hypothesis, albeit less strongly so compared to Pilot 1. Nevertheless, the imperfect design of this pilot and the imbalance between the number of French and German participants prevent us from accepting the null hypothesis on the basis of this value alone. Only if the proposed Experiment 2 also returns a high Bayes factor (despite being more biased towards a positive result), would we feel confident about this null result.

Appendices

Appendix A. Stan models

A1. Model for Experiment 1

```
data {
 int<lower=0> N;                // number of data points
 int<lower=1> N_subj;           // number of subjects
 int<lower=1> N_pair;           // number of pairs
 vector[N] fqa;                 // dependent variable (scaled)
 vector[N] gender;  // gender (sum-coded: masc=-0.5, fem=+0.5)
 vector[N] type;                // noun type (treatment-coded: object=0, person=1)
 int<lower=1,upper=N_subj> subj[N];   // subject identifier
 int<lower=1,upper=N_pair> pair[N];   // pair identifier
}

parameters {
 vector[4] beta; // fixed effects: intercept, gender, type, interaction
 real<lower=0> sigma_e;                 //error sd
 vector<lower=0>[2] sigma_u;            // subj sd for intercept and slope
 cholesky_factor_corr[2] L_u;           // subj correlation matrix
 matrix[2,N_subj] z_u;                  // subjects normed intercepts and slopes
 vector<lower=0>[2] sigma_w;            // pair sd for intercept and slope
 cholesky_factor_corr[2] L_w;           // pair correlation matrix
 matrix[2,N_pair] z_w;                  // pair normed intercepts and slopes
}

transformed parameters {
 matrix[2,N_subj] u;                                //subj random effects
 matrix[2,N_pair] w;                                // item random effects
 u = diag_pre_multiply(sigma_u, L_u) * z_u;
 w = diag_pre_multiply(sigma_w, L_w) * z_w;
}

model {
 vector[N] mu;                          // mean predictor
 // Priors:
 beta ~ normal(0,3);
 L_u ~ lkj_corr_cholesky(2.0);
 L_w ~ lkj_corr_cholesky(2.0);
 sigma_u ~ gamma(2,0.1);                // keep sigma away from 0
 sigma_w ~ gamma(2,0.1);
 to_vector(z_u) ~ normal(0,1);
 to_vector(z_w) ~ normal(0,1);
 // Likelihood:
 for (i in 1:N) {
            mu[i] = beta[1] + u[1,subj[i]] + w[1,pair[i]]
            + (beta[2] + u[2,subj[i]] + w[2,pair[i]]) * gender[i]
            + beta[3] * type[i]+ beta[4] * gender[i] * type[i];
 }
 fqa ~ normal(mu, sigma_e);
}
```

A2. Model for Experiment 2

```
data {
  int<lower=0> N;                    // number of data points
  int<lower=1> N_subj;               // number of subjects
  int<lower=1> N_pair;               // number of pairs
  vector[N] fqa;                     // dependent variable (scaled)
  vector[N] gender;      // gender predictor (sum-coded: masc=-0.5, fem=+0.5)
  vector[N] language;   // language (sum-coded: DE=-0.5 or FR=+0.5)
  int<lower=1,upper=N_subj> subj[N]; // subject identifier
  int<lower=1,upper=N_pair> pair[N]; // pair identifier
}

parameters {
  vector[4] beta;          // fixed effects: intercept, gender, language, interaction
  real<lower=0> sigma_e;             //error sd
  vector<lower=0>[2] sigma_u;      // subj sd for intercept and slope
  cholesky_factor_corr[2] L_u;// subj correlation matrix
  matrix[2,N_subj] z_u;              // subjects normed intercepts and slopes
  vector<lower=0>[3] sigma_w;     // item sd for intercept and slopes
  cholesky_factor_corr[3] L_w;       // item correlation matrix
  matrix[3,N_pair] z_w;                          // item normed intercepts and slopes
}

transformed parameters {
  matrix[2,N_subj] u;    //subj random effects
  matrix[3,N_pair] w;    //item random effects
  u = diag_pre_multiply(sigma_u, L_u) * z_u;
  w = diag_pre_multiply(sigma_w, L_w) * z_w;
}

model {
  vector[N] mu;                      // mean predictor
  // Priors:
  beta ~ normal(0,3);
  L_u ~ lkj_corr_cholesky(2.0);
  sigma_u ~ gamma(2,0.1);
  to_vector(z_u) ~ normal(0,1);
  L_w ~ lkj_corr_cholesky(2.0);
  sigma_w ~ gamma(2,0.1);
  to_vector(z_w) ~ normal(0,1);
  // Likelihood:
  for (i in 1:N) {
              mu[i] = beta[1] + u[1,subj[i]]  + w[1,pair[i]]
              + (beta[2] + u[2,subj[i]] + w[2,pair[i]]) * gender[i]
              + (beta[3] + w[3,pair[i]]) * language[i]
              + beta[4] * gender[i] * language[i];
  }
  fqa ~ normal(mu, sigma_e);
}
```

Appendix B. Stimuli

B1. Person nouns selected for Experiment 1, tested in Pilot 1.

Feminine	Masculine
mères	pères
sœurs	frères
tantes	oncles
belles-mères	beaux-pères
mamans	papas
femmes	hommes
dames	messieurs
nanas	mecs
filles	garcons
reines	rois
princesses	princes
grands-mères	grands-pères

B2. Object nouns selected for Experiment 1

Feminine	Masculine	$NSyll_{Fem}$	$NSyll_{Masc}$	MeanRating	SDRating
pantoufle	chausson	2	2	5.90	0.30
bicyclette	vélo	3	2	5.88	0.33
main	doigt	1	1	5.81	0.51
pierre	caillou	1	2	5.76	0.44
soupe	potage	1	2	5.76	0.54
boutique	magasin	2	3	5.75	0.55
chaussure	soulier	2	2	5.71	0.56
courge	potiron	1	3	5.68	0.48
poubelle	déchet	2	2	5.62	0.67
flamme	feu	1	1	5.59	0.71
valise	bagage	2	2	5.58	0.77
forteresse	château	3	2	5.57	0.68
lampe	luminaire	1	3	5.55	0.69
maison	logement	2	3	5.55	0.60
veste	blouson	1	2	5.53	0.61
voiture	véhicule	2	3	5.53	0.51
lame	couteau	1	2	5.52	0.81
page	cahier	1	2	5.48	0.68
balle	ballon	1	2	5.47	0.84
peinture	pinceau	2	2	5.47	0.70
carafe	pichet	2	2	5.41	0.87
prairie	champ	2	1	5.41	0.62
terre	sol	1	1	5.41	0.71
baguette	pain	2	1	5.40	0.94
cigarette	cigare	3	2	5.38	0.67
roue	pneu	1	1	5.38	0.97
tasse	mug	1	1	5.37	1.12
bougie	cierge	2	1	5.35	0.88
revue	journal	2	2	5.32	0.67
touche	clavier	1	2	5.32	0.82
voile	bateau	1	2	5.32	0.82
boisson	breuvage	2	2	5.29	0.85
clé	cadenas	1	3	5.24	0.94
lettre	courrier	1	2	5.24	0.56
parka	manteau	2	2	5.24	0.83
tarte	gâteau	1	2	5.24	0.77
douche	savon	1	2	5.21	1.13
sucette	bonbon	2	2	5.21	0.98
mitaine	gant	2	1	5.20	0.77
cheminée	toit	3	1	5.19	0.98
laine	fil	1	1	5.19	0.68
chaise	siège	1	1	5.18	1.07
commode	meuble	2	1	5.18	0.88
ceinture	pantalon	2	3	5.16	0.90
brosse	peigne	1	1	5.14	0.79
tisane	thé	2	1	5.12	0.93
fenêtre	rideau	2	2	5.11	1.24
pluie	nuage	1	1	5.11	1.33
plage	sable	1	1	5.10	1.02
route	chemin	1	2	5.10	0.85
peluche	doudou	2	2	5.06	0.75
rivière	ruisseau	2	2	5.00	1.00

B3. Noun pairs for Experiment 2

French	German	gender	$NSyll_{Fr}$	$NSyll_{De}$	Type	English
pomme	Apfel	f/m	1	2	nature	apple
plage	Strand	f/m	1	1	nature	beach
poutre	Balken	f/m	1	2	object	beam
cathédrale	Dom	f/m	3	1	object	cathedral
chaise	Stuhl	f/m	1	1	object	chair
boussole	Kompass	f/m	2	2	object	compass
marmite	Topf	f/m	2	1	object	cooking-pot
digue	Deich	f/m	1	1	other	dike
colline	Hügel	f/m	2	2	nature	hill
clé	Schlüssel	f/m	1	2	object	key
lavande	Lavendel	f/m	2	3	nature	lavender
pelouse	Rasen	f/m	2	2	other	lawn
foudre	Blitz	f/m	1	1	nature	lightning
liqueur	Likör	f/m	2	2	object	liquor
lune	Mond	f/m	1	1	nature	moon
montagne	Berg	f/m	2	1	nature	mountain
moutarde	Senf	f/m	2	1	other	mustard
planète	Planet	f/m	2	2	nature	planet
raquette	Schläger	f/m	2	2	object	racket
pluie	Regen	f/m	1	2	nature	rain
biscotte	Zwieback	f/m	2	2	other	rusk
neige	Schnee	f/m	1	1	nature	snow
cuillère	Löffel	f/m	2	2	object	spoon
courge	Kürbis	f/m	1	2	nature	squash
pierre	Stein	f/m	1	1	other	stone
valise	Koffer	f/m	2	2	object	suitcase
table	Tisch	f/m	1	1	object	table
tuile	Ziegel	f/m	1	2	object	tile
tour	Turm	f/m	1	1	object	tower
tablier	Schürze	m/f	3	2	object	apron
haricot	Bohne	m/f	3	2	nature	bean
pont	Brücke	m/f	1	2	object	bridge
beurre	Butter	m/f	1	2	other	butter
canon	Kanone	m/f	2	3	object	cannon
chocolat	Schokolade	m/f	3	4	other	chocolate
cigare	Zigarre	m/f	2	3	object	cigar
nuage	Wolke	m/f	1	2	nature	cloud
cordon	Schnur	m/f	2	1	object	cord
concombre	Gurke	m/f	2	2	nature	cucumber
dôme	Kuppel	m/f	1	2	object	dome
ferry	Fähre	m/f	2	2	object	ferry
drapeau	Fahne	m/f	2	2	object	flag
fruit	Frucht	m/f	1	1	nature	fruit
pistolet	Pistole	m/f	3	3	object	gun
citron	Zitrone	m/f	2	3	nature	lemon
melon	Melone	m/f	2	3	nature	melon
lait	Milch	m/f	1	1	other	milk
journal	Zeitung	m/f	2	2	object	newspaper
chêne	Eiche	m/f	1	2	nature	oak tree
palmier	Palme	m/f	2	2	nature	palm tree
persil	Petersilie	m/f	2	4	nature	parsley
coing	Quitte	m/f	1	2	nature	quince
ravin	Schlucht	m/f	2	1	nature	ravine
savon	Seife	m/f	2	2	object	soap
potage	Suppe	m/f	2	2	other	soup
soleil	Sonne	m/f	2	2	nature	sun
vase	Vase	m/f	1	2	object	vase
caveau	Gruft	m/f	2	1	object	vault
yacht	Jacht	m/f	1	1	object	yacht

B4. Object nouns tested in Pilot 1

Feminine	Masculine
chaise	fauteuil
tasse	verre
balle	ballon
plume	stylo
boîte	carton
vis	clou
scie	couteau
bouteille	flacon
guitare	violon
pelle	balai
table	bureau
pierre	rocher

B5. Nouns pairs tested in Pilot 2

French	German	Gender	English
pont	Brücke	m/f	bridge
fontaine	Brunnen	f/m	fountain
château	Burg	m/f	castle
chat	Hütte	m/f	cottage
collier	Halskette	m/f	necklace
bague	Ring	f/m	ring
rivière	Fluss	f/m	river
montagne	Berg	f/m	mountain
nuage	Wolke	m/f	cloud
pluie	Regen	f/m	rain
clé	Schlüssel	f/m	key
fourchette	Gabel	f/m	fork
flèche	Pfeil	f/m	arrow
pistot	Pistole	m/f	gun
moustique	Mücke	m/f	mosquito
serpent	Schlange	m/f	snake
pomme	Apfel	f/m	apple
raisin	Traube	m/f	grape
chaussure	Schuh	f/m	shoe
ceinture	Gürtel	f/m	belt
mariage	Hochzeit	m/f	wedding
voyage	Reise	m/f	trip
pensée	Gedanke	f/m	thought
besoin	Notwendigkeit	m/f	need

Chapter 3. How fair is gender-fair language? Insights from gender ratio estimations in French

Hualin Xiao[1,2,4], Brent Strickland[1,3,4] & Sharon Peperkamp[1]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS–PSL, EHESS, CNRS), Paris, France

[2] Institut Jean Nicod (ENS–PSL, EHESS, CNRS), Paris, France

[3] Africa Business School, UM6P, Rabat, Morocco

[4] School of Collective Intelligence, UM6P, Rabat, Morocco

This section presents the manuscript "How fair is gender-fair language? Insights from gender ratio estimations in French", with two additional pilot studies that were run on pronouns, using a similar design as the one in the described experiments. The manuscript is currently under review at *Journal of Language and Social Psychology*.

Abstract

Heated societal debates in various countries concern the use of so-called “gender-fair language”, meant to replace the generic use of grammatically masculine forms. Advocates and opponents of gender-fair language disagree on – among other things – the question of whether masculine forms leave women underrepresented in people’s minds. We investigated the influence of linguistic form on the mental representation of gender in French. Participants read a short text about a professional gathering and estimated the percentage of women present at the gathering. Results showed higher estimates in response to two gender-fair forms relative to the masculine form. Comparisons with normed data on people’s perception of real-world gender ratios additionally showed that the gender-fair forms removed or reduced a male bias for neutral- and female-stereotyped professions, respectively, yet induced a female bias for male-stereotyped professions. Thus, gender-fair language increases the prominence of women in the mind, but has varying effects on the consistency of mental representations.

Keywords:

Grammatical gender, Masculine generics, Gender-fair language, *Ecriture inclusive*, French

Introduction

Languages belonging to various language families have sex-based grammatical gender systems (Corbett, 1991; Gygax et al., 2019). In these languages, each noun has either masculine or feminine (or, in some languages, neutral) gender, triggering agreement in words such as articles, adjectives, and pronouns. When referring to groups of humans, the masculine and feminine plural genders are used asymmetrically. That is, masculine plural forms are typically used to refer to: 1/ groups of men only, 2/ groups of men and women, and 3/ groups for whom the gender of the referents is unknown. In the latter two cases the masculine gender has a generic meaning. Feminine plural forms, by contrast, can only be used to refer to groups unambiguously and exclusively composed of women.

An example from French is shown in (6). As the subject is presented in a masculine plural form (6a), the sentence can be interpreted in three ways: 1/ male cashiers are on strike; 2/ male and female cashiers are on strike; 3/ cashiers whose gender is unknown are on strike. However, if presented in the feminine form (6b), it unambiguously indicates that only female cashiers are on strike.

(6) **a.** *Les caissiers sont en grève.*

'The cashiers$_{masc}$ are on strike.'

b. *Les caissières sont en grève.*

'The cashiers$_{fem}$ are on strike.'

The asymmetric roles of the masculine and feminine forms have intersected with heated social debates about gender equality in France and other countries (Bodine, 1975; Elmiger, 2008),

as this asymmetry in language might reproduce and perpetuate an unequal social status between men and women (Menegatti & Rubini, 2018; see also Ng, 2007). According to one view, the default for masculine forms to represent a mixture of both genders leaves women underrepresented in language and hence underrepresented in the mind. The idea is that in seeing or hearing the masculine form, people are less likely to think of women, which in turn affects the way that women's roles in society are mentally represented. Proponents of this view thus argue for the replacement of the generic use of the masculine by alternative, gender-fair, linguistic forms in order to increase the visibility of females in language, and consequently – by hypothesis – in people's mental representations (for a review, see Sczesny et al., 2016).

There are several gender-fair alternatives to the generic masculine form (Abbou, 2011). One common one is a "double-gender" form, illustrated in (7).

(7) *Les caissiers et caissières sont en grève.*

'The cashiers$_{masc}$ and cashiers$_{fem}$ are on strike'.

Another type of alternative is limited to written language only. It consists of the use of a contracted form, for instance by means of parentheses (*les caissiers(ères)*) or a slash (*les caissiers/ères*). In French, a more radical innovation in this area was instigated in 2015 by the *Haut Conseil à l'Egalité entre les femmes et les hommes* ('High Council for Equality between women and men'), namely a contracted form featuring a middot, as shown in (8).

(8) *Les caissier·ère·s sont en grève.*

'The cashiers$_{masc.fem}$ are on strike'.

This form is largely known as *écriture inclusive* ('inclusive writing'), but here we will use the term "middot form", as *écriture inclusive* is really an umbrella term for multiple strategies against gender-stereotyped communication (see, for instance, the *Manuel d'écriture inclusive* by the French communication agency *Mots-Clés*[6]).

Replacing the generic masculine gender with an innovative and inclusive language form is not restricted to French. For instance, German has seen the introduction of gender-neutral nominalized adjectives and participles, e.g. *die Studierenden* 'the students' (cf. *studieren* 'to study') (Sato et al., 2016), as well as that of a word-internal capital 'I', as in *LeserInnen* 'readers$_{masc\text{-}fem}$' (cf. *Leser* 'readers$_{masc}$ and *Leserinnen* 'readers$_{fem}$'; note that all German nouns are spelled with an initial capital). The latter is a spelling innovation that to some extent resembles the French middot.

A possible skeptical position concerning the use of gender-fair alternatives to the masculine form argues that linguistic forms have little influence on how people think about gender roles. According to this view, the generic use of the masculine gender does not bias people's mental representation against women and hence it is unnecessary to use a longer double-gender or an unconventional, deliberately invented form. In France, stronger forms of skepticism are targeted specifically towards the middot form, which is particularly controversial, including among linguists.[7] This form is argued to damage orthography, creating confusion and obstacles in learning to read and write. In its solemn declaration, *L'Académie Française* admonished French society for the idea of "inclusive writing" (meant is the middot form), warning people about its potential to estrange future generations from France's written heritage,

[6] https://www.motscles.net/ecriture-inclusive
[7] See for instance https://www.marianne.net/agora/tribunes-libres/une-ecriture-excluante-qui-s-impose-par-la-propagande-32-linguistes-listent-les

undermine the status of French as a world language, and even put French in mortal danger.[8] Accordingly, in 2017 the prime minister recommended not to use it in official texts.[9] Going one step further, a number of parliamentarians proposed a bill in 2021, aiming to prohibit and penalize the use of the middot form in public administrations and organizations in charge of public services,[10] and shortly after that, the minister of National Education ordered that it not be taught in schools.[11] Despite these governmental restrictions, the middot form has become more and more widespread over the years. For instance, it appears in most Parisian universities' brochures for some of their 2019–2020 undergraduate programs (Burnett & Pozniak, 2020).

Societal debates notoriously take place without reference to empirical evidence. Yet, there is a growing body of experimental work assessing the interpretation of both the masculine generic form and gender-fair alternatives in languages such as English, French and German (for a review, see Menegatti & Rubini, 2018). Below, we will first review this research and then introduce our approach to addressing some outstanding questions in the current study.

Previous research

Previous studies on the interpretation of various linguistic forms have employed a variety of paradigms, with dependent measures such as sentence reading time, sentence plausibility judgment, proportion of women estimated to be present in a group of people described in a short text, and proportion of favorite women mentioned for a given profession. With regards to English, a so-called natural gender language (i.e. a language that marks gender on personal pronouns only), the generic singular pronoun *he* was found to favor the presence of men in people's mental representations compared to singular *they* and the alternative *he/she* (Gastil,

[8] https://www.academie-francaise.fr/actualites/declaration-de-lacademie-francaise-sur-lecriture-dite-inclusive
[9] https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000036068906
[10] https://www.assemblee-nationale.fr/dyn/15/textes/l15b4003_proposition-loi
[11] https://www.education.gouv.fr/bo/21/Hebdo18/MENB2114203C.htm

1990; Hamilton, 1988; Martyna, 1978). As for the masculine plural form of nouns, several studies have provided evidence that it likewise disfavors the presence of women in mental representations (Brauer & Landry, 2008; Braun et al., 1998; Gabriel & Mellenberger, 2004; Gygax et al., 2008, 2012; Gygax & Gabriel, 2008; Horvath et al., 2016; Irmen, 2007; Irmen & Roßberg, 2004; Kollmayer et al., 2018; Stahlberg et al., 2001). Most of them manipulated the gender stereotype of the noun – typically a role name (e.g. golfer, cashier, or spectator) (Brauer & Landry, 2008; Braun et al., 1998; Gygax et al., 2008, 2012; Gygax & Gabriel, 2008; Horvath et al., 2016; Irmen, 2007; Irmen & Roßberg, 2004). As such stereotypes are activated during reading (Banaji & Hardin, 1996; Cacciari & Padovani, 2007; Carreiras et al., 1996; Garnham et al., 2002; Gygax & Gabriel, 2008; Kennison & Trofe, 2003; Reynolds et al., 2006), the generic - as opposed to the specific - interpretation of the masculine gender should be especially plausible for female-stereotyped groups. Yet, even for those groups the results revealed male biases. For instance, Gygax et al. (2008) presented French and German participants with two sentences. The first one named a group of professionals in the masculine plural form (e.g. *les espions* 'the spies$_{masc}$'); the second sentence provided explicit gender information about one or more of the people in the group, and participants had to decide whether it was a sensible continuation of the first one. The authors compared three types of gender-stereotyped professions, chosen from a norming study they had run previously (Gabriel et al., 2008): masculine (e.g. spies), feminine (e.g. beauticians) or neutral (e.g. singers). Results showed that more positive responses were given when the second sentence explicitly referred to men rather than to women, regardless of the profession's stereotype. Thus, French and German speakers were more likely to match professionals presented in a masculine plural form with men than with women, even for female-stereotyped professions. (The experiment was also run in English, which does not have

grammatically gendered nouns. English participants did show an effect of gender stereotype, such that continuation sentences referring to women or to men were deemed more likely for female- or male-stereotyped professions, respectively). Gygax et al. (2012) additionally found that this biased interpretation of masculine generics could be reduced but not suppressed if participants were explicitly reminded of the generic meaning of masculine forms.

Other studies have compared masculine plurals with gender-fair alternatives (Brauer & Landry, 2008; Braun et al., 1998; Gabriel & Mellenberger, 2004; Gygax & Gabriel, 2008; Horvath et al., 2016; Kollmayer et al., 2018; Sato et al., 2016; Stahlberg et al., 2001; Stahlberg & Sczesny, 2001). Crucially, double-gender forms and – in German – the innovative capital-I and nominalized forms typically yield a stronger representation of women than masculine forms. Only a few studies have examined whether this effect of grammatical gender is modulated by stereotype, with mixed results: Braun et al. (1998) presented German participants with a short text about an annual meeting of a professional group, and asked them a few hours later to estimate the percentage of women present in the group. They found that compared to the masculine form, the double-gender form yielded higher estimated percentages of women for male- but not for female-stereotyped professional groups. Applying the same paradigm to French, however, Brauer & Landry (2008) observed a global increase in estimated percentage of women when a double-gender form was presented, relative to when a masculine form was displayed, but this effect was not modulated by the gender stereotype.

Finally, there have been some empirical studies on the potential difference between double-gender and innovative gender-fair forms. For German, Stahlberg et al. (2001) asked participants to name famous people in a given category (e.g. singers or politicians). They found that the use of the capital-I form in the question yielded higher proportions of women than that of

a double-gender form (the latter yielding no higher proportion of women than the masculine form – a rare result). In Stahlberg & Sczesny (2001), participants were presented with the written name of a professional category, followed by a picture of a famous person; their task was to indicate whether the person belonged to the category. For female pictures, reaction times were faster when the name of the category was written in the capital-I form compared to the double-gender form. (Reaction times were slowest for the masculine gender condition.)

Concerning the effect of innovative forms, two more studies are worth mentioning. Both deal with Swedish, a largely non-gendered language but which, like English, has gender marking on personal pronouns. Interestingly, Sweden has introduced a gender-neutral pronoun, *hen*, complementing *han* 'he' and *hon* 'she'. Following a few years of debate, *hen* was officially adopted in 2015, quickly reaching widespread adherence and use (Gustafsson Sendén et al., 2015). As shown by Tavits & Pérez (2019) and Lindqvist et al., (2019), *hen* boosts the presence of women in mental representations. For instance, participants in Lindqvist et al., (2019) read a job advertisement for a profession that according to official Swedish statistics is gender-balanced, followed by the description of an applicant. They then had to choose a picture of either a man or a woman, indicating who they thought the applicant was. When the applicant was referred to by means of a non-gendered noun (*den sökande* 'the applicant') the results showed a male bias; that is, a picture of a man was chosen more than 50%. By contrast, there was no such bias when the applicant was referred to by means of the paired pronouns *han/hon* 'he/she' or the gender-neutral pronoun *hen*, both yielding about 50% choices of a picture of a man.

This last study is important for another reason: To the best of our knowledge it is the only one with a clear quantitative inference concerning the presence or absence of a *bias* induced by a given linguistic form in mental representations. Indeed, as participants made a forced choice

between a male and a female applicant to a gender-neutral job, the data from each condition could be compared not only to that of the other conditions but also to the objective real-world baseline chance level of 50%. Sweden fares particularly well on closing the gender-equality gap (according to the 2020 Global Gender Gap Report of the World Economic Forum, it is ranked 4th), which makes the presence of the male bias when a neutral, non-gendered, noun is used particularly striking. (Note, though, that their experiment tested only a single profession.)

In summary, previous research with gendered languages such as French and German has shown that compared to the masculine plural form, the double-gender and innovative forms boost the presence of women in mental representations. Whether the effect of linguistic form is modulated by stereotype, though, is largely an open question. Furthermore, while in German there is a difference between double-gender forms and the innovative capital-I, with the latter yielding the largest difference compared to the masculine form, no research has yet examined the effect of the middot writing form in French, and its potential difference in comparison with a double-gender form. Finally, the question of whether in these languages a given linguistic form induces mental representations that consistently reflect the proportions of men and women – real or perceived – in specific societal groups or in the society as a whole is yet unanswered. That is, establishing that the presence of women in mental representations differs depending on linguistic form is one thing; a quantified comparison with a benchmark (i.e. census data or normed estimates of real-world gender ratios) would additionally allow one to establish which linguistic form – if any – induces *consistent* mental representations, and to estimate the size of the bias (male- or female-oriented) induced by the other forms.

Current study

In order to shed light on the three open questions mentioned above, we conduct two on-

line experiments on the mental representation of gender in text-based inferences in French. We compare the masculine, double-gender, and middot plural forms. Focusing on gender ratios in professional groups, we consider both gender-neutral professions (Experiment 1) and gender-stereotyped ones (Experiment 2). Importantly, we use a ratio estimation task, which allows us to compare our participants' responses to independently normed estimates of the proportion of men and women in the relevant professions (Misersky et al., 2013), Comparison to such a benchmark is of critical importance to the question whether a given linguistic form induces a bias in mental representations.

Our experimental paradigm is an adapted version of the one used by Braun et al. (1998) and Brauer & Landry (2008), in which participants read a short text on a professional gathering and are asked to estimate the percentage of women present at the gathering. Our most important modifications are that we test a variety of professions, such as to ensure that any observed effect generalizes across professions, and that, for practical reasons, we ask participants to provide their estimate immediately after having read the text rather than a few hours later. Additionally, we control for a possible effect of question framing by asking participants to estimate the percentages of men and women on a response slider, counterbalancing the order of appearance of the words 'men' and 'women' and the corresponding slider layout. We also use a shorter text, with two instead of four occurrences of the crucial piece of information. Like in these previous studies, though, we test participants on a single trial, since exposure to multiple trials might make them become aware of the experimental manipulations and develop a response strategy. Lastly, the professions are chosen from the French part of a norming study in which native speakers estimated the proportions of men and women in a great many professions (Misersky et al., 2013). As in this norming study the influence of linguistic form was controlled at best (the endpoints of

the rating scale showed the masculine and feminine forms, respectively, and the direction of the rating scale was counterbalanced across participants), these same norms also serve as the benchmark against which we compare our participants' estimates. (We acknowledge the fact that Misersky et al. collected their French data in Switzerland, while we test participants in France. Yet, this difference is unlikely to impact the benchmark's validity, given that high correlations were obtained across all seven languages investigated in the norming study (the six others were English (UK), German, Norwegian, Italian, Czech, and Slovak).

Experiment 1

In this experiment, we compare three plural forms in French that can be used to refer to a group of mixed genders: the masculine, double-gender and middot forms. Importantly, we focus on professions with a perceived neutral stereotype, which should encourage participants to interpret the masculine plural form as generic, referring to both men and women, and hence potentially treat it on a par with the double-gender and middot forms. By comparing the masculine form to these alternatives, we test whether differences in linguistic form alter language users' inferences about gender ratios in the described scenarios. Specifically, we examine whether due to the ambiguity of the masculine form and its lack of explicit inclusion of female referents, this form disfavors women in mental representations relative to the other two forms. If this were the case, we should observe lower estimates of %-women in response to the masculine compared to the double-gender and middot forms. The latter two are both gender-fair but differ in two respects: First, the middot is a more recent and more militant form than the double-gender, and second, the double-gender can be read aloud straightforwardly while the middot is essentially a spelling convention. We might therefore observe differential effects of these gender-fair forms, although

it is unclear which form would be expected to boost the mental representation of women more. Finally, we examine for each form the extent to which the mean estimate of %-women deviates from people's perception of real-world gender ratios in the professions at hand. We expect that in this respect, the double-gender and middot forms fare better (i.e., closer to a consistent representation) than the masculine form.

Unless otherwise specified, all aspects of the stimuli, procedure and analyses were preregistered (10.17605/OSF.IO/K649W).

Method

Stimuli

We selected six neutral-stereotyped professions whose French names have grammatical gender marking from the French part of Misersky et al. (2013)'s norming study. With one exception (*employé de banque* 'bank employee'), the masculine and feminine forms differ not only orthographically but also phonologically (The pattern of results was the same when we removed the trials with this noun from the analyses). The estimated proportions of women in these professions are between .47 and .51 (M = .49, SD = .01) (see Appendix A).

We constructed the short passage shown in (9), describing a fictitious scenario where an annual gathering of some professionals took place. The profession name appeared twice in the text and no referential pronoun was used.

(9) *Le rassemblement régional des PROFESSION NAME a eu lieu cette semaine à Amiens. La localisation centrale de cette ville a été particulièrement appréciée. Les PROFESSION NAME ont aussi adoré l'apéro offert à l'hôtel de ville le premier jour.*

'The regional gathering of PROFESSION NAME took place this week in Amiens. The central location of this city was particularly appreciated. The PROFESSION NAME also loved the aperitif offered at City Hall on the first day.'

The words régional and Amiens were replaced with européen and Francfort respectively for three professions, i.e. mathématicien 'mathematician', douanier 'customs officer' and astronaute 'astronaut', for it would be more plausible for these professionals to have a Europe-wide gathering in a more internationally-oriented city than a regional one in a provincial city. We constructed three versions of this passage, varying the linguistic form of the profession name (masculine vs. double-gender vs. middot plural), as exemplified for one of the professions in (10).

(10) Three forms of sample profession *musicien* 'musician'

a. masculine: *musiciens*

b. double-gender: *musiciens et musiciennes*

c. middot[12]: *musicien.ne.s*

We also constructed two multiple-choice questions about the text, to serve as attention checks:

(11) Attention check questions

a. Qu'est-ce qui a été apprécié à propos d'Amiens ?

[12] The middot "·" was replaced with the normal dot "." in the two experiments reported here, as is often the case in everyday language use.

‘What was being appreciated about Amiens?’

b. Qu'est-ce qui a été offert à l'hôtel de ville ?

‘What was offered at the City Hall?’

Procedure

The experiment was run via Qualtrics online software (https://www.qualtrics.com). Each participant was tested on a single trial and was paid 0.50 € for their time.

Participants were randomly assigned to one of the three linguistic forms (masculine, double gender, or middot). Within each group, they were randomly presented with one profession from the relevant list.

Once they agreed on the informed consent, participants were told that in the survey, they would be shown two very short texts to read and answer a few questions about. After reading the first text, which unbeknownst to the participants was a warm-up text, they had to answer two multiple-choice questions related to its contents (see Appendix C for the text and the questions), with the text still being visible. Depending on their responses, they received either positive or negative written feedback. Before they moved on to the next page of the survey and were presented with the target text on the professional gathering, they were told that questions about the following text would be more difficult. This was done to prompt them to read the following text with full attention. Then, they first read the text at their own pace; after a button press, the text disappeared and three questions were shown. The first two were the attention check questions shown in (11) above; if participants made at least one error, they were still able to finish the experiment but their data were excluded from the analyses. The third question asked them to provide their estimate of the gender ratio in the fictional gathering. There were two

versions of it, depending on the order of the words for men and women in the sentence. One version was framed as *Selon vous, quels étaient les pourcentages d'hommes et de femmes dans ce rassemblement?* (‘In your opinion, what were the percentages of men and women in the gathering?’) (men-women version) and the other, *Selon vous, quels étaient les pourcentages de femmes et d'hommes dans ce rassemblement?* (‘In your opinion, what were the percentages of women and men in the gathering?’) (women-men version). There were likewise two versions of the response slider, such that the labels for the left and right endpoints, i.e. pictograms of a man and a woman, reflected the framing of the test question. Instructions on how to use the slider were shown with the question and remained visible to participants during the test. (See Appendix D for the slider and instructions.) Slider version (and hence, question framing), was counterbalanced within each of the three groups.

At the end of the survey, participants were asked to fill in information about their native language, country of residence, gender, and age.

Participants

We recruited 195 participants. The data from 42 of them were removed from the analysis for the following reasons: one participated in a related experiment not reported on here, one did not complete the survey, 17 responded incorrectly at one or both attention check questions, and 23 did not satisfy all of our recruitment criteria (three were non-native speakers, two did not live in France, and among the ones recruited on Foulefactory, 18 reported an age outside of the requested range). For the 41 participants who took the survey more than once, we only kept their first response.

The data analysis thus included 153 participants (67 women and 86 men). They were native French speakers living in France, aged between 22 and 39 years (M = 30, SD = 2.7).

Three of them were recruited on the crowd-sourcing platform Clickworker (https://www.clickworker.com/), and all others on Foulefactory (https://www.foulefactory.com), the preregistered recruitment platform, has a large number of French workers, but presents two disadvantages compared to Clickworker. First, it only offers pre-specified age ranges, although a customized age range can be obtained for an extra 500€. Here, we chose the pre-specified range 25-34. Second, it lacks a good screening function: workers can do a task more than once, leading to considerable data loss. We therefore completed our sample by collecting the last three datapoints on Clickworker, with a customary age range of 20-40 years).

Their random assignment to one of six conditions (three linguistic forms x two slider layouts) yielded a mean number of 25 participants (min = 24, max = 26) per condition.

Results and Discussion

Boxplots of the estimated percentages of women as a function of linguistic form are shown in Figure 3.

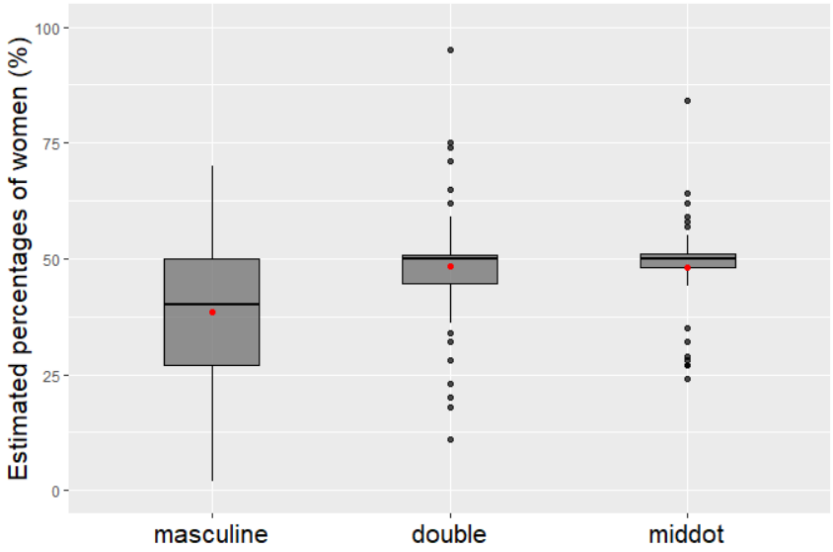


Figure 3. Boxplots of estimated percentages of women as a function of linguistic form. Medians are indicated by black lines, means by red dots.

These data were fit with a linear mixed-effects model by using the *lme4* package (Bates et al., 2015) in the programming software *R* (R Core Team, 2020) and *Rstudio* (RStudio Team, 2020) Statistical significance was assessed by means of the Anova function in the *Car* package (Fox & Weisberg, 2019), and effect sizes were computed using the eta_squared function in the *effectsize* package (Ben-Shachar et al., 2020). Linguistic form was contrast-coded and set as fixed effect; a random intercept was added for Profession. As we collected only one datapoint per person, we did not include a random factor for Participant. Note that this model differs from the one we preregistered in that it does not contain fixed factors for Slider version and its interaction with Linguistic form. As neither in this nor in the next experiment this counterbalancing factor or any of its interactions affected the estimated %-women, and as omitting these terms did not change any of the results, we report the simpler models in both

experiments for the reader's convenience. (The only other studies we know of that looked at order effects are the norming studies by Gabriel et al. (2008) and Misersky et al. (2013). For some languages, these studies reported higher estimates for women when the question framing and response slider showed a women-men order, but there was no such effect for French.)

The results of the mixed-effects model, shown in Table 1, revealed an effect of Linguistic form. Restricted analyses with corrections for multiple comparison (mvt method), carried out with the *emmeans* package (Lenth et al., 2020), showed that compared to the masculine plural form, higher estimates of %-women were obtained for the double-gender ($\beta = 9.92$, SE = 2.64, $t(146) = 3.76$, $p < .001$) and the middot form ($\beta = 9.63$, SE = 2.63, $t(146) = 3.67$, $p < .001$). By contrast, there was no difference between the double-gender and middot forms ($|t| < 1$).

Table 1. Results of linear mixed-effects model for Experiment 1

	β	*SE*	*t*	χ^2	*Df*	*p*	*partial* η^2
Form				18.54	2	< .0001	0.11
double	3.41	1.53	2.23				
middot	3.11	1.52	2.04				

Next, we carried out non-preregistered post-hoc analyses to compare the results to Misersky et al.'s (2013) norming data shown in Appendix A. To do this, we subtracted the normed %-women from the participant's estimate for each profession and each participant, and constructed intercept-only models with this difference score as dependent measure and a random intercept for Profession. In these models, a positive estimate for the intercept would thus indicate an overestimation of the presence of women compared to the benchmark and a negative estimate

an underestimation. We used the lmer function of the *lmerTest* package (Kuznetsova et al., 2017) such as to obtain p-values. We found that the masculine form yielded an underrepresentation of women compared to the benchmark (β = -10.66, SE = 2.73, t = -3.91, $p < .02$), while estimates for the double-gender and middot forms did not differ from the benchmark values (double: β = -0.80, $|t| < 1$; middot: β = -1.01, $|t| < 1$). In other words, these results suggest that masculine plural induces an 11% point of male bias, while both the alternative forms induce a consistent representation of the proportion of women.

No previous study has focused specifically on a neutral stereotype. Yet, our finding that participants inferred a higher percentage of women when the double-gender or middot form was presented relative to the masculine form meshes well with the results of Gygax et al. (2008, 2012), who examined neutral-stereotyped role names alongside male- and female- stereotyped ones. Indeed, using a different paradigm they observed a male bias regardless of stereotype in both French and German. Similarly, the finding that the masculine form induces a male bias is in accordance with the Swedish study of Lindqvist et al. (2019).

Experiment 2

Experiment 1 showed an influence of linguistic form on estimations of gender ratios for professions without a gender stereotype. In this experiment, we focus on male- and female-stereotyped professions, and compare the same linguistic forms as before, i.e. masculine, double-gender and middot. This design allows us to examine whether the effect of linguistic form is modulated by stereotype. As double-gender and middot forms might promote women's visibility especially in cases where they are typically a minority gender, we expect – if anything – a larger effect of linguistic form for male-stereotyped professions than for female-stereotyped ones.

Furthermore, given the lack of a difference between the double-gender and middot forms in Experiment 1, we expect these two forms likewise to yield similar results. Finally, as in Experiment 1, we also examine to what extent participants' estimates in the experimental context reflect people's perceived gender ratios in the real world, by comparing the results to the norming data of Misersky et al. (2013). If the double-gender and middot forms yield more consistent mental representations, they should fare better than the masculine form.

Unless otherwise specified, all aspects of the stimuli, procedure and analyses were preregistered on OSF (10.17605/OSF.IO/FCEWA).

Method

Stimuli

We selected six male-stereotyped professions (e.g. *électricien* 'electrician$_{masc}$' – *électricienne* 'electrician$_{fem}$') and six female-stereotyped ones (e.g., *caissier* 'cashier$_{masc}$' – *caissière* 'cashier$_{fem}$') from the same norming study as used in Experiment 1 (Misersky et al., 2013). For all professions, the feminine plural form of their French name differs from the masculine one not only orthographically but also phonologically. The mean estimated proportions of men or of women, respectively, were above .70 (male-stereotyped: M_{men} = .81 SD = .03; female-stereotyped: M_{women} = .78, SD = .05; t(10) = 1.12, p = 0.3). The 12 professions, together with the estimates of %-women in those professions – according to Misersky et al.'s norming study – are shown in Appendix A.

We used the same testing passage in the same three versions as in Experiment 1, shown in (9) and (10) above.

Procedure

Participants were randomly assigned to one of 12 groups obtained by crossing the two stereotypes, three linguistic forms, and two slider versions. Within each group, participants were randomly shown one of the six professions from the relevant list (male- or female-stereotyped). The procedure was otherwise identical to the one for Experiment 1.

Participants

We recruited 438 participants. The data from 133 of them were removed from the analysis for the following reasons: 33 did not complete the survey, 36 participated in Experiment 1 or a related experiment not reported in this article, 28 responded incorrectly at one or both attention check questions, and the remaining ones did not satisfy all of our recruitment criteria: 15 were non-native speakers, two did not live in France, 12 recruited on Foulefactory were older than 34, and seven recruited on Clickworker were younger than 20 (For the same reason as in Experiment 1, the selected age range for participants on Foulefactory was 25-34, while the range for those on Clickworker was 20-40. The data from the 12 on Foulefactory who indicated being older than 34 were again removed because of the conflict with the registered age in Foulefactory's database, even though they indicated being younger than 40). We also excluded the second response of 38 participants who took the survey twice.

The data analysis thus included 305 participants (158 women, 145 men, and 2 other gender). They were native French speakers living in France, aged between 20 and 40 years (M = 28, SD = 5.3). They participated on the crowd-sourcing platforms Foulefactory (N = 79) and Clickworker (N = 226) and none of them had participated in the previous experiment.

The random assignment of participants to one of 12 groups (two stereotypes x three linguistic forms x two slider layouts) resulted in a mean number of 25 participants per condition

(min = 24, max = 28).

Results and Discussion

Boxplots of estimated percentages of women as a function of linguistic form and stereotype are shown in Figure 4.

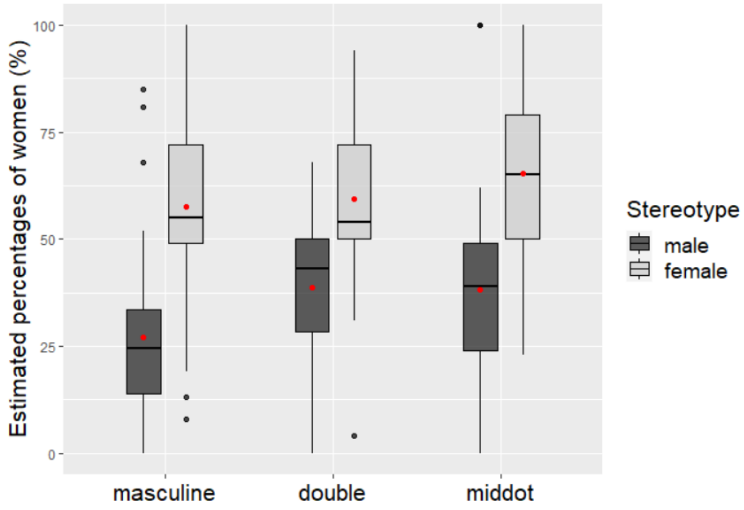


Figure 4. Boxplots of estimated percentages of women as a function of linguistic form and stereotype. Medians are indicated by black lines, means by red dots.

As in Experiment 1, the data were fit with a linear mixed-effects model. The model contained fixed factors for contrast-coded Stereotype, Linguistic form, and their interaction, as well as a random intercept for Profession. The results, shown in Table 2, revealed effects of Stereotype and Linguistic form, but no interaction.

Table 2. Results of linear mixed-effects model for Experiment 2

	β	*SE*	*t*	χ^2	*Df*	*p*	*partial* η^2
Stereotype(male)	-13.2	1.73	-7.64	58.4	1	< .0001	0.85
				14.8	2	< .001	0.05
Form	1.62	1.48	1.10				
double	3.91	1.49	2.62				
inclusive				3.82	2	0.15	0.01
Stereotype × Form	2.66	1.48	1.80				
male:double	-0.38	1.49	-0.26				
male:middot							

Relative to female-stereotyped professions (M = 60.8, SD = 19.6), lower percentages of women were obtained for male-stereotyped ones (M = 34.6, SD = 18.7; β = -13.2, SE = 1.73, t = -7.64, p < .001). Restricted analyses with corrections for multiple comparison (mvt method) showed that compared to the masculine form, higher estimates of %-women were obtained for the double-gender form (β = 7.16, SE = 2.55, t(290) = 2.81, p < .02) and the middot form (β = 9.44, SE = 2.57, t(290) = 3.67, p < .001), while there was no difference between the latter two (t < 1).

In order to compare the results of this experiment to the ones by Braun et al. (1998) and Brauer & Landry (2008), who tested only masculine and double gender forms across both stereotypes, we also performed the same regression analysis without the data for the middot form. The results of this analysis, which was not preregistered, did reveal an interaction of small effect size ($\beta = 2.45$, SE = 1.27, t = 1.98, $\chi^2 = 3.92$, $p < .05$, partial $\eta^2 = 0.02$), such that for male-stereotyped professions, the double form yielded higher estimates of %-women than the masculine form ($\beta = 12.0$, SE = 3.52, t(194) = 3.42, $p < .001$), while for the female-stereotyped professions no difference was found between the linguistic forms ($t < 1$).

Finally, following the same procedure as in Experiment 1, we carried out non-preregistered post-hoc analyses to compare the results to Misersky's et al. (2013) norming data shown in Appendix A. (In two of the models, i.e. male stereotype with double-gender form, and female stereotype with middot form, the random factor was not taken into account by lmerTest because its estimated variance was zero or close to zero.) For the male-stereotyped professions, we found that the double-gender and middot forms yielded an overrepresentation of women (double-gender: $\beta = 19.26$, SE = 2.02, t = 9.52, $p < .0001$; middot: $\beta = 19.18$, SE = 3.00, t = 6.40, $p < .003$), with the masculine form trending in the same direction ($\beta = 7.47$, SE = 3.19, t = 2.34, $p < .07$). For the female-stereotyped professions, conversely, we found that all three forms induced an underrepresentation of women, which – as indicated by the values of the beta coefficient – is numerically largest for the masculine and smallest for the middot form (masculine: $\beta = -20.24$, SE = 3.63, t = -5.57, $p < .003$; double-gender: $\beta = -17.89$, SE = 2.97, t = -6.02, $p < .002$; middot: $\beta = -12.39$, SE = 2.83, t = -4.37, $p < .0001$).

These results show that both stereotype and linguistic form affect inferences about gender ratios in male- and female-stereotyped professions. Note that the overall effect of linguistic form

was similar to the one observed for neutral-stereotyped professions in the previous experiment: the double-gender and middot forms yield higher estimates of %-women than the masculine form, and they do so to the same extent. The lack of a global interaction between linguistic form and stereotype is unexpected, since we did observe such an interaction in an almost identical pilot experiment with less participants (N=142); that is, the increase in the estimations with the gender-fair forms was restricted to male-stereotyped professions. Here, we observed the interaction only in a restricted analysis without the data for the middot form. The same interaction pattern was present in the German data of Braun et al. (1998), but not in the French data of Brauer & Landry (2008). As the effect sizes of the interactions in the present experiment and in our pilot experiment are both small, statistical power might be at issue to explain the overall now-you-see-it-now-you-don't pattern. We return to this point in the General Discussion.

As to the comparison with the norming data, the results suggest that none of the linguistic forms induce a consistent representation of the proportion of women for either male- or female-stereotyped professions. This contrasts with results for neutral-stereotyped professions in Experiment 1, where the gender-fair language forms indeed matched the normed ratios. Table 3 provides an overview of the comparisons with the norming data both for this experiment and the previous one, by showing the models' estimated differences, measured in percentage points, between our participants' estimates and the norms. A positive value indicates a female bias, a negative one a male bias.

Table 3. Estimated percentage point differences between participants' estimates in Experiments 1 and 2 and Misersky et al.'s (2013) French norms. Note: Significant values are shown in boldface type, with positive ones indicating a female bias, and negative ones a male bias.

	Linguistic form		
Stereotype	Masculine	Double-gender	Middot
Male	+7.5	**+19.3**	**+19.2**
Neutral	**-10.7**	-0.8	-1.1
Female	**-20.2**	**-17.9**	**-12.4**

These data suggest that gender-fair language forms can rectify the male-biased representation induced by the masculine form for neutral-stereotyped professions, while they create a female bias for male-stereotyped professions and fail to correct the male bias for female-stereotyped professions.

General Discussion

In two on-line experiments, we investigated the influence of linguistic form and gender stereotype on the presence of women in mental representations of groups of people. French participants read a short text on a professional gathering and estimated the percentage of women present in the gathering. We deliberately opted for a complete between-participants design in which each participant was tested in a single trial, such as to avoid the emergence of response strategies. In each experiment we compared the masculine form – which is ambiguous since its interpretation can be both specific (i.e., referring to men only) and generic (i.e., referring to men and women) – to two unambiguous alternatives for mixed-sex groups, i.e. double-gender and

middot form. In Experiment 1 we tested neutral-stereotyped professions, and in Experiment 2 male- and female-stereotyped ones. In addition, we compared all experimental results to norming data from Misersky et al. (2013). These comparisons allowed us to establish for each stereotype which linguistic forms yield a mental representation in accordance with people's perception of gender ratios in the real world and which ones generate a biased representation. Our results can be summarized as follows.

In both experiments we observed lower estimates of %-women for the masculine form than for the double-gender and middot forms. Experiment 2 additionally revealed an effect of stereotype, with lower estimates for male- compared to female-stereotyped professions. This effect interacted with that of linguistic form only in a post-hoc analysis that left out the data for the middot form, showing higher estimates of %-women for male- but not for female-stereotyped professions when the double-gender form was shown. Compared to the norming data, for neutral-stereotyped professions we found a male bias with the masculine form but consistent estimates with the gender-fair forms. However, for male-stereotyped professions, this comparison showed consistent estimates with the masculine form but a female bias with the gender-fair forms, and for female-stereotyped professions it suggested a male bias with all linguistic forms.

As to the relatively increased representation of women induced by the gender-fair plural forms compared to the masculine plural, our results are in accordance with previous studies on both French (Brauer & Landry, 2008; Gygax & Gabriel, 2008) and German (Braun et al., 1998; Stahlberg & Sczesny; 2001; Stahlberg et al., 2001; Gabriel & Mellenberger, 2004; Horvath et al., 2016; Kollmayer et al., 2018). Both these languages have an innovative neutralizing form alongside the more conventional double-gender, i.e. middot in French and capital-I in German.

For German, there is evidence that capital-I yields an even higher representation of women than the double-gender form (Stahlberg & Sczesny, 2001, Stahlberg et al. 2001), while for French, no previous research has examined the effect of the middot form. Our study, however, suggests that there is no gradient effect of linguistic form, as we found no difference between the estimates for the double-gender and middot conditions in either experiment. Possibly, the diverging results are due to the fact that in German the innovative form is highly similar to the feminine form in writing and indistinct from it in pronunciation (German, e.g.: *LeserInnen* versus *Leserinnen*; French, e.g.: *électricien·ne·s* versus *électriciennes*, but *caissier·ère·s* versus *caissières* and *éboueur·euse·s* versus *éboueuses*). In this respect, we can also recall the case of Swedish, a language that has introduced an innovative, neutral, personal pronoun *hen*, which complements male *han* 'he' and female *hon* 'she'. *Hen* reached widespread adherence in a few years, before its official adoption in 2015 (Gustafsson Sendén et al., 2015), which contrasts sharply with the situation in France, where the use of middot has remained an ideologically divisive issue. Yet, similar to what we found for French, Lindqvist el al. (2019) observed no difference in responses to the neutral *hen* compared to the double-gender form *hen/hon* regarding the presence of women in mental representations. More research, though, is necessary to examine possible differing effects of double-gender and middot forms in French.

There is one caveat to be mentioned concerning the results for the masculine form. Gygax & Gabriel (2008) showed that the interpretation of the masculine plural as specifically referring to men is enhanced when participants have just read short, unrelated, texts containing double-gender forms. (They did not test if using a middot has the same effect, but there is no reason to think it would not.) As participants come to an experiment with all their previous language experience, the co-existence of gender-fair forms and generically intended masculine

forms might have made our participants on average less inclined to embrace the generic interpretation than would have been the case before the rise of gender-fair language. Hence, the estimates of %-women in response to the masculine form might have been lower than what we would have seen one or more decades ago. As long as generically intended masculine forms co-exist with gender-fair forms, this pragmatic backlash effect is expected to similarly be present in people's text interpretations in real-life situations. For future research, it would be interesting to examine whether participants' use of, familiarity with, or even adherence to gender-fair language affects their interpretation of the masculine plural.

The effect of stereotype has been shown many times (Brauer & Landry, 2008; Braun et al., 1998; Gygax et al., 2008, 2012; Gygax & Gabriel, 2008; Horvath et al., 2016; Irmen, 2007; Irmen & Roßberg, 2004). Given that none of the three linguistic forms imply anything about the gender ratio in groups of mixed gender, it is easy to see why stereotype information influences judgments on this ratio. One might expect, though, that male and female stereotypes have differential effects when one of the gender-fair forms is presented. Specifically, the use of a double-gender or middot form might boost the presence of women in mental representations to a larger extent for male-stereotyped than for female-stereotyped professions. We found only limited evidence for this, despite the clear presence of such an interaction – albeit with a small effect size – in our Pilot 3. Might this latter result have been a strike of luck? In the absence of a tool for computing power in linear mixed-effects models with interactions, this is hard to know. We present here the results of a pooled analysis, for which we combined the data from Experiment 2 with that of the pilot. (None of the participants in Experiment 2 had participated in the pilot.) The pooled dataset contains a total of 447 participants aged between 20 and 40 (mean = 29.3, SD=5.6), with on average 71 participants per condition (min = 32, max = 42). The same

regression model as the one for the analysis of Experiment 2 revealed not only effects of Stereotype and Linguistic form, but also an interaction, as shown in Table 4.

Table 4. Results of linear mixed-effects model for pooled data analysis (Experiment 2 and Pilot 3)

	β	*SE*	*T*	χ^2	*Df*	*p*	partial η^2
Stereotype(male)	-14.3	1.69	-8.48	71.9	1	< .0001	0.87
				36.2	2	< .0001	0.08
Form	3.66	1.27	2.90				
double	3.88	1.25	3.12				
middot				10.6	2	.005	0.02
Stereotype × Form	4.08	1.27	3.23				
male:double	-1.54	1.25	-1.24				
male:middot							

The pattern of results of this pooled analysis is identical to that of the pilot analyzed separately: First, the effect of Linguistic form reveals that compared to the masculine form, higher estimates of %-women are obtained for the double-gender form (β = 11.2, SE = 2.19, t(435) = 5.11, p < .0001) and the middot form (β = 11.4, SE = 2.16, t(435) = 5.29, p < .0001), while there is no difference between the latter two (t < 1). Second, restricted analyses of the interaction show that this pattern is most prominent for male-stereotyped professions (masculine vs. double: β = 17.8, SE = 3.05, t(438) = 5.84, p < .0001; masculine vs. middot: β = 12.4, SE = 3.07, t(435) = 4.05, p < .0003; double vs. middot: β = -5.40, SE = 3.05, t(438) = -1.74, p > .1);

for female-stereotyped professions, the differences between the masculine form on the one hand and the gender-fair forms on the other hand, are indeed smaller and significant only for the middot form (masculine vs. double: $\beta = 4.59$, SE = 3.15, $t(432) = 1.46$, $p > .1$; masculine vs. middot: $\beta = 10.4$, SE = 3.04, $t(435) = 3.47$, $p < .002$; double vs. middot: $\beta = 5.84$ SE = 3.06, $t(435) = 1.91$, $p > .1$). Compared to the results of the pilot experiment, the effect size (partial η^2) of the interaction is smaller (0.02 vs. 0.05) but the χ^2 statistic is higher (10.6 vs. 6.53). This pooled data analysis, then, adds evidence supporting the hypothesis that gender-fair language forms increase the presence of women in mental representations especially for male-stereotyped professions. Yet, given its small effect size, we conjecture that a large sample size is needed in order to reliably observe the relevant interaction. Recall that the two previous studies that examined the effect of linguistic form across two stereotypes compared the masculine to a double-gender form only (Braun et al. 1998; Brauer & Landry, 2008). Both adopted in essence the same paradigm as we did, but with just a single profession per stereotype, such that it is unknown whether their findings are generalizable across different professions. For French, Brauer & Landry (2008) observed no interaction; with 73 participants, though, their sample size was probably too small. For German, by contrast, Braun et al. (1998) did report the expected interaction: the double-gender form increased the percentage of estimated women for male- but not for female-stereotyped professional groups. Further research would be welcome to shed more light on the issue of differential effects of various gender-fair language forms depending on the associated gender stereotype. One specific question in this respect concerns a possible difference between the double-gender and middot forms, as suggested by the fact that it is the middot form that prevents a global interaction in Experiment 2. That is, as the middot is still relatively infrequent, its processing might take up resources that would otherwise be allocated to

processing information on stereotype.

Next, we turn to the question as to whether gender-fair language forms induce consistent representations. We have assumed throughout that representations are consistent if they reflect people's perceived gender ratios, as indicated by Miserky et al.'s (2013) norming data, rather than real-world gender ratios (which are currently not available for France). There is some evidence that people can reliably estimate real-world gender distributions. That is, in a study on the distribution of men and women in professions in the UK, Garnham et al. (2015) found a good correlation between the norming data provided by the English sample in Misersky et al. (2013) and real-world data from governmental sources. Yet, as English marks gender only on personal pronouns, it would be useful to run a similar study in French (or German). Indeed, one might wonder to what extent norming data are themselves influenced by linguistic representations, either as encountered in daily life or as used in the norming questionnaire. As to the latter, the endpoints of the rating scale for gender-marked profession names in Misersky et al. (2013) showed the masculine and feminine forms, respectively, and the direction of the rating scale was counterbalanced across participants. The influence of language forms on perceived gender ratios was thus experimentally controlled at best. We tentatively conclude that the choice of benchmark data should not make much of a difference.

Recall that our results showed that the gender-fair forms yield consistent representations only for professions with a neutral stereotype. For male-stereotyped professions they overshoot their objective, while for female-stereotyped ones they fail to provide enough of a boost. It would be interesting to consider professions in a larger range of stereotypicality, as we only tested professions for whom the norming data either showed an almost perfect gender balance (estimated proportions of women between .47 and .51) or a largely unbalanced one (estimated

proportions of women below .30 or above .70). For female-stereotyped professions, one may also wonder if putting the feminine form before the masculine, e.g., *les caissières et caissiers* 'the cashiers$_{fem/masc}$', would render the presence of women in mental representations more salient, thus yielding more consistent representations. In everyday communication this 'feminine before masculine' presentation is standard in a few cases where a mixed-sex group of people is addressed (e.g., *Mesdames et messieurs* 'Ladies and gentlemen', and *Bonjour à toutes et à tous* 'Hello to all$_{fem-masc}$ of you'), but otherwise quite unusual. Previous research suggests that the order of words in a binomial phrase concurs with differential cognitive accessibility and relevance to a context (Kesebir, 2017; Tachihara & Goldberg, 2020). For instance, when asked to name familiar couples, people tended to mention first the person to whom they felt close (thus easier to be brought to mind) (Tachihara & Goldberg, 2020). Kesebir (2017) showed when "women" was mentioned before "men" in a binomial phrase, people were more likely to think of women as member of a group of activists, relative to when "women" was mentioned after "men". Thus, reversing the order might indeed render the presence of women in mental representations more salient. (It's worth noting that the 'masculine before feminine' word order, more than a linguistic convention, maps stereotypical beliefs about the two sexes (Hegarty et al., 2011). Mentioning women before men may therefore also help fight traditional gender stereotype.)

Finally, we briefly turn to the social impact of gender-fair language. Depending on one's view on gender roles, it might be argued that the mismatches we observed between our experimental results and people's perception of gender ratios in the real world – with gender-fair language forms resulting in more balanced gender ratios for male- and females-stereotyped professions than those found in the norming data of Miserksy et al. (2013) – are *desirable*. Specifically, the practice of using gender-fair language could weaken stereotypes (Chatard et al.,

2005; Horvath & Sczesny, 2016; Vervecken et al., 2015; Vervecken & Hannover, 2015; but see Merkel et al., 2012), and hence over time, traditionally male-dominated professions might attract more women and female-dominated ones, more men. In other words, the use of gender-fair language could play a normative role in promoting more balanced real-world gender ratios in the long term. In the short run, there may be some side effects from the backlash against the use of gender-fair language. For instance, female job applicants were evaluated less favorably in a hiring process when introduced with feminine job titles than with masculine ones (Budziszewska et al., 2014; M. Formanowicz et al., 2013) and professions presented with gender-fair language were estimated to earn lower salaries than with masculine forms (Horvath et al., 2016). (The above-mentioned studies were run with Polish, German and Italian participants. A fourth, smaller-scale, study with French participants did not find any influence of gender-fair language on the evaluation of professions (Gygax & Gesto, 2007). As observed by Gustafsson Sendén et al., (2015), people's attitudes toward gender-fair language become positive over time, thus any backlash effects might diminish with more exposure to the presence of gender-fair language forms.

To conclude, we showed that the generic use of the masculine plural and gender-fair alternatives differentially impact how people mentally represent and estimate gender ratios. In addition to adding important data to fuel public debate around gender-fair language, our results also potentially lead to new questions. For example, how do people prioritize consistency of mental representations vs. gender fairness when these come apart in their views about which linguistic forms are most desirable? And how do proponents of gender-fair language weigh the obvious orthographic drawbacks of the middot form against its potential advantage in terms of representational consistency compared to double-gender forms? Whatever the answers to these

questions end up being, it is clear that existing and future empirical data should be a driving force in informing the trade-offs that must be considered in arriving at coherent policy decisions.

Pilot Studies

Pilot 1

This pilot (not included in the submitted manuscript) is similar to the two experiments described in the chapter with respect to its design and procedure. It has three aims. First, we seek to validate our paradigm by focusing on the French feminine grammatical form exclusively, which unambiguously refers to women-only groups. Thus, if the grammatical information is processed correctly, we should observe overall high estimates of %-women. In order to minimize the chances that the profession name is erroneously processed as if it were shown in the masculine form, we use names for which the feminine plural is not only written but also pronounced differently than the masculine one.

Second, we investigate whether the influence of gender stereotype can override that of grammatical gender. To do this, we compare male- vs. female-stereotyped professions. For female-stereotyped professions the stereotype is in accordance with the grammatical gender, but for male-stereotyped professions the two types of information conflict. Thus, if stereotype information overrides linguistic information, we should observe lower estimates of %-women for male-stereotyped than for female-stereotyped professions, where the latter should be at or near ceiling, i.e. 100%.

Third, we examine a potential difference according to whether the linguistic information is present on a noun or on a pronoun. Compared to content words, function words – including pronouns – are shorter, more frequent, more predictable and more often redundant, and there is evidence that they are processed in less depth during reading (Carpenter & Just, 1983; Healy,

1994; Saint-Aubin & Poirier, 1997; Staub et al., 2019). Thus, one might expect for grammatical manipulations on pronouns to have less of an impact than similar manipulations on profession names. In order to test this, we exploit the fact that not all French profession names have different forms for male and female gender. Specifically, we contrast two minimally different scenarios: In one, the profession name has gender marking (e.g., *caissier* 'cashier$_{masc}$' – *caissière* 'cashier$_{fem}$', used in the examples above) and there is no referential pronoun. In the other one, the profession name has no gender marking (e.g. *artiste* 'artist$_{masc/fem}$') but the grammatical gender information is present on a referential pronoun, i.e. *elles* 'she$_{pl}$'. Moreover, we examine more directly to what extent the grammatical gender information is processed, by adding a multiple-choice question to test participants' recall of the crucial word after they have provided their estimate. We expect that if there is a difference, recall will be better when the crucial word is the profession name than when it is the pronoun.

Method

Stimuli

We selected 24 professions, twelve female-stereotyped and twelve male-stereotyped, from the French part of the same norming study (Misersky et al., 2013) as mentioned before. The 12 professions for the noun condition that have varying forms for the masculine and feminine gender (i.e. gender-marked) were the same items as used for Experiment 2. The other 12 professions have a unique form (i.e. not gender-marked) (e.g. male-stereotyped: *bagagiste* 'porter$_{masc/fem}$'; female-stereotyped: *fleuriste* 'florist$_{masc/fem}$') and are used in the pronoun condition. All professions were selected from among those with mean estimated proportions of men or of women, respectively, above .70 (male-stereotyped, gender-marked: M_{men} = .81 SD

= .03; male-stereotyped, not gender-marked: M_{men} = .80 SD = .02; female-stereotyped, gender-marked: M_{women} = .78, SD = .05; female-stereotyped, not gender-marked: M_{women} = .78, SD = .08). An Anova with the factors Stereotype and Gender-marking showed no main effect and no interaction (Stereotype: $F(1,20) = 1.07$, $p > .1$; Gender-marking and Stereotype × Gender-marking: $F < 1$). The 24 professions, together with the estimates of %-women in those professions – according to Misersky et al.'s norming study – are shown in the Appendices A and B.

The short text about the professional gathering was similar to the one used in the two experiments. Same as the previous experiments, in the passage for the noun condition, exemplified in (12a), the feminine plural form of the profession name appeared twice in the text and no referential pronoun was used. In the passage for the pronoun condition, exemplified in (12b), the profession name had no gender marking and its plural form was shown only once, but a referential feminine plural pronoun appeared twice; this pronoun thus revealed the gender information. Hence, in both conditions the relevant grammatical information appeared twice.

(12) a. Passage for noun condition with sample profession:

Le rassemblement régional des caissières a eu lieu cette semaine à Amiens. La localisation centrale de cette ville a été particulièrement appréciée. Les caissières ont aussi adoré l'apéro offert à l'hôtel de ville le premier jour.

'The regional gathering of cashiers$_{fem}$ took place this week in Amiens. The central location of this city was particularly appreciated. The cashiers$_{fem}$ also loved the aperitif offered at the City Hall on the first day.'

b. Passage for pronoun condition with sample profession:

Le rassemblement régional des fleuristes a eu lieu cette semaine à Amiens. Elles ont particulièrement appréciée la localisation centrale de cette ville. De plus, elles ont adoré l'apéro offert à l'hôtel de ville le premier jour.

'The regional gathering of florists$_{masc/fem}$ took place this week in Amiens. They$_{feminine}$ particularly appreciated the central location of this city. They$_{feminine}$ also loved the aperitif offered at the City Hall on the first day.'

The words *régional* and *Amiens* were replaced with *européen* and *Francfort* respectively for three professions, *mathématicien* 'mathematician', *douanier* 'customs officer' and *astronaute* 'astronaut', for it would be more plausible for these professionals to have a Europe-wide assembly in a more internationally-oriented city than a regional one in a provincial city.

Procedure

The pilot followed the same procedure as for Experiment 1 and 2 except in the aspects described below.

Participants were randomly assigned to one of four groups defined by crossing word type (noun or pronoun) and stereotype (male or female). Within each group, participants were randomly assigned to one profession from the relevant list.

After the rating task, participants were asked a multiple-choice question to test how well they recalled the gender-marked profession name or the pronoun from the text they had seen. The question was framed as *Quel mot était présent dans le texte ?* ('Which word was shown in the text?') and three choices were shown below the question. For participants in the noun condition, the three choices were: profession name in masculine plural form (e.g., *caissiers*), profession name in feminine plural form (e.g., *caissières*), and *Je ne sais pas* ('I don't know');

for those in the pronoun condition, the three choices were: masculine plural pronoun (*ils*), feminine plural pronoun (*elles*) and *Je ne sais pas*.

Participants

The analysis included 201 (134 women and 67 men) native French speakers living in France, aged between 20 and 40 years (M = 28.2, SD = 6.2), who participated on the crowd-sourcing platform Clickworker. Their random assignment to one of eight conditions (two gender stereotypes, two word types, and two slider layouts) yielded a mean number of participants per condition of 25 (min = 24, max = 26).

The data from an additional 41 participants were removed from analysis for the following reasons: one did not complete the survey, five took the survey for the second time, 17 responded incorrectly at one or both attention check questions, and the remaining ones did not satisfy all of our recruitment criteria (eight were non-native speakers, one did not live in France, and four were younger than 20, and five older than 40).

Results and discussion

Boxplots of the estimated percentages of women as a function of word type and gender stereotype are shown in Figure 5.

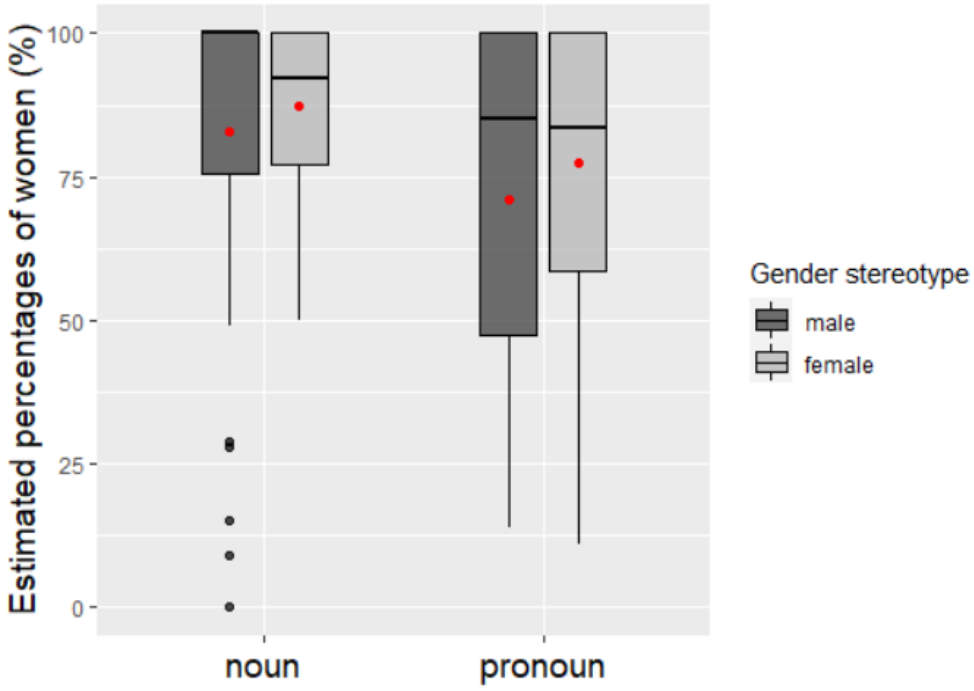


Figure 5. Boxplots of estimated percentages of women as a function of word type and gender stereotype. Medians are indicated by black lines, means by red dots.

These data were fit with a linear mixed-effects model by using the *lme4* package (Bates et al., 2015) in the programming software *R* (R Core Team, 2020) and *Rstudio* (RStudio Team, 2020) Statistical significance was assessed by means of the Anova function in the *Car* package (Fox & Weisberg, 2019). Stereotype (male vs. female) and Word type (nouns vs. pronoun) were contrast-coded and set as fixed effects together with their interaction term. A random intercept was added for Profession. As we collected only one datapoint per person, we did not include a random factor for Participant. Note that this model differs from the one we preregistered in that it does not contain fixed factors for Slider version and its interaction terms. As neither in this nor in the following experiments this counterbalancing factor or any of its interactions affected the

estimated %-women, and as omitting these terms did not change any of the results, we report the simpler models throughout the article for the reader's convenience.

The results, shown in Table 5, revealed an effect of Word type, with higher estimates of %-women in the noun compared to the pronoun condition. By contrast, there was no effect of Stereotype nor an interaction.

Table 5. Results of linear mixed-effects regression

	β	*SE*	*t*	χ^2	*Df*	*p*	*partial* η^2
Stereotype(male)	-2.71	1.92	-1.41	1.98	1	0.16	0.10
Word type(noun)	5.36	1.92	2.79	7.77	1	< .01	0.29
Stereotype(male):**Word type**(noun)	0.42	1.92	0.22	0.05	1	0.83	0

Next, we consider the responses to the memory question. Five participants, all in the pronoun condition, could not choose between the male- or female-inflected forms of the crucial word: they replied that they were unsure. Their responses were coded as incorrect. The mean percentages of gender recall accuracy as a function of word type and stereotype are shown in Figure 6.

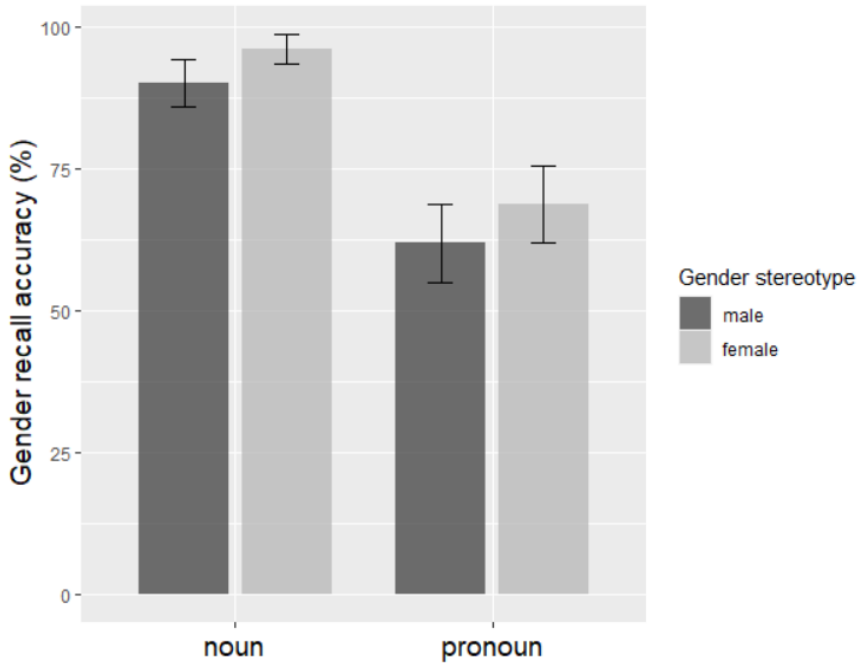


Figure 6. Mean percentages of gender recall accuracy as a function of word type and stereotype.

We fit the data with a logistic mixed-effects model with contrast-coded fixed factors Stereotype, Word type, and their interaction, and with a random intercept for Profession. The results, shown in Table 6, revealed an effect of Word type, as nouns were more likely to be correctly remembered than pronouns. By contrast, there was no effect of Stereotype nor was there an interaction.

Table 6. Results of logistic mixed-effects regression

	β	*SE*	*Z*	χ^2	*Df*	*p*
Stereotype(male)	-0.33	.26	-1.27	1.23	1	0.27
Word type(noun)	1.07	0.27	4.01	15.9	1	< .0001
Stereotype(male) × **Word type**(noun)	-0.16	0.26	-0.61	0.38	1	0.54

These results show that when grammatical information is unambiguous, i.e. indicating that a group of professionals consists entirely of women, participants' estimates of the %-women in the group are not influenced by the profession's stereotype. Thus, in this case gender stereotype does not override grammatical gender information, regardless of whether the gender marker appears on the noun or on a referential pronoun. The estimates, however, were not at ceiling (as in theory they should be), and, more importantly, they were lower when the grammatical information was present on a referential pronoun (mean: 74%) than when it was present on the noun (mean: 85%). Furthermore, gender recall of the crucial word was worse for pronouns (mean accuracy: 65%) than for nouns (mean accuracy: 93%), suggesting better processing of nouns than of pronouns, in accordance with previous research on differences between content and function words in reading (Healy, 1994; Saint-Aubin & Poirier, 1997; Schindler, 1981; Staub et al., 2019).

Although with a single datapoint per participant we did not have sufficient statistical power to directly assess the effect of accuracy in the memory question on the estimate of %-women, the data do suggest that incorrect responses to the memory question are associated with lower estimates of %-women, especially for male-stereotyped professions. This can be seen in Table 7, which shows the number of participants and mean estimates of %-women as a function of Word type, Stereotype, and Recall response.

Table 7. Number of participants and mean estimates of %-women as a function of Word type, Stereotype, and Recall response

Word type	Stereotype	Recall response	Nb	Mean %-women	SE
Noun	Male	correct	46	89.5	2.63
		incorrect	5	20.4	8.56
	Female	correct	50	88.9	1.75
		incorrect	2	50.5	0.50
Pronoun	Male	correct	31	89.7	4.08
		incorrect	19	40.9	3.68
	Female	correct	33	86.9	3.62
		incorrect	15	56.6	5.63

Note that on average, estimates of %-women by participants who were correct on the memory question still did not exceed 90%. Various factors might be involved in this lack of ceiling performance: Some of these participants may have taken into consideration persons who were not denoted by the profession names but were present at the gathering (e.g. organizers, hotel staff, or accompanying family members, all of whom might be male). For others, their answer to the memory question might have been a guess which happened to be correct. Or, some might have misinterpreted the response slider when providing their estimate (six of them indeed estimated the % of women to be lower than 50%, ranging from 11% – for a female-stereotyped profession! – to 43%, with a mean of 27.9%) Last but not least, some might not have paid enough attention to the task.

These factors could of course also be at play for the participants who were incorrect on the memory question. Despite the presence of such noise in the data, our paradigm appears to be adequate for testing the influence of grammatical gender on the perception of biological gender, especially when the gender information is present on the noun denoting the profession rather than on a referential pronoun. In the remaining experiments, we therefore focus on grammatical gender-marked profession names, and hence use the kind of noun passage exemplified in (5a) above.

Pilot 2

Pilot 2 (not included in the submitted manuscript) was similar to Pilot 1 in terms of the materials and procedure. In this pilot experiment, we compared the masculine plural pronoun *ils* 'they$_{masc}$' and its double-gender alternative *ils et elles* 'they$_{masc}$ and they$_{fem}$'. We did not include the inclusive writing form *iels* as it seems extremely unfamiliar to French speakers.

Method

Stimuli

The stimuli used in this pilot were the same as for the pronoun condition of Pilot 1. They were 12 profession names with a unique form for the masculine and feminine gender (see Appendix B).

Procedure

The procedure for this pilot was identical to that of Pilot 1 except in the following aspects. First, we only had a pronoun condition as we removed the noun condition described in Pilot 1. Second, no memory question was being asked after the estimation task. Third, participants were randomly assigned to one of four groups (2 gender stereotype x 2 linguistic form), and within each group, they were randomly shown one profession name from the relevant list.

Participants

Participants were native French speakers living in France (N = 146), 80 women and 66 men, aged between 20 and 68 years (M = 43, SD = 12). They participated on the crowd-sourcing platform Foulefactory.

The mean number of participants per condition was 18 (min = 8, max = 25).

Results

Boxplots of estimated percentages of women as a function of linguistic form and stereotype are shown in Figure 7.

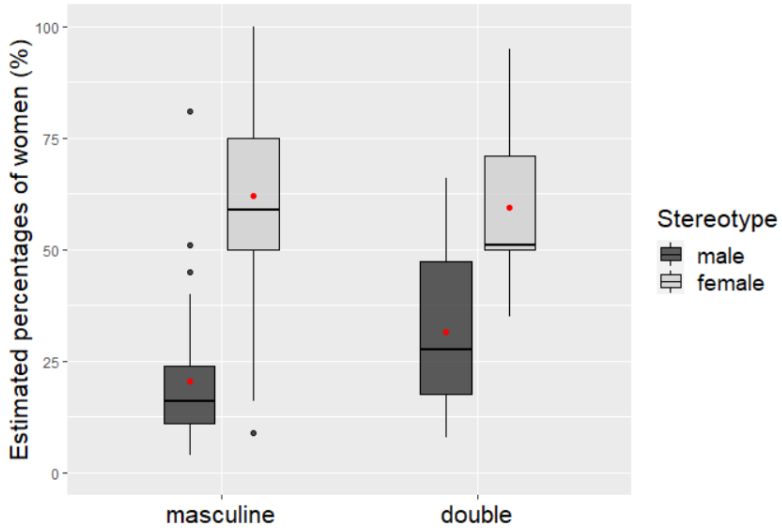


Figure 7. Boxplots of estimated percentages of women as a function of linguistic form and stereotype. Medians are indicated by black lines, means by red dots.

As in Pilot 1, the data were fit with a linear mixed-effects model. The model contained fixed factors for contrast-coded Stereotype, Linguistic form, and their interaction, as well as a random intercept for Profession.

Table 8. Results of linear regression for Pilot 3

	β	*SE*	*t*	χ^2	*Df*	*p*	*partial* η^2
Stereotype(male)	-17.0	2.56	-6.64	44.1	1	< .0001	0.81
Form(double)	2.56	1.35	1.89	3.58	1	0.058	0.03
Stereotype(male) × **Form**(double)	2.96	1.35	2.18	4.77	1	0.03	0.03

The results, shown in Table 8, revealed effects of Stereotype and an interaction. However, only a marginal effect of linguistic form was found.

As found in previous experiments, lower percentages of women were estimated for male-stereotyped professions (M = 26.5, SD = 17.3) than for female-stereotyped ones (M = 60.8, SD = 17.5). Conversely, double-gender pronoun (M = 47.3, SD = 21.3) only slightly increased the perceived proportion of women compared to the masculine form (M = 46.1, SD = 27.0). The influence of linguistic form, however, was dependent on stereotype. Restricted analyses with corrections for multiple comparison (mvt method) showed that for male-stereotyped professions, double-gender form increased the representation of women [$\beta = 11.0$, SE = 4.13, $t(136) = 2.67$, $p < .01$], while for female-stereotyped ones, no difference was found between the two linguistic forms ($t < 1$, $p = 0.82$).

Pilot 3

This pilot was included in the submitted manuscript. It is identical to Experiment 2 in the aspects of design, materials, and procedure.

Method

Stimuli

The stimuli were identical to those for Experiment 2.

Procedure

The procedure remained identical to that of Experiment 2 with one exception: the number of datapoints per profession within each group defined by stereotype, linguistic form, and slider direction was more variable.

Participants

Participants were native French speakers living in France (N = 142), 73 women and 69 men, aged between 21 and 40 years (M = 32, SD = 5.2). The mean number of participants per condition was 11 (min = 6, max = 17).

Results

Boxplots of estimated percentages of women as a function of linguistic form and stereotype are shown in Figure 8.

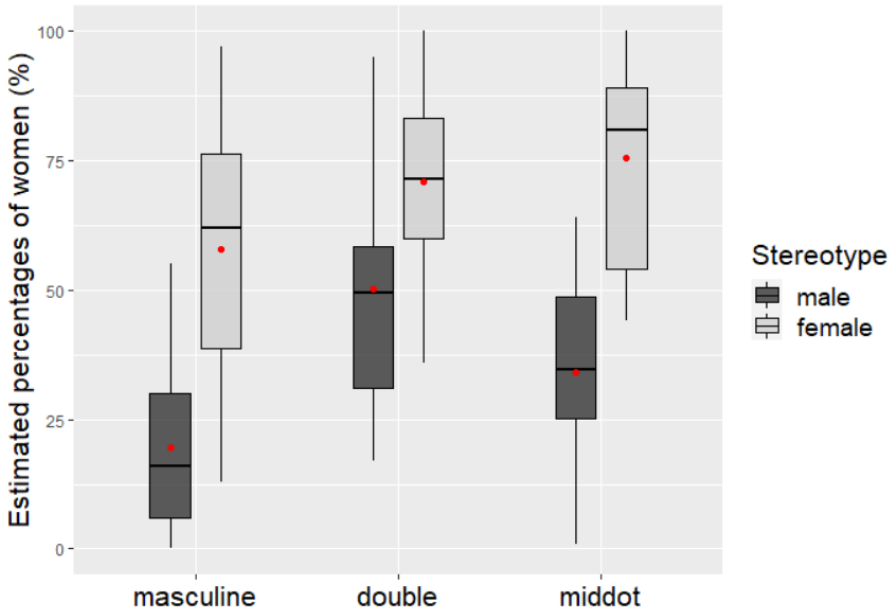


Figure 8. Boxplots of estimated percentages of women as a function of linguistic form and stereotype. Medians are indicated by black lines, means by red dots.

As in Experiment 2, the data were fit with a linear mixed-effects model. The model contained fixed factors for contrast-coded Stereotype, Linguistic form, and their interaction, as well as a random intercept for Profession. The results, shown in Table 9 revealed effects of Stereotype, Linguistic form, and an interaction.

Table 9. Results of linear mixed-effects model for the Pilot study

	β	*SE*	*t*	χ^2	*Df*	*p*	*partial* η^2
Stereotype(male)	-17.0	1.92	-8.88	78.9	1	< .0001	0.91
				27.9	2	< .0001	0.17
Form	9.00	2.40	3.75				
double	3.07	2.25	1.36				
middot				6.53	2	0.04	0.05
Stereotype × Form	6.06	2.40	2.53				
male:double	-3.58	2.25	-1.59				
male:middot							

Compared to female-stereotyped professions, lower percentage of women was estimated for male-stereotyped ones.

As to the main effect of linguistic form, the pattern of results is consistent with that of Experiment 2 and the pooled analysis described above. Compared to the masculine form, higher estimates of %-women were obtained for the double-gender form (β = 21.1, SE = 4.23, t(136) = 4.98, p < .0001) and the middot form (β = 15.1, SE = 4.0, t(131) = 3.78, p < .001). However, there was no significant difference between the two gender-fair forms (β = -5.94, SE = -4.10, t(135) = -1.45, p > .1).

Concerning the interaction, restricted analyses showed that this pattern is most prominent for male-stereotyped professions (masculine vs. double: β = 29.6, SE = 5.57, t(122) = 5.32, p < .0001; masculine vs. middot: β = 14.0, SE = 5.78, t(127) = 2.43, p < .05; double vs. middot: β = -15.6, SE = -5.69, t(132) = -2.74, p < .02); while for female-stereotyped professions, the

differences between masculine and the gender-fair forms are smaller and significant only for the middot form (masculine vs. double: $\beta = 12.5$, SE = 6.38, $t(133) = 1.97$, $p > .1$; masculine vs. middot: $\beta = 16.2$, SE = 5.54, $t(134) = 2.93$, $p < .02$; double vs. middot: $t < 1$).

Appendices

Appendix A. French profession names used in Experiments 1, 2, and Pilots 1 (noun condition) and 3, with mean percentages of women as rated by participants in Misersky et al. (2013).

Stereotype	**Profession name**			**English translation**	**%-women**
	Masculine	Double-gender	Middot		
Neutral	Musiciens	Musiciens et musiciennes	Musicien.ne.s	Musicians	47%
(Exp. 1)	Greffiers	Greffiers et greffières	Greffier.ère.s	Law clerks	49%
	Acupuncteurs	Acupuncteurs et acupunctrices	Acupuncteur.trice.s	Acupuncturists	49%
	Employés de banque	Employés et employées de banque	Employé.e.s de banque	Bank clerks	49%
	Bijoutiers	Bijoutiers et bijoutières	Bijoutier.ère.s	Jewelers	51%
	Comédiens	Comédiens et comédiennes	Comédien.ne.s	Comedians	50%
Male	Éboueurs	Éboueurs et éboueuses	Éboueur.euse.s	Rubbish collectors	14%
(Exp. 2,	Charpentiers	Charpentiers et charpentières	Charpentier.ère.s	Carpenters	18%
Pilot 1)	Électriciens	Électriciens et électriciennes	Électricien.ne.s	Electricians	20%
	Mécaniciens	Mécaniciens et mécaniciennes	Mécanicien.ne.s	Mechanics	20%
	Douaniers	Douaniers et douanières	Douanier.ère.s	Customs officers	22%
	Mathématiciens	Mathématiciens et mathématiciens	Mathématicien.ne.s	Mathematicians	22%
Female	Couturiers	Couturiers et couturières	Couturier.ère.s	Dressmakers	73%
(Exp. 2,	Caissiers	Caissiers et caissières	Caissier.ère.s	Cashiers	75%
Pilot 1)	Diéteticiens	Diéteticiens et diététiciennes	Diéteticien.ne.s	Dieticians	75%
	Maquilleurs	Maquilleurs et maquilleuses	Maquilleur.euse.s	Make-up artists	77%
	Assistants maternels	Assistants maternels et assistantes maternelles	Assistant.e.s maternel.le.s	Nursery assistants	82%
	Esthéticiens	Esthéticiens et esthéticiennes	Esthéticien.ne.s	Beauticians	86%

Appendix B. French profession names used in Pilots 1 (pronoun condition) and 2, with mean percentages of women as rated by participants in Misersky et al. (2013).

Stereotype	Profession name	English translation	%-women
Male	Astronautes	Astronauts	18%
	Bagagistes	Porters	19%
	Garde-forestiers	Forest rangers	20%
	Métallurgistes	Metallurgists	21%
	Haltérophiles	Weight lifters	21%
	Croque-morts	Undertakers	23%
Female	Mannequins	Fashion models	68%
	Nutritionnistes	Nutritionists	74%
	Réceptionnistes	Receptionists	74%
	Fleuristes	Florists	81%
	Secrétaires	Secretaries	82%
	Manucures	Manucurists	90%

Appendix C. Warm-up text and questions in the experiments and pilots

Text

La visite du couple royal et de leurs trois enfants a été préparé avec beaucoup de soin. La mairie a notamment commandé un grand nombre de fleurs jaunes pour décorer la place du village.

'The visit of the royal couple and their three kids were carefully prepared. The City Hall especially ordered a great number of yellow flowers to decorate the village square'

Question 1

Combien d'enfants le couple royal a-t-il ?

'How many kids does the royal couple have?'

Options: [1], [2], [3], [4]

Question 2

De quelle couleur étaient les fleurs sur la place du village ?

'In what color were the flowers at the village square?'

Options: [Blanc] 'white', [Rouge] 'red', [Violet] 'purple', [Jaune] 'yellow'

Appendix D. Women-men version of the response slider and instructions for use. (Note: instructions for the men-women version were identical except that the words for women and men were switched.)

Déplacez l'indicateur pour répondre :

Si l'indicateur est complètement à gauche, vous estimez qu'il y a 100% de femmes et 0% d'hommes.

Si l'indicateur est complètement à droite, vous estimez qu'il y 0% de femmes et 100% d'hommes.

Si ll'indicateur est au milieu, vous estimez qu'il y a autant de femmes que d'hommes.

Bien sûr, votre choix n'est pas restreint à ces trois positions ; vous pouvez mettre l'indicateur à n'importe quelle position intermédiaire.

‘Move the indicator to respond:

If the indicator is at the utmost left end, you estimate that there is 100% women and 0% men.

If the indicator is at the utmost right end, you estimate that there is 0% women and 100% men.

If the indicator is at the midpoint, you estimate that there are women as many as men.

Of course, your choice is not restricted to the three positions; you can place the indicator whichever position.’

Chapter 4: The Role of Morality in The Evaluations of Research on Gender Bias

Hualin Xiao[1,2,3], Antoine Marie[5] & Brent Strickland[2,3,4]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS–PSL, EHESS, CNRS), Paris, France

[2] Institut Jean Nicod (ENS–PSL, EHESS, CNRS), Paris, France

[3] School of Collective Intelligence, UM6P, Rabat, Morocco

[4]Africa Business School and School of Collective Intelligence, UM6P, Rabat, Morocco

[5] Department of Political Science, Aarhus University, Denmark

This section presents the manuscript "The Role of Morality in The Evaluations of Research on Gender Bias" (work in progress).

Abstract

In a context of increasing gender equality in Western societies, systematic differences and possible biases in how people weight evidence of gender discrimination against women have garnered much public attention. However, the ultimate sources of those are unclear. Some previous work (Handley et al., 2015) suggests that participant's gender is a key factor, such that men rate less favorably than women strong (experimental) scientific evidence suggesting discriminatory hiring practices against women. Here, we explore a potentially more powerful source of variation in how people evaluate evidence of gender discrimination: their level of moral commitment to gender equality.

Across a series of six experiments, we focus on perceptions of discriminatory hiring practices based on gender in academic contexts. We find that people's degree of moral commitment to gender equality is a robust predictor of their trust in statistical evidence of gender discrimination against women, and that the correlation between moral commitment and evaluations cannot be explained by factual prior beliefs. Moreover, this holds whether the evidence of hiring bias is strong (experimental) or weak (only correlational, thus leading to confusion between gender imbalance and discriminatory hiring). Our results additionally show, in contrast to previous work in this area, that participant sex does not predict evaluations of evidence of discrimination. Taken together, our findings suggest a new picture of the origins of systematic differences in people's appreciation of evidence of gender bias in academia.

Keywords: Gender bias, Moral commitment, STEM, Prior beliefs, Trust, Research evaluations

Introduction

The fields of Science, Technology, Engineering, and Mathematics (STEM) are indisputably male-dominated (Shen, 2013). Despite the implementation of recent compensatory measures in STEM, women are still outnumbered by their male colleagues, especially in senior positions (James et al., 2019). The persistent underrepresentation of women has been attributed to multiple interacting factors, ranging from external barriers such as discriminatory hiring practices against women, to internal factors including gender differences in math abilities, lifestyle, and career choices (see Ceci et al., 2009; Charlesworth & Banaji, 2019). In particular, previous research has provided evidence of sex-based discrimination in academia, but it seems that the directionality of the bias varies depending on the context. For example, Moss-Racusin et al. (2012) asked professors at research-intensive universities to evaluate the CV of an undergraduate student for a lab manager position. All professors received the exact same CV while the experimenters manipulated the sex of the applicant by assigning them either a male or female name. Results revealed a significant bias in favor of the male student, who was rated as more competent and hirable and was offered more mentoring and a higher starting salary than the female applicant. Similarly, a wealth of other studies found biased hiring processes in favor of males over female candidates, despite their comparable backgrounds (Begeny et al., 2020; Moss-Racusin et al., 2012; Régner et al., 2019; Reuben et al., 2014). Conversely, some other studies observed a gender bias favoring females (Breda & Hillion, 2016; Breda & Ly, 2015; Williams & Ceci, 2015). For instance, employing a similar CV assessment method, Williams & Ceci (2015) found that women were actually preferred over men for a tenure-track position across both math-intensive and non-math-intensive fields.

While a wealth of evidence suggests that women do suffer sex-based discrimination in academia, there is significant variability in how people react to this fact (Danbold & Huo, 2017; Handley et al., 2015; Moss-Racusin et al., 2015). For instance, Handley et al. (2015) asked participants to rate the quality of the research described in Moss-Racusin et al's (2012) article as mentioned above. The authors found that males rated the research less favorably than female participants, and that this gender gap in evaluations was more pronounced among STEM faculty than in the general public. In a similar vein, Moss-Racusin et al. (2015) conducted a content analysis of the comments written by members of the public in response to press articles reporting on evidence of gender bias as demonstrated in Moss-Racusin et al. (2012); again, they found more positive reactions (e.g., calls for social change) among females, and more negative reactions (e.g. justifications of gender bias) among males. The authors interpreted this gender difference in reactions as in-group favoritism, a mechanism by which men protect their identity as the dominant gender group in STEM by resisting information that threatens this identity (Danbold & Huo, 2017; Handley et al., 2015).

In this work, we approach people's differing reactions to evidence of gender bias from a novel angle: by examining the possible influence of moral commitment to gender equality. Moral commitment (or conviction) to a cause refers to the degree to which individuals deem the issue as an unnegotiable moral imperative. People highly committed to an issue typically see it as objectively and universally important, and as central to their sense of moral identity. They display strong emotional reactions—such as anger and disgust—at people, practices, representations, and institutions that they think to impede the advancement of their cherished cause (Skitka, 2010; Skitka et al., 2005). Drawing on these considerations, we hypothesized that individuals' *moral commitment to gender equality* may be a key moderator of their trust in

research on the issue of gender bias in academia. A long tradition of research has shown that people's moral attitudes modulate their processing of new information—whether this springs from prior beliefs only, or involves emotional processes as in "motivated thinking" (Kunda, 1987, 1990; Lord et al., 1979). For example, scientific articles highlighting the role of human activities in global warming are more likely to have the credibility of their methods or the probity of their authors questioned by conservatives than liberals (Kahan et al., 2011a). Proponents of the death penalty put studies suggesting its inefficiency in deterring murder under closer scrutiny than people who forcefully oppose it (Edwards & Smith, 1996; Lord et al., 1979). Scientists tend to discount manuscripts under review that run against their favored theory than those supporting their theoretical convictions (Koehler 1997).

Current study

Following these lines of research, we expected that individual differences in moral commitment to gender equality may be associated with differences in how persuasive evidence of gender discrimination in academia is perceived as being. Specifically, we hypothesized that greater commitment to gender equality should predict increased trust in research reporting evidence of hiring discrimination against females, regardless of participants' sex. Experiments 1a, 1b, 1c, and 2 look specifically at this question.

Influences of moral attitudes on judgments raise the important question of whether those influences can be reduced to people's prior factual beliefs/expectations on the issue at hand (Lord et al., 1979; Pennycook, 2020; Tappin et al., 2020a, 2020b). In order to shed light on this question, we also explored in Experiment 3 if the (theorized) relationship between individuals' moral commitment to gender equality and their trust in research reporting evidence of hiring

discrimination against females can be explained by their prior expectations about the degree to which hiring practices in academia are biased against women.

Finally, in Experiment 4 we examined one particular way in which moral commitment to gender equality can be expected to influence perceptions of scientific research on the issue: by making participants more likely to make an inadequate inference when the conclusion is pro-attitudinal. When "people believe a conclusion is true, they are also very likely to believe the arguments that appear to support it, even when these arguments are unsound" (Kahneman, 2011)—a fallacious form of reasoning called the "belief bias" (Pennycook, 2020). Specifically, we tested whether the increased moral commitment to gender equality would be associated with higher chances of viewing merely observational data that women are fewer than men in academia as proof that women are discriminated against because of their sex.

Experiment 1a

Using a research evaluation method adapted from Handley et al. (2015), Experiment 1a was designed to answer the question of whether individuals' moral commitment to gender equality affects their trust in evidence related to gender bias. Specifically, we tested whether individuals higher on moral commitment to gender equity would perceive research findings showing hiring discrimination against women as more accurate and the methods employed as more reliable, regardless of whether they were male or female. Moreover, we examined whether any hypothetical sex difference in the reception of evidence as reported in Handley et al. (2015), would disappear after the variance explained by moral commitment was factored out.

All aspects of the experiment including the materials, procedure, and statistical analyses were pre-registered on OSF (10.17605/OSF.IO/UGPE4).

Method

Participants

The data analysis included 268 UK participants (171 women and 97 men) aged between 18 and 74 years (M = 35, SD = 13.1). They were recruited on the crowd-sourcing platform Prolific (https://prolific.co/) and were paid £0.50 for their participation. We determined the sample size by running power analysis using the applications "mc_power_med" and "schoam4" (Schoemann et al., 2017) in R (R Core Team, 2020) via Rstudio (RStudio Team, 2020). The applications suggested a sample size of 250 for a statistical power of 0.87. As data were collected in batches and we applied the exclusion criteria after each batch was complete, we ended up with a sample size slightly larger than the suggested one.

In total, we recruited 288 participants. Data from 20 participants were removed for the following reasons: one did not meet our recruitment criterion (reported sex being “other”), 17 failed one of the two attention checks, and two took the survey more than once (only their first response was kept).

Materials and procedure

We used a summary of the scientific research article by Moss-Racusin et al., (2012) as stimulus. As described above, Moss-Racusin et al., (2012) reported evidence of gender bias against female job applicants (favoring males) for a lab manager position at research intensive universities. We exposed participants to a structured summary of the article divided into several short sections: the research background, a description of the study’s methods, main results (in both text and graphic formats), and a short conclusion. The summary also mentioned the original study’s title, authors, and the name of the scientific journal (*PNAS*) where the article was published in order to make it clear to participants that it was a real scientific study. Our method

was adapted from that used by Handley et al. who however exposed participants to the original abstract of Moss-Racusin et al., (2012). In addition, we employed different dependent measures: the perceived accuracy of research findings and the perceived reliability of research methods (see Electronic Supplementary Materials for exact stimulus).

Participants first gave their consent to participate in the study. Then they were presented with a language check question ("To make sure you're not a bot and that you speak English fluently, please describe in two sentences the current Covid-19 crisis"). Those who provided nonverbal or incomprehensible answers to this question were removed from data analyses.

On the following page, participants saw the research summary and responded to two questions immediately following the summary. The first question read: "To what extent do you think these research findings are likely to be accurate?". Participants responded on a slider scale ranging from [0] "Extremely unlikely" to [10] "Extremely likely", with [5] "Neither likely nor unlikely" being the default slider position. The second question asked: "To what extent do you find the research methods reliable?". The scale ranged from [0] "Not at all reliable" to [10] "Very much reliable", with [5] "Uncertain" as the default slider position. The two questions were presented in random order.

On the next page, participants passed a second attention check by answering the question: "The summary you've just read is about". They could choose from "Poverty economics", "Modern literature", "Gender bias", and "Quantum physics". Participants who chose answers other than "Gender bias" were removed from the analyses.

Then, participants were asked to report their degree of moral commitment to gender equality by indicating the extent to which they agreed with the statement: "Achieving gender

equality in society is a moral imperative" on a slider scale ranging from [0] "Strongly disagree" to [10] "Strongly agree" with [5] "Neither agree nor disagree" as default slider position.

Next, they reported their sex and gender identity in random order. The question for biological sex was "Assuming sex is a biological notion that refers to your physical anatomy, what is your sex?". Participants chose from "Male", "Female" and "Other" as responses. There were two versions of the question on gender identity in order to counterbalance the order in which the words "masculine" and "feminine" appeared in the text. One version of the question was framed as "Assuming gender refers to how masculine or feminine you feel with respect to your identity, where would you position your gender?" (Masculine-Feminine version), and the other version, as "Assuming gender refers to how feminine or masculine you feel with respect to your identity, where would you position your gender?" (Feminine-Masculine version). Consistent with the question framing, there were two versions of response scales depending on whether the left [0] and right [10] endpoints were labeled as "Strongly masculine" and or "Strongly feminine", respectively (Masculine-Feminine version), or the reverse (Feminine-Masculine version). The slider was initially positioned at [5] "Gender neutral" on both scales. Participants were randomly assigned either version of the gender identity question / scale. Responses on the Feminine-Masculine scale were reverse coded.

The survey ended with a question about participants' age, and a 1-item measure of political orientation.

Results

All data cleaning and analyses reported in this paper were conducted in R (R Core Team, 2020) via RStudio (RStudio Team, 2020). Linear regression analyses were run using the *lme4* package (Bates et al., 2015). The categorical variable (i.e. sex) was contrast-coded and the

continuous variables (i.e. accuracy and reliability ratings, moral commitment and gender identity) were centered by applying the *scale* function of R.

We first ran two simple linear regressions with sex as the sole predictor in the models. Results showed no sex difference in the ratings of accuracy of the findings, despite numerically lower ratings of research accuracy provided by male participants (M = 6.40, SD = 2.14) than by female participants (M = 6.87, SD = 2.16; t = -1.70, p = 0.09). No sex difference was observed for the perceived reliability of the reported methods (men: M = 6.45, SD = 2.03; women: M = 6.42, SD = 1.96; t <1).

Then, we ran two linear regressions with moral commitment, sex and gender identity as predictors. Results showed that moral commitment positively predicted the perceived accuracy of the findings ($\beta = 0.32$, SE = 0.06, t = 5.17, $p < .0001$, $\eta^2 = 0.10$) and the perceived reliability of the methods ($\beta = 0.18$, SE = 0.06, t = 3.0, $p < .003$, $\eta^2 = 0.10$) of the summary reporting hiring bias against women in STEM. Again no sex difference nor any effect of gender identity was observed (all |t| <1).

We did not run the pre-registered mediation analysis as we did not observe any significant effects of sex and gender identity.

Plots of the research evaluations as a function of moral commitment and participant sex are shown in Figure 9.

Experiment 1b

To establish if the effects observed in Experiment 1a with U.K. participants can be generalized to another cultural context, Experiment 1b replicated Experiment 1a with U.S. participants. Adopting the same design, we tested whether individuals higher on moral commitment to gender

equality were more likely to trust research showing hiring discrimination against women, regardless of their sex. All aspects of the current study were pre-registered on OSF (10.17605/OSF.IO/2UG95).

Method

Participants

The data analysis included 419 U.S. participants (166 women and 253 men) aged between 18 and 88 years (M = 38, SD = 12.9) recruited on Amazon Mechanical Turk (https://www.mturk.com/). They were paid $0.50 for their participation. We determined the sample size by running a power analysis in G*Power 3.1 (Faul et al., 2009) which suggested a sample of 415 for a statistical power of 0.90. As data were collected in batches and we applied the exclusion criteria after each batch was complete, we ended up with a sample size slightly larger than the suggested one.

In total, we recruited 768 participants. Data from 349 participants were excluded as 287 failed one of the two attention checks (Note that we performed the same analyses with the 287 participants included, and a similar pattern of results was observed. Removing or including these participants did not qualitatively change the results.) and 62 took the survey twice (only their first response was kept).

Materials and procedure

Materials and procedure of Experiment 1b were identical to those of Experiment 1a, except in the following aspects. First, in order to mitigate chances that participants would be distracted by information of secondary importance, we removed the name of the scientific journal (*PNAS*) where Moss-Racusin et al., (2012)'s article was originally published, as well as the names of the authors except the lead author (see Appendix A).

Second, a short introduction was added before showing the stimulus summary in order to better prepare participants for the task: “In this survey, we are interested in how people think about scientific findings. You will be asked to read very carefully a brief summary of a research article that was published in a scientific journal and then give your opinions as non-expert about the research”.

Finally, we replaced our 1-item measure of moral commitment to gender equality with a 3-item scale in order to increase the reliability of the measure. The three items, inspired from (Skitka et al., 2005) were now: 1) “Achieving gender equality in society is an absolute moral imperative”; 2) “The conviction that we need to fight for gender equality is central to my identity”; and 3) “Furthering gender equality should be the government's utmost priority”. Response scales ranged from [0] “Strongly disagree” to [10] “Strongly agree”, with [5] “I don’t mind” as the default slider position. The three items ($\alpha = 0.89$) were averaged as our measure of moral commitment to gender equality.

Results

We performed the same analyses here as we did for Experiment 1a. Similar to what was observed in Experiment 1a, men ($M = 7.33$, $SD = 1.96$) and women ($M = 7.54$, $SD = 1.61$) did not differ in the perceived accuracy of findings ($t = 1.14$, $p = 0.26$), nor in how reliable they perceived the methods to be (men: $M = 7.44$, $SD = 1.83$; women: $M = 7.47$, $SD = 1.65$; $|t| < 1$). Replicating results of Experiment 1a, moral commitment to gender equality positively predicted perceived accuracy of research findings ($\beta = 0.38$, $SE = 0.04$, $t = 10.6$, $p < .0001$, $\eta^2 = 0.22$) and perceived reliability of the methods ($\beta = 0.30$, $SE = 0.04$, $t = 8.30$, $p < .0001$, $\eta^2 = 0.22$). Again, compared to moral commitment, sex and gender identity were not significant predictors of

research evaluations (all $|t| < 1$). For the same reason as described in Experiment 1a, the pre-registered mediation analysis was again not run.

Plots of the research evaluations as a function of moral commitment and participant sex are shown in Figure 9.

Experiment 1c

Unlike Handley et al (2015) who reported that males were less receptive to evidence demonstrating a hiring bias against women than female participants, we found an effect of moral commitment on individuals' research evaluations across cultures in Experiments 1a and 1b, while no sex difference was observed. There were a few possible explanations for the absence of sex difference in the perception of evidence in our experiments. One is that our dependent measures, which were different from those of Handley et al's (2015), failed to capture the sex difference (if it exists). Another possibility is that men and women reached more consensus on the issue of gender bias over the last few years and thus any sex difference observed before has diminished. To answer this question, here we ran an exact replication of Handley et al's design again with a U.S. sample, by using their original dependent measures in order to allow for a more accurate comparison. All aspects of the current study were pre-registered on OSF (10.17605/OSF.IO/7UW85).

Method

Participants

The data analysis included 467 U.S. participants (233 women and 234 men) aged between 18 and 68 years (M = 35, SD = 11.8) recruited on Prolific. They were paid £0.50 for their participation. We determined the sample size by running a power analysis in G*Power 3.1

(Faul et al., 2009) which suggested a sample of 466 for a statistical power of 0.90. As data were collected in batches and we applied the exclusion criteria after each batch was complete, we ended up with a sample size slightly larger than the suggested one.

In total, we recruited 483 participants. Data from 16 participants were excluded for the following reasons: four did not meet our recruitment criteria (reported sex being “other”), 10 failed one of the two attention checks and two took the survey twice (only their first response was kept).

Materials and procedure

The materials and procedures were identical to those of Experiment 1b, except in the following respects. First, participants were presented with the exact same introduction to the study as in Handley et al., (2015) before they saw the stimulus summary (see the ESM).

On the top of the following page, participants saw another short instruction on the evaluation task: “Please read the following abstract from a 2012 published research study then provide your opinion with the items below” (identical to Handley et al.'s). The same research summary as used in Experiment 1b was then shown to participants. After reading the summary, participants were asked to provide their evaluations of the research by responding to four question items as in Handley et al.: 1) “To what extent do you agree with the interpretation of the research results?”; 2) “To what extent are the findings of this research important?”; 3) “To what extent was the abstract well written?”; 4) “Overall, my evaluation of this abstract is favorable.” Same as in the original study, responses were collected on 6-point Likert scales ranging from [1] "Not at all" to [6] "Very much". Following Handley et al., scores on the scales were averaged to create a composite measure of participants' evaluations of the research ($\alpha = .89$).

A second change from our previous experiments was that we counterbalanced the order in which the moral commitment items and the research evaluation questions were presented. Finally, in this study and the other studies presented below, we kept only the Masculine-Feminine version of the gender identity question and response scale.

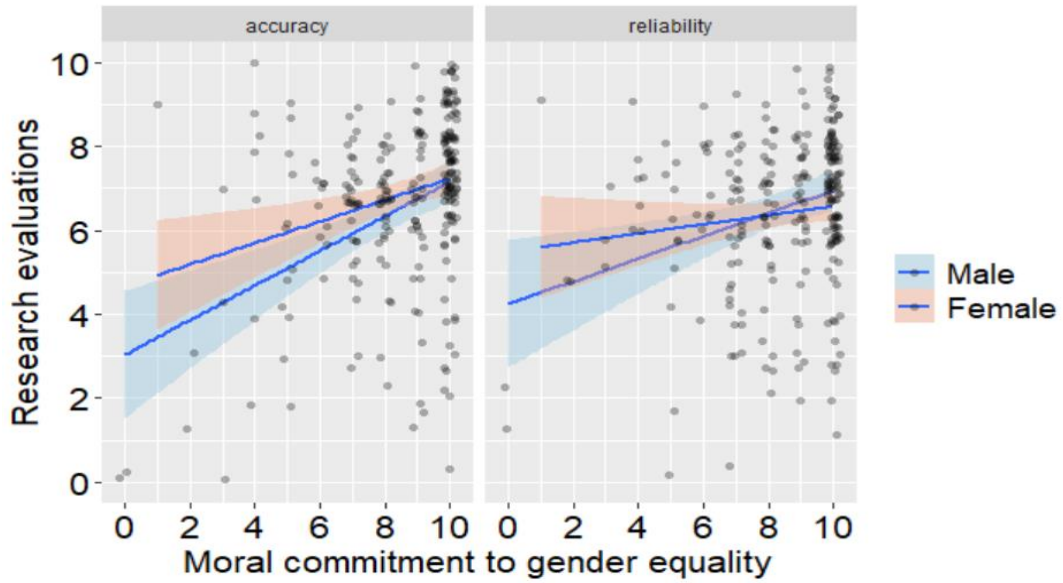
Results

We ran the same analysis as before. This time, the sex effect in Handley et al (2015) was replicated such that men (M = 4.60, SD = 1.04) rated the summary reporting gender bias against female applicants less favorably than women (M = 4.86, SD = 0.96; t = - 2.82, p < .01), albeit the smaller effect size (d = 0.26) compared to what Handley et al. found (d = 0.45).

That said, the sex difference vanished (β = 0.05, SE = 0.04, t = 1.10, p = 0.27) when moral commitment was factored in. Consistent with results of previous experiments, moral commitment positively predicted the perceived overall quality of the research summary (β = 0.19, SE = 0.02, t = 12.06, p < .0001, η^2 = 0.26). Additionally, unlike Experiment 1a and 1b, here we observed an interaction between moral commitment and participant sex – the effect of moral commitment was more pronounced among males than females (β = 0.04, SE = 0.02, t = 2.67, p < .01, η^2 = 0.02). Results also suggested a three-way interaction between moral commitment, presentation order and sex (β = 0.04, SE = 0.02, t = 2.81, p < .01, η^2 = 0.02). Restricted post-hoc analysis revealed that the effect of commitment on evaluations was greater in male than in female participants when the moral commitment scale was presented before the research summary (t = 4.34, p < .001); however, the effect was the same on both sexes when the research summary preceded the moral commitment items (t < 1, p > .9).

Plots of research evaluations (i.e. the perceived accuracy of findings and the perceived reliability of methods in Experiments 1a and 1b, and the perceived research quality in Experiment 1c) as a function of moral commitment and participant sex are shown in Figure 9.

a.



b.

c.

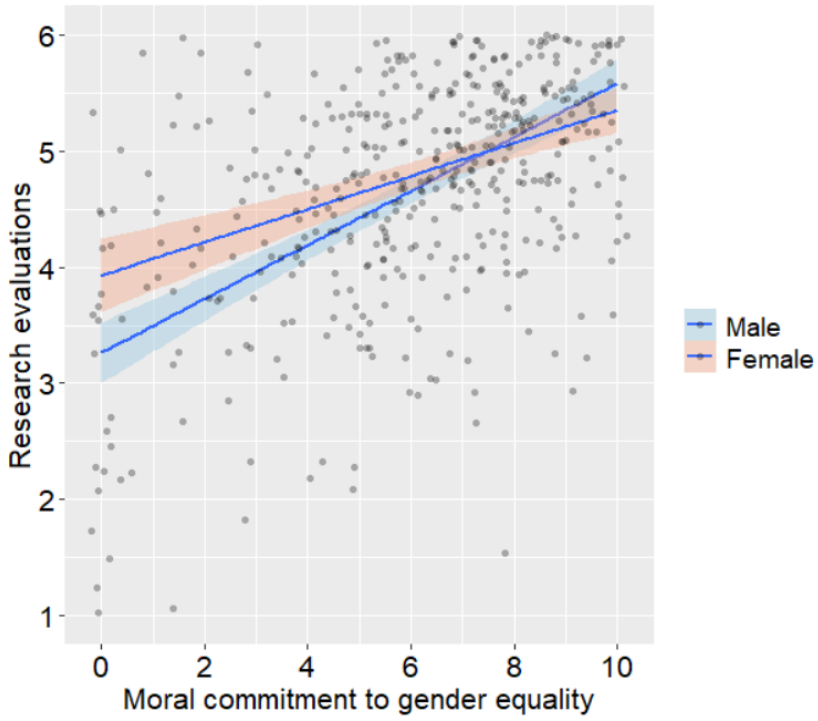


Figure 9. Research evaluations as a function of participants' degree of moral commitment to gender equality and sex. **a**: Experiment 1a (U.K.); **b**: Experiment 1b (U.S.); **c**: Experiment 1c (U.S.)

Discussion of Experiments 1 a, b, c

Similar patterns of results were observed in Experiments 1a -1c. Across U.K. and U.S. samples, the more participants deemed pursuing gender equality in society as a moral imperative, the more they trusted research findings suggesting hiring bias against women in STEM. This result was found both when the research quality measures focused on the perceived accuracy of

the findings and the perceived reliability of the methods employed (Experiments 1a and 1b), and when the composite measure of research quality of Handley et al (2015) – participants' agreement with the interpretation of the results, the perceived importance of the findings, the quality of the writing and their overall evaluation - was used (Experiment 1c).

By contrast, participants' sex had little effect on their trust in the research reporting hiring bias against women. While men judged the findings and the methods slightly less positively than women, the difference was not significant in Experiments 1a and 1b. It was only when adopting Handley et al's composite index of research quality, in Experiment 1c, that the sex difference reached statistical significance - men rated the research less favorably than women. This sex difference, however, was smaller (d = 0.26) than in Handley et al's (d = 0.45) and the sex effect appears to be entirely explained by participants' moral commitment as the effect of participant sex disappeared when the variance accounted for by moral commitment was factored out. This overall pattern may be due to greater awareness of gender discrimination against females among men in 2020 than in 2015. After all, feminism has been on the rise during this period, especially among young people (e.g. the #MeToo movement took place https://metoomvmt.org/).

Experiment 2

Experiments 1a – 1c showed that the more participants were morally committed to gender equality, the more they trusted evidence of gender bias against women in hiring in STEM sciences. The first goal of Experiment 2 was to gauge the generality of this finding by testing an additional research summary reporting bias against women. Second, Experiment 2 explored what would happen if the evidence reported gender bias *favoring* women in hiring, as has been found in some academic contexts (e.g. Williams & Ceci, 2015; Breda & Hillion, 2016). We expected

evidence of hiring bias favoring women to face more skepticism than that of women being discriminated against, and that the skepticism would be amplified by participants' moral commitment to gender equality. All aspects of the current experiment were pre-registered on OSF (10.17605/OSF.IO/W9KUS).

Method

Participants

The data analysis included 636 U.K. participants (319 women and 317 men) aged between 18 and 76 years (M = 36, SD = 13.2), recruited on Prolific. They were paid £0.50 for their time. Again, we determined the sample size by applying the same criterion as described in Experiments 1b and 1c. We ran a power analysis in G*Power 3.1 (Faul et al., 2009) which suggested a sample of 618 for a statistical power of 0.90. For the same reason as in previous experiments, we ended up with a sample size slightly larger than the suggested one.

In total, we recruited 660 participants. Data from 24 participants were excluded for the following reasons: two did not meet our recruitment criteria (reported sex being "other"), 20 failed one of the two attention checks and two took the survey twice (only their first response was kept).

Materials and procedure

The materials used in Experiment 2 were four summaries reporting evidence of gender discrimination that in two cases described a hiring bias "against women" and in the other two cases a hiring bias "favoring women". One of the "against women" summaries was based on the article by Moss-Racusin et al., (2012), as used in previous experiments, and one of the "favor women" summaries was based on the article by Williams and Ceci, (2015) which reported evidence of hiring bias *favoring* women in STEM. In order to vary our stimuli while controlling

for a potential influence of wording, we also created a counterfactual version of each of these two (real) summaries. The counterfactual summaries were identical to the originals except that the results were reversed so that they indicated evidence of gender bias in the opposite direction than the originals (i.e. original: *against* women -> counterfactual: *favoring* women; original: *favoring* women -> counterfactual: *against* women). All summaries had the same structure as in Experiment 1b (See Appendix A).

The procedure was identical to that of Experiment 1b except in the following aspects. In a 2 x 2 between-subjects design, participants were randomly presented with one of the four summaries (against women original; against women counterfactual; favoring women original; favoring women counterfactual). Participants who saw the counterfactual summaries were informed at the end of the survey that they were in fact fictitious, and presented with the actual version. As in Experiment 1c, the order in which the moral commitment items and the research summary evaluation questions were presented was counterbalanced.

Results

We performed two stepwise linear regressions (direction being "both") with accuracy and reliability ratings as the dependent variables, respectively, and reported bias, moral commitment, sex, gender identity, presentation order, summary version, and all interactions as predictors. Below we report results of the converged models.

Here, we added a random intercept for each individual summary in the linear model as there were four summaries tested. Results of the linear mixed-models were shown in Table 10.

Table 10. Results of linear regression

DV	IV	β	SE	t	p
Accuracy	Bias (favoring)	-0.77	0.07	-10.4	**<.0001**
	Moral commitment	0.14	0.04	3.74	**<.001**
	Bias (favoring)*Moral commitment	-0.12	0.04	-3.37	**<.001**
Reliability	Bias (favoring)	-0.54	0.07	-7.32	**<.0001**
	Moral commitment	0.10	0.04	2.70	**0.007**
	Order (com->summary)	-0.17	0.07	-2.31	**0.02**
	Bias (favoring)*Moral commitment	-0.10	0.04	-2.68	**0.007**

The results revealed a main effect of bias direction such that the findings of the "bias favoring women" summaries were rated as less accurate (M = 5.53, SD = 2.05) than those of the "bias against women" summaries (M = 7.07, SD = 1.79; β = 0.77, SE = 0.07, t = -10.4, p <.0001). Likewise, the methods of the "bias favoring women" summaries were perceived as less reliable (M = 5.61, SD = 1.94) than those of the "bias against women" summaries (M = 6.69, SD = 1.86; β = 0.54, SE = 0.07, t = -7.32, p <.0001).

Overall, higher degree of moral commitment predicted higher perceived accuracy of findings (β = 0.14, SE = 0.04, t = 3.74, p <.001), and perceived reliability of the methods (β = 0.10, SE = 0.04, t = 2.70, p <.01). However, these effects of moral commitment were qualified by interactions with the reported direction of bias, as the post-hoc analyses showed significant effect of moral commitment on the perceived accuracy of the findings (β = 0.26, SE = 0.05, t = 5.61, p <.0001, η^2 = 0.09), and perceived reliability of the methods (β = 0.20, SE = 0.05, t = 4.02,

p <.0001, η^2 = 0.05) when the summaries reported bias "against" women. By contrast, when the reported bias was "favoring" women, the effect of moral commitment disappeared (both p >.5). Research methods were judged to be less reliable when the commitment items were presented before (M = 5.96, SD = 2.01) than after the evaluation task (M = 6.34, SD = 1.92; β = 0.17, SE = 0.07, t = 2.31, p = .02). Nevertheless, this order effect was characterized by its interaction with the direction of bias, as it was found marginally significant only for the "bias favoring women" summaries (β = 0.23, SE = 0.11, t = 2.09, p = .04), but not for the "bias against women" summaries (β = 0.12, SE = 0.10, t = 1.14, p = .26). No order effect was observed on ratings of accuracy of the findings (p > .2) for both types of summaries.

The converged model showed no sex difference nor effect of gender identity (all p > .5). To test if there was any sex effect when the variance explained by moral commitment was factored out, we performed the regression analyses with bias direction, participant sex and their interaction term as predictors. Consistent with results shown above, the findings of the "bias favoring women" summaries were rated as less accurate than those of the "bias against women" summaries (β = 0.76, SE = 0.08, t = -10.2, p <.0001). Likewise, the methods of the "bias favoring women" summaries were perceived as less reliable than those of the "bias against women" summaries (β = 0.54, SE = 0.08, t = -7.17, p <.0001). In addition, we found an interaction between sex and bias direction (accuracy rating: β = 0.26, SE = 0.08, t = 3.48, p < .001, reliability rating: β = 0.17, SE = 0.08, t = 2.26, p = .02). For "bias against women" summaries, men (M = 6.75, SD = 1.86) rated the findings as less accurate than women (M = 7.39, SD = 1.65; t = -3.21, p <.001); similarly, men (M = 6.49, SD = 2.04) rated the methods as marginally less reliable than the women counterparts (M = 6.89, SD = 1.64; t = -1.96, p < .05).

Conversely, for the "bias favoring women" summaries, men rated the findings as equally accurate as women ($t = 1.72$, $p > .05$), and the same for the ratings of methods ($t = 1.23$, $p > .1$).

Plots of research evaluations (i.e. the perceived accuracy of findings and the perceived reliability of methods) as a function of moral commitment and participant sex are shown in Figure 10.

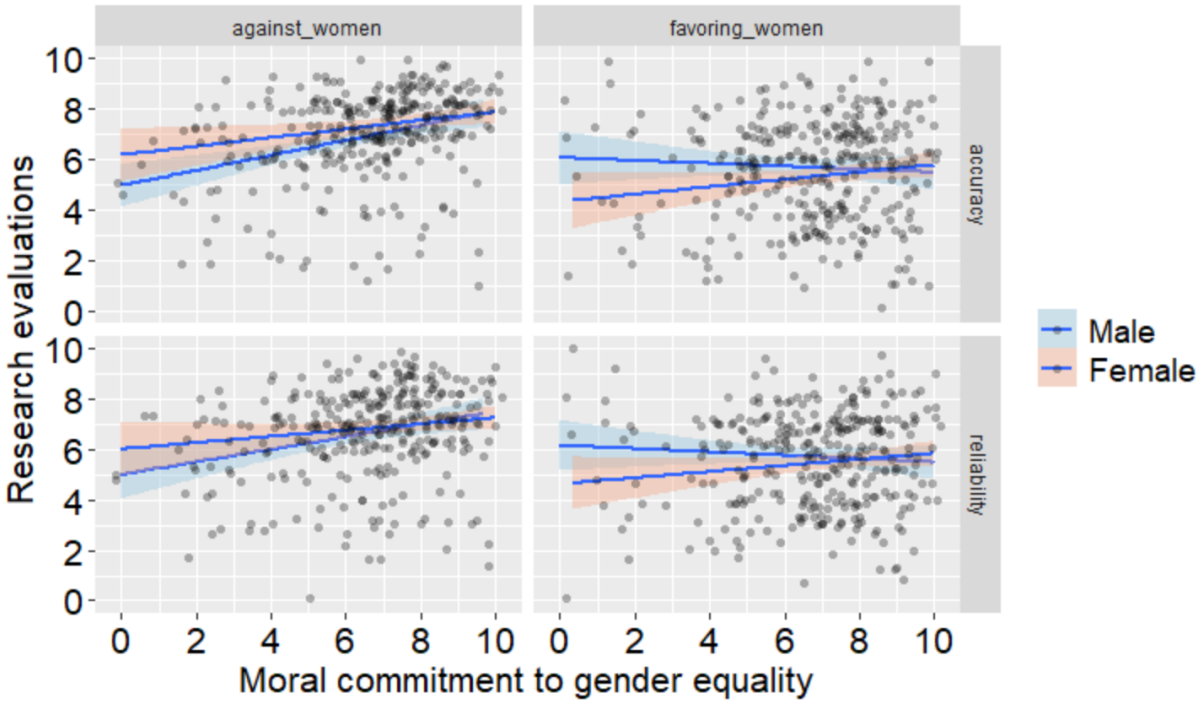


Figure 10. Evaluations of research showing evidence of "bias against women" and "bias favoring women" as a function of moral commitment to gender equality and participant sex. Top: the perceived accuracy of findings; Bottom: the perceived reliability of methods.

Discussion

Replicating results from Experiments 1a-1c, Experiment 2 found that individuals' moral commitment to gender equality affected their evaluations of the research demonstrating a hiring bias against women in STEM. People with positive moral attitudes on gender equity were more likely to accept evidence of women being discriminated against, while those with negative attitudes were more resistant to this evidence.

The influence of moral commitment seems to be restricted to the scenario where women were described as the victims of hiring bias but not when they were the beneficiaries, as participants scoring higher on moral commitment did not exhibit stronger resistance to the evidence of hiring preference for women (bias against men). One might argue that the effects of moral commitment on the evaluations of "bias against women" evidence could be attributed to the possibility that people who are morally concerned for gender equity also have overall higher trust in science. If this was true, we should have observed a similar pattern of effects for the "bias favoring women" summaries. However, this was not what we observed and this possibility should thus be ruled out.

Experiment 3

Experiments 1-2 found that moral commitment to gender equality positively predicted participants' level of trust in scientific evidence of hiring discrimination against females in academia. While this type of attitude-evaluation association is typically observed on polarizing issues, the psychological mechanism that underpins it is unclear. According to a narrow view, the positive correlation between participants' moral commitment to gender equality and their evaluations of specific evidence of discriminatory hiring practices against women may be

explained entirely by their prior expectations about the degree of discrimination that women would face in the relevant scenario. Alternatively, it may be that the effect of moral commitment on evaluations of scientific evidence goes over and above these prior expectations in such a way that we can consider any such effects to be conceptually separate from those of specific prior expectations.

In order to examine the extent to which an association between moral commitment and research evaluations, when it is found, can be reduced to participants' expectations about the likely size of the hiring bias, Experiment 3 adopted a design that allowed participants to both predict how much discrimination against women they thought would be observed in the study before it was conducted, and to evaluate how credible they perceived the study once its results were available. Similar to Experiment 2, it also manipulated the direction of the bias described in the summaries: hiring bias against vs. favoring women.

All aspects of the current study concerning the materials, procedure, and analyses were pre-registered on OSF (10.17605/OSF.IO/BVKJC)

Method

Participants

The data analysis included 517 UK participants (265 women and 252 men) aged between 18 and 80 years (M = 36, SD = 13), recruited on Prolific. They were paid £0.50 for their time. Again, we determined the sample size by running a power analysis in G*Power 3.1 (Faul et al., 2009) which suggested a sample of 502 for a statistical power of 0.90. For the same reason as in previous experiments, we ended up with a sample size slightly larger than the suggested one.

In total, we recruited 520 participants. Data from three participants were excluded for the following reasons: two failed one of the two attention checks and one took the survey more than once (only their first response was kept).

Materials and procedure

As in Experiment 2, this experiment used a between-subjects design in which participants were randomly exposed to a research summary reporting evidence either of gender bias *against* or *favoring* women, with the former based on the article of Moss-Racusin et al., (2012), and the latter on that of Williams & Ceci (2015). Contrary to Experiment 2 however, only the two original summaries were used (and thus no "counterfactual" scenarios were employed). Moreover, each condition was composed of two tasks. First, participants read the summary of a scientific study about to be conducted on the issue of gender discrimination in hiring in academic sciences and were asked to *predict* its likely results. The research summaries were the same as in previous experiments, except that the study's methods were described in the future tense and no results were shown. Participants' predictions about the study's likely results were collected in the form of average scores they expected the male and female candidates to receive in the study, respectively (on a 0-100 slider scale with the slider initially positioned at 0) (see Appendix A).

Immediately after having made their predictions, participants saw the following message: "The study that you just read about has been recently run by researchers. The research findings were published as an article in a scientific journal. On the next page, you are going to see a summary of the published article. Please read carefully and answer two questions". This time, participants saw the full research summary including its results (both in text and graphical format), presented in the past tense. Following previous experiments, they were asked to rate

how accurate they found the research summary's results, and how reliable they found its methods to be.

As in Experiment 2, the order in which the moral commitment items and the two tasks (prediction and evaluation) were presented was counterbalanced across participants. The design was thus a 2 (bias against vs. favoring women) x 2 (commitment items before vs. after the prediction and evaluation tasks) between-subjects design.

Results

Linear regression analysis

We performed two stepwise linear regressions (direction being "both") with accuracy and reliability ratings as the dependent variables, and reported bias, moral commitment, prior beliefs, sex, gender identity, presentation order, and all interactions as predictors. Below we report results of the converged models.

Results of linear regressions are shown in Appendix B1. As observed in Experiment 2, the "bias favoring women" summary (M = 5.86, SD = 2.05) received lower accuracy ratings than the "bias against women" summary (M = 7.36, SD = 1.81; β = 0.77, SE = 0.08, t = -9.49, p <.0001); and the method reported in the "bias favoring women" summary (M = 5.99, SD = 1.95) were also rated as less reliable than that of the "bias against women" summary (M = 6.98, SD = 1.81; β = 0.51, SE = 0.08, t = -6.27, p <.0001).

Again, overall moral commitment positively correlated with the perceived accuracy of research findings (β = 0.10, SE = 0.04, t = 2.53, p =.01) and the reliability of methods (β = 0.11, SE = 0.04, t = 2.74, p <.01). However, these effects of moral commitment were again qualified by their interactions with the reported direction of bias (accuracy: β = 0.13, SE = 0.04, t = 3.31, p = .001; reliability: β = 0.10, SE = 0.04, t = 2.33, p = .02), such that when the summary reported

hiring bias against women, participants' with higher degree of moral commitment rated the findings as more accurate (β = 0.24, SE = 0.05, t = 4.73, p <.0001, η^2 = 0.11), and the methods as more reliable (β = 0.22, SE = 0.05, t = 4.15, p <.0001, η^2 = 0.06). Conversely, when the summary reported hiring bias favoring women, the effect of moral commitment disappeared (both p > .5).

Next, higher prior beliefs of gender bias against women predicted higher ratings on the accuracy of research findings (β = 0.01, SE = 0.006, t = 2.03, p = .04), and this effect was moderated by the reported direction of bias (β = 0.02, SE = 0.006, t = 3.72, p <.001). For the "bias against women" summary, expectations about the likelihood of gender discrimination happening positively predicted ratings on the accuracy of research findings (β = 0.04, SE = 0.008 t = 4.55, p <.0001, η^2 = 0.07), but again the effect of priors disappeared for the "bias favoring women" summary (β = 0.01, SE = 0.008, t = 1.30, p = .20).

As before, participants who reported their moral commitment before the prediction and evaluation tasks tended to rate the findings as marginally less accurate (β = 0.15, SE = 0.08, t = 1.82, p = .07) and the methods as less reliable (β = 0.19, SE = 0.08, t = 2.40, p = .02). This order effect was conditioned on the reported direction of bias (accuracy: β = 0.19, SE = 0.08, t = 2.38, p = .02; reliability: β = 0.22, SE = 0.08, t = 2.73, p <.01). The results of the "bias favoring women" summary were rated as less accurate (β = 0.36, SE = 0.13 t = -2.87, p < .01, η^2 = 0.03), and its methods as less reliable (β = 0.42, SE = 0.12 t = -3.50, p <.001, η^2 = 0.05) when participants saw the commitment items before the prediction and evaluation tasks than in the reversed order. Conversely, for the "bias against women" summary, no order effect was observed (both p >.5).

Similar to results of Experiment 2, when moral commitment was factored in, no effect of sex or gender identity, nor any interaction with other factors was found (all $p > .05$). However, when the variance explained by moral commitment was factored out, an interaction between sex and the reported direction of bias was found (accuracy: $\beta = 0.30$, SE = 0.08, $t = 3.64$, $p < .001$; reliability: $\beta = 0.23$, SE = 0.08, $t = 2.81$, $p < .01$). For the "bias against women" summary, men rated the findings as significantly less accurate (Men: M = 6.96, SD = 1.74; Women: M = 7.73, SD = 1.73; $t = 3.28$, $p < .01$) and the methods as slightly less reliable (Men: M = 6.86, SD = 1.81; Women: M = 7.25, SD = 1.79; $t = 1.69$, $p = 0.09$) than did women. In comparison, for the "bias favoring women" summary, men rated the findings as slightly more accurate (Men: M = 5.99, SD = 2.05; Women: M = 5.55, SD = 2.04; $t = 1.87$, $p = 0.06$) and the methods as significantly more reliable (Men: M = 6.28, SD = 2.01; Women: M = 5.75, SD = 1.89; $t = 2.29$, $p = 0.02$) as did women.

Correlational mediation analyses

To further assess the triangular relationships between prior beliefs, moral commitment and research evaluations, we further performed correlational mediation analyses by following the steps set forth by Baron and Kenny (1986), using the R package *mediation* (Tingley et al., 2014). These analyses were run only for the "bias against women" condition, aiming to test two (theorized) mediation effects: 1) the relationship between moral commitment and research evaluations is mediated by priors, and 2) the relationship between priors and research evaluations is mediated by moral commitment. The results, obtained with 1000 times bootstrap simulations, showed two significant partial mediation effects (see Appendices B2 and B3)

First, participants' prior beliefs partially mediated the positive correlations between moral commitment and the ratings on the accuracy of the research findings (Average Causal Mediating

Effect [13] of priors: β = 0.05, p <.0001; Proportion of the total effect that is mediated: β = 0.16, p <.0001), and between moral commitment and the ratings on the reliability of research methods (Average Causal Mediating Effect of priors: β = 0.04, p <.01; Proportion of the total effect that is mediated: β = 0.16, p <.01).

Similarly, participants' moral commitment partially mediated the positive correlations between moral commitment and the ratings on the accuracy of the research findings (Average Causal Mediating Effect of priors: β = 0.007, p <.01; Proportion of the total effect that is mediated: β = 0.17, p <.01), and between prior beliefs and the ratings on the reliability of research methods (Average Causal Mediating Effect of priors: β = 0.03, p <.01; Proportion of the total effect that is mediated: β = 0.17, p <.01).

Taken together, these findings suggest the distinct effects of prior beliefs and moral commitment on people's evaluations of "bias against women" evidence.

Plots of research evaluations (i.e. the perceived accuracy of findings and the perceived reliability of methods) as a function of moral commitment and participant sex are shown in Figure 11, and plots of research evaluations as a function of predicted gender bias (i.e. the likelihood of hiring bias against women) and participant sex are shown in Figure 12.

[13] Despite the parameter name, the analyses were correlational as the data were observational and we did not have a treatment condition.

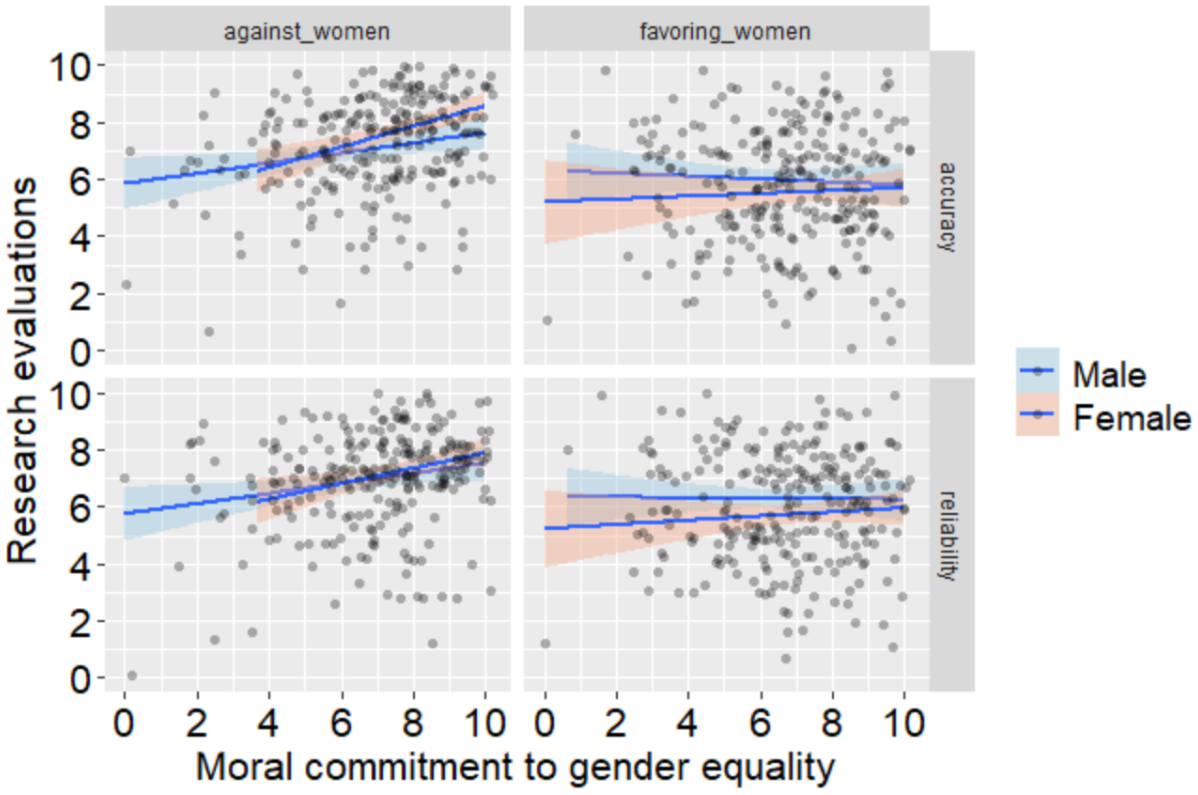


Figure 11. Evaluations of research showing evidence of "bias against women" and "bias favoring women" as a function of moral commitment to gender equality and participant sex. Top: the perceived accuracy of findings; Bottom: the perceived reliability of methods.

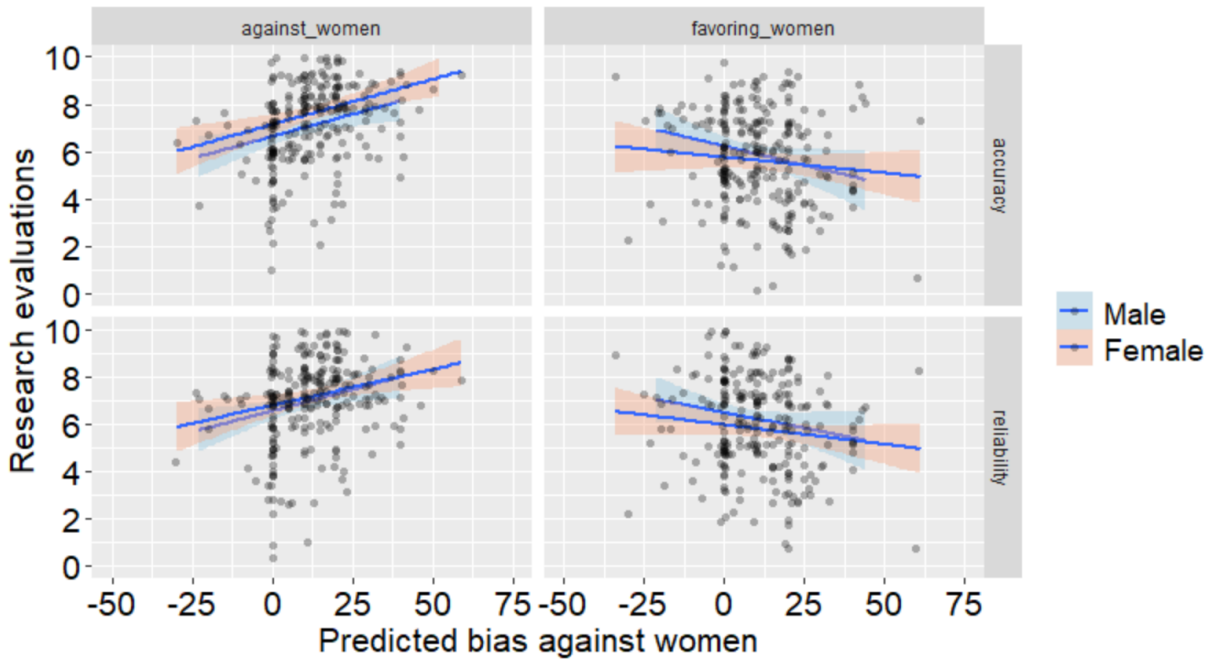


Figure 12. Evaluations of research showing evidence of "bias against women" and "bias favoring women" as a function of predicted gender bias and participant sex. Top: the perceived accuracy of findings; Bottom: the perceived reliability of methods.

Discussion

In line with results from previous experiments, Experiment 3 found that higher moral commitment to gender equality predicted greater trust in evidence of hiring bias against women in STEM in men and women alike. This suggests that (in 2021) individuals' moral concern for gender equality more than their sex explains the level of credence they give to evidence of discrimination against women. Moreover, the results suggested that the effect of moral commitment on evaluations is not reducible to people's prior expectations about the extent to which women are discriminated against in hiring and that people's moral attitudes and prior beliefs have concurrent but differing impacts on their appraisal of evidence.

Experiment 4

Experiments 1-3 showed that participants' moral commitment to gender equality predicts how much they trust evidence of hiring discrimination against women. Experiment 4 focused on another likely consequence of moralization: the propensity to make an imprecise inference when one agrees with the conclusion. More specifically, we hypothesized that increased moral commitment to gender equality would make people more prone to infer sex-based hiring discrimination against women from the mere observation that the profession comprises more males than females.

The literature remains controversial as to the role of external bias, relative to innate or socialized individual differences in math abilities, lifestyle, and career choices in the underrepresentation of women in STEM fields (Breda & Napp, 2019; Ceci et al., 2009; Ceci & Williams, 2010, 2011; Charlesworth & Banaji, 2019). These factors are often intertwined and interacting with each other. To detect the effect of external bias, one needs to have the other factors maximally constant as any relationships observed in correlation studies are likely to be confounded, thus not necessarily causal. For example, Moss-Rocusin et al. (2012) provided clear, unequivocal evidence of sex-based discrimination as factors other than the sex of the applicant were well controlled for in the experiment. However, the mere observation of larger proportions of male research trainees in science labs is not statistical evidence of external bias as factors other than hiring bias (e.g. the base rate of male and female applicants for the positions, and their credentials) can also be in play (Sheltzer & Smith, 2014).

Method

Participants

The data analysis included 264 UK participants (198 women and 66 men) aged between 18 and 62 years (M = 29, SD = 10.9) recruited on Prolific. They were paid £0.50 for their time. As in previous experiments, we determined the sample size by running a power analysis in G*Power 3.1 (Faul et al., 2009) which suggested a sample of 255 for a statistical power of 0.90. For the same reason as described above, we ended up with a sample size slightly larger than the suggested one.

In total, we recruited 280 participants. Data from 16 participants were excluded for the following reasons: one did not meet our recruitment criteria (reported sex being "other"), 13 did not complete the survey and two failed one of the two attention checks.

Materials and procedure

The stimulus summary was based on the research article of (Sheltzer & Smith, 2014), presented in the same format as the summaries tested before (see Appendix A). In the original study, the authors examined the gender distribution of biomedical scientists in academia by collecting information on post-doctoral researchers and professors employed in 39 departments at 24 of the highest-ranked research institutions in the United States. They focused on departments studying molecular biology, cell biology, biochemistry, and/or genetics. The original results reported that the labs employed fewer female than male post-docs in labs run by senior professors. Importantly, the report did not take this observed gender imbalance as evidence of gender discrimination specifically targeted at women. This would imply committing a fallacy, as women may be underrepresented in a profession because they simply are less interested in the job, for instance, and consequently apply less. Nonetheless, we intentionally

wrote the conclusion of the summary in such a way that it did commit the fallacy. The key passage of the summary was formulated as follows (see Appendix A):

> *Results*: [Graph showing a greater proportion of male than female postdocs] "On average, laboratories comprised significantly fewer female postdocs than male postdocs."
> *Conclusion*: "The study provides clear evidence of women being discriminated against in academia because of their sex, a characteristic that should not matter for research work."

The experimental procedure of Experiment 4 was identical to that of Experiment 1b except that an additional item was introduced to measure participants' ability to spot the fallacy: "To what degree do you think the researchers' conclusion is justified by the research results?". Participants responded on an 11-point scale (0 "Not at all justified – 10 "Absolutely justified").

Results

As shown in Table 11, people with higher commitment to gender equality were more likely to take sex imbalance as evidence of gender discrimination ($\beta = 0.44$, SE = 0.08, $t = 5.62$, $p < .001$, $\eta^2 = 0.10$). Replicating the results of previous experiments, the more people regarded gender equality as a moral imperative, the more they believed the reported finding that the research labs comprised a larger proportion of male than female post-docs to be accurate ($\beta = 0.34$, SE = 0.06, $t = 5.68$, $p < .001$, $\eta^2 = 0.10$), and the research methods to be reliable ($\beta = 0.42$, SE = 0.07, $t = 6.11$, $p < .001$, $\eta^2 = 0.12$).

Table 11. Results of linear regression

DV	IV	*β*	SE	t	*p*
Accuracy	Moral commitment	0.34	0.06	5.68	**< .001**
	Order (com ->task)	0.27	0.12	2.30	**0.02**
Reliability	Moral commitment	0.42	0.07	6.11	**< .001**
	Order (com ->task)	0.33	0.13	2.47	**0.01**
Justification	Moral commitment	0.44	0.08	5.62	**< .001**
	Order (com ->task)	0.26	0.15	1.71	0.09

Contrary to the order effect observed before in Experiments 2 and 3, the perceived accuracy of findings was higher this time when participants reported their moral commitment before seeing the summary ($\beta = 0.27$, SE = 0.12, t = 2.30, p = .02, $\eta^2 = 0.02$), and the same trend was observed for the perceived reliability of methods ($\beta = 0.33$, SE = 0.13, t = 2.47, p = .01, $\eta^2 = 0.02$). However, participants' propensity for imprecise inference was not affected by the presentation order of the commitment items and stimulus summary ($\beta = 0.26$, SE = 0.15, t = 1.71, p = .09, $\eta^2 = 0.01$).

No effect of participant sex was found either when moral commitment was factored in or out (all $p > .05$).

As an exploratory analysis, we performed a one-way Anova test comparing the three types of research ratings: accuracy, reliability, justification. Results showed that ratings on the

three items were significantly different ($F(2, 789) = 16.5$, $p < .0001$), such that the conclusion was thought of being justified ($M = 5.61$, $SD = 2.57$) to a lesser degree than the findings were rated to be accurate ($M = 6.75$, $SD = 1.99$; $p < .0001$) and the methods to be reliable ($M = 6.09$, $SD = 2.28$; $p < .04$); also, the methods were judged as reliable to a lesser degree than the findings were considered as accurate ($p < .003$).

Plots of research evaluations as a function of moral commitment to gender equality and participant sex are shown in Figure 13.

Figure 13. Research evaluations as a function of moral commitment to gender equality and participant sex. Left: the perceived accuracy of findings; Middle: the perceived reliability of methods; Right: the extent to which the conclusion is perceived as justified by the results.

Discussion

Consistent with results of previous experiments, individuals having higher moral commitment to gender equity were more inclined to accept the finding that women are outnumbered by men in academia and to consider the research methods employed to obtain this finding as reliable. When faced with weak evidence, participants overall showed more skepticism about the reliability of the research methods and the extent to which the conclusion was supported by the reported results. However, as predicted, individuals who saw gender equality as a moral imperative were more likely to make an imprecise inference (i.e. the existence of gender bias against women) from the observation of lower proportions of female than male post-docs in research labs.

General discussion

Across six survey-based experiments, we investigated the influences of moral commitment to gender equality on people's trust in scientific research on gender bias. Overall, the experiments consistently revealed a positive relationship between the moralization of gender equality and reactions to evidence of women being discriminated against in academia. Specifically, Experiments 1 – 3 found that participants, regardless of their sex, who scored higher on moral commitment to gender equity tended to rate research findings demonstrating a hiring bias against women as more accurate and the research methods adopted as more reliable. This effect of moral commitment was limited to research summaries reporting a gender bias to the disadvantage of women. By comparing two opposing directions of hiring bias (i.e. *against* vs. *favoring* women), Experiment 2 additionally showed that moral conviction moderated people's evaluations of "bias against women" evidence, but not when the evidence pointed to a "bias favoring women". Regarding the role of prior beliefs in the relationship between moral commitment and trust in

scientific research, results of Experiment 3 suggest that people's prior expectations about the likelihood of women being discriminated against in hiring processes because of their sex and their moral concern for gender equity exert distinct influences on their appraisals of evidence that females suffer sex-based hiring bias in academia. Participants' research evaluations reflected both their pre-existing beliefs and their initial moral attitudes. Furthermore, Experiment 4 demonstrated that people's reasoning is subject to the influence of their moral stand on an issue such that participants with greater moral commitment to gender equality conflated correlation with causation.

Uncovering the reason why people have diverging reactions to scientific evidence of gender bias disfavoring women is critical to the establishment of gender equality in society. Relative to previous research showing that males were more resistant than females to evidence of hiring bias against women in STEM fields (Handley et al., 2015; Moss-Racusin et al., 2015), here we found that people's moral convictions about gender equality, no matter if they are biologically male or female, affect their acceptance of such evidence. An individual can have multiple social identities defined by categories including their sex, profession, ideology and moral views and these identities may not be mutually exclusive. For example, being a male does not preclude one from sympathizing with females in their suffering of unequal treatment and in consequence, seeing gender equality as a moral imperative. Given the objectivity and universality nature of moral convictions (Skitka, 2010), it is not surprising that individuals' moral identity has more profound impacts on their judgments than any other non-moral identities (e.g. biological sex, profession).

By comparing people's evaluations of research reporting a hiring bias *against* vs. *favoring* women, Experiment 2 and 3 suggested that the influence of moral commitment to

gender equity is dependent on the depicted direction of bias. The extent to which individuals feel morally concerned about gender equality affected their reception of "bias against women" evidence, but moral concern did not influence individuals' judgment about "bias favoring women" research results. The differing impacts of moral commitment may be explained by the ambivalent feelings triggered in participants after seeing the unexpected or unfamiliar information that women were actually favored in academia. That is, for individuals who were morally engaged in improving women's status, the message that women have an advantage in academic hiring could be "good" and "bad" news at the same time. It is good news because, for people who value gender equality, any practice that helps to increase women's representation in male-dominated fields seems desirable, consistent with their goals for equality. Nevertheless, it can also be bad news as counter-attitudinal information could undermine collective mobilization for the cause (Petersen, 2020). Participants who interpreted the "bias favoring women" evidence as "good news" might rate the research more favorably than those regarding it as "bad news". In this way, the positive and negative effects of moral commitment simply evened out.

Considering the role of moral commitment and prior beliefs in people's treatment of new information, results of Experiment 3 suggest the concurrent but distinct functioning of the two factors in influencing people's reactions to evidence of gender bias against females in the STEM fields. Individuals' responses to the "bias against women" research are likely to result from the combinatory works of "motivated thinking" (Kunda, 1987, 1990) and Bayesian reasoning. People who hold positive moral attitudes on gender equality are more receptive of "bias against women" research findings not just because they find the information palatable, but also that it fits with their prior knowledge and understanding of the state of affairs. When asked to make judgments pertaining to a moralized issue, individuals rely on both how they feel about the

information and how likely according to their prior knowledge the information appears to be true. Put in another way, they calibrate their trust in evidence by coordinating the works of 'cold' cognition and 'hot' emotion on the mind.

Furthermore, the influences of moral commitment and prior expectations are distinct from each other as reflected by the partial mediation effects shown in Experiment 3. Indeed, individuals' prior factual beliefs about a situation do not always correspond to how much they care about a cause and their moral involvement in it. A person can be well aware of the dire consequences of gender discrimination for women but remain indifferent to it for reasons such as they are not identified with the female gender group or they see the existence of inequality as justified (Napier et al., 2010).

By varying the strength of evidence, we ruled out the mundane explanation for the effects of moral attitudes on people's trust in science that participants scoring higher on moral commitment happen to be better at recognizing reliable scientific evidence. Results of Experiment 4 indicated that individuals with higher moralization of gender equality are inclined to make imprecise inferences when they find the conclusion congenial to their moral stand. However, the results of Experiment 4 do not imply that the participants who did not confuse the correlational data of gender imbalance with sex-based discrimination succeeded at detecting the inherent logical flaw, as they too may have simply distrusted any evidence that appeared to justify a counter-attitudinal conclusion without deliberation. It is likely that once people agree/disagree with a claim, they tend to accept/reject any evidence seemingly in support of it.

The current study has its own limitations. For example, it revealed a correlation between moralization and people's research evaluations, but it remains unclear whether or not moralization actually caused the differing perception of evidence. To answer this question, future

studies can manipulate participants' level of moral concern through moral priming before exposing them to research summaries. In addition, our study showed that people with high moral commitment tend to make an invalid inference to reach a congenial conclusion, but we are unsure if those who resisted such an inference actually identified its potentially fallacious nature or they simply responded based on attitudes. Future research can address this question by eliciting reasons and arguments from participants after they complete the evaluation task and see if they have justifications for their judgments or they make attitude-motivated decisions. Finally, results of the study should be interpreted with one caveat: responses were collected with non-representative samples. Future studies are welcome to look if the effects hold with national representative samples.

To conclude, the study showed that people's moral stand on gender equality affects their trust in scientific evidence of gender bias and this effect is dependent on the described direction of bias. The effects of moral commitment are not entirely reducible to people's prior factual beliefs about the likelihood of a gender bias disfavoring women happening in a hiring process. Furthermore, the moralization of an issue leads to inadequate inferences when people see invalid evidence as valid when it seems to support a conclusion congenial to their attitudes.

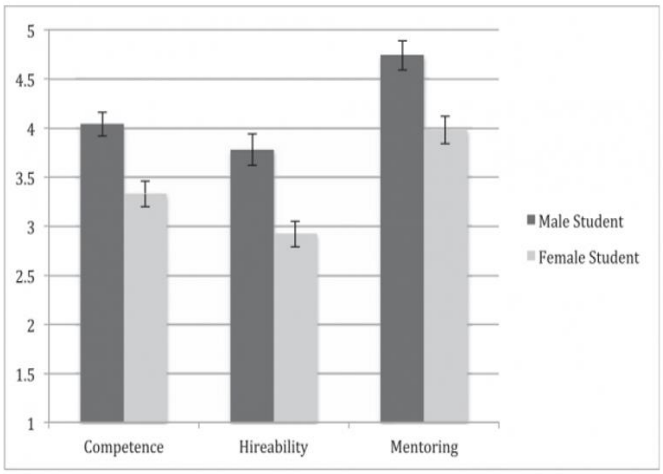
Appendices

Appendix A. Stimuli

A1. Actual "bias against women" research summary based on the article of Moss-Rocusin (2012) used in Experiment 1a

Proceedings of the National Academy of Sciences USA 109(41):16474–16479
Science faculty's subtle gender biases favor male students
C. A. Moss-Racusina, J. F. Dovidio, V. L. Brescoll, M. J. Grahama, and J. Handelsman

Background:
Despite efforts to recruit and retain more women, a stark gender disparity persists within academic science. Abundant research has demonstrated gender bias in many demographic groups, but has yet to experimentally investigate whether science faculty specifically exhibit a bias against female students that could contribute to the gender disparity in academic science.
Design:
The present study asked science faculty members (N = 127) from research-intensive universities to rate the application materials of a student for a laboratory manager position (which, in scientific labs, is often occupied by a student). The student applicant was randomly given a male or female name, but was described as having exactly the same skills in both groups.
In both groups, faculty members were asked to: rate the student's perceived competence and hireability, propose a starting salary, and offer a certain amount of career mentoring to the student.
A comparison between average ratings of student's perceived competence, hireability, proposed salary and career mentoring was conducted between the male student and the female student groups.
Results:



- Science faculty members rated the male student as significantly more competent and hireable than the (identical) female student.

- Faculty members also offered more career mentoring to the male student and selected a higher starting salary.
- The gender of the faculty did not affect their ratings, such that female and male faculty were equally likely to exhibit bias against the female student.

Conclusions:
These results suggest that well-documented bias against females also exists in academia

A2. Actual “bias against women” research summary based on the article of Moss-Rocusin (2012) used in Experiments 1b, 1c, 2, and 3

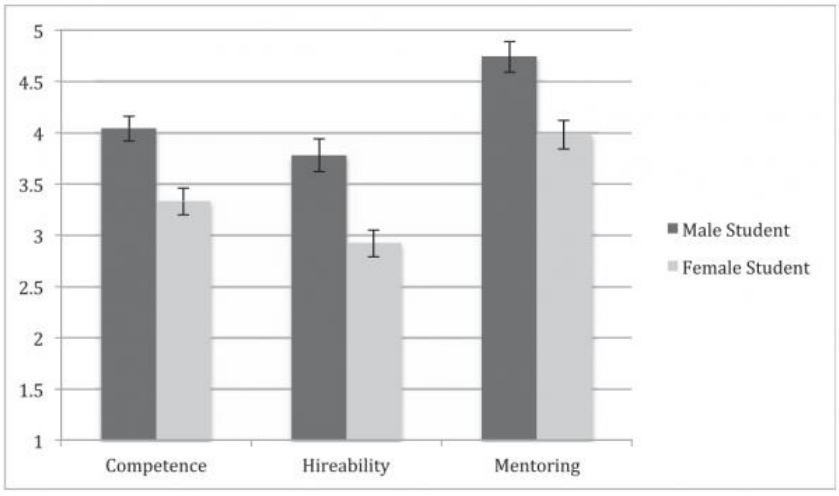
Science faculty’s subtle gender biases favor male students

Moss-Racusin et al.
Yale University, New Haven, CT 06520

Background:
Despite efforts to recruit and retain more women, a stark gender disparity persists within academic science. Abundant research has demonstrated gender bias in many demographic groups, but has yet to experimentally investigate whether science faculty specifically exhibit a bias against female students that could contribute to the gender disparity in academic science.

Design:
The present study asked science faculty members (N = 127) from research-intensive universities to rate the application materials of a student for a laboratory manager position (which, in scientific labs, is often occupied by a student). The student applicant was randomly given a male or female name, but was described as having exactly the same skills in both groups.
In both groups, faculty members were asked to: rate the student's perceived competence and hireability, propose a starting salary, and offer a certain amount of career mentoring to the student.
A comparison between average ratings of student's perceived competence, hireability, proposed salary and career mentoring was conducted between the male student and the female student groups.

Results:



Science faculty members rated the male student as significantly more competent and hireable than the (identical) female student.

- Faculty members also offered more career mentoring to the male student and selected a higher starting salary.
- The gender of the faculty did not affect their ratings, such that female and male faculty were equally likely to exhibit bias against the female student.

Conclusions:
These results suggest that well-documented bias against females also exists in academia.

A3. Introductory text preceding the stimulus summary (identical to Handley et al. 2015) in Experiment 1c

In the scientific world, peer experts judge the quality of research and decide whether or not to publish it, fund it, or discard it. But what do everyday people think about these articles that get published? We are conducting an academic survey about people's opinions about different types of research that was published back in the last few years. You will be asked to read a very brief research summary and then answer a few questions about your judgments as non-experts about this research. There is no right or wrong answer and we realize you don't have all the information or background. But just like in the scientific world, many judgments are made on whether something is quality science or not after just reading a short abstract summary. So to create that experience for you, we ask that you just provide your overall reaction as best you can even with the limited information. You will also be asked to provide demographic information about yourself.

A4. Actual "bias favoring women" research summary based on the article of Williams and Ceci (2015) used in Experiment 2

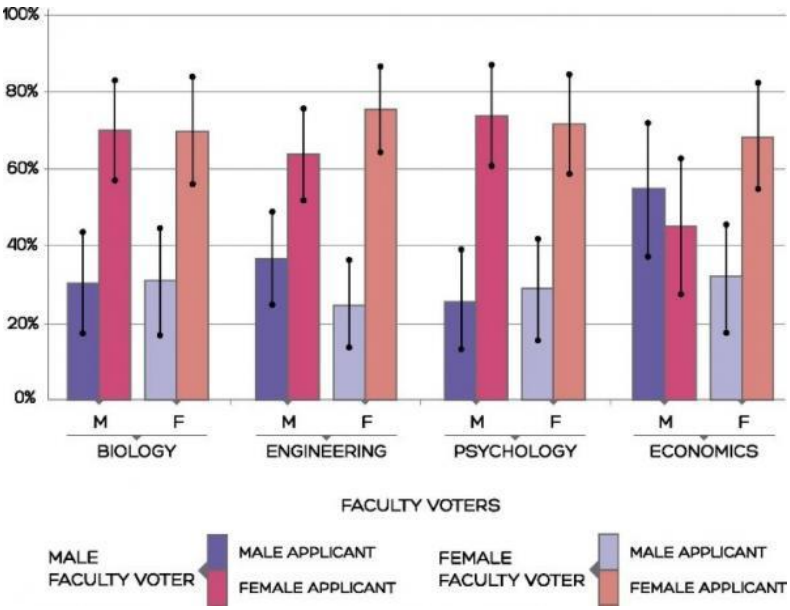
National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track

W. M. Williams and S. J. Ceci
Cornell University, Ithaca, NY 14853

Background:
In life and social sciences, women now earn the majority of doctorates, but they make up a minority of assistant professors. The underrepresentation of women in academic science is typically attributed, both in scientific literature and in the media, to sexist hiring.

Design:
In the present study, 363 faculty members (182 women, 181men) were asked to evaluate hypothetical narrative summaries describing identically qualified female and male applicants for tenure-track assistant professorships in biology, engineering, economics, and psychology. The profiles were systematically varied to disguise identical academic credentials; applicants shared the same lifestyle (e.g., single without children, married with children); and the profiles were counterbalanced by gender across faculty.

Results:



Contrary to prevailing assumptions, our data revealed that men and women faculty members from all four fields preferred female applicants 2:1 over identically qualified males with matching lifestyles (single, married, divorced), with the exception of male economists, who showed no gender preference.

Conclusions:
Our findings, supported by real-world academic hiring data, suggest advantages for women to launch careers in academic science.

A5. Counterfactual "bias against women" research summary based on the article of Williams and Ceci (2015) used in Experiment 2

National hiring experiments reveal 2:1 faculty preference for men on STEM tenure track

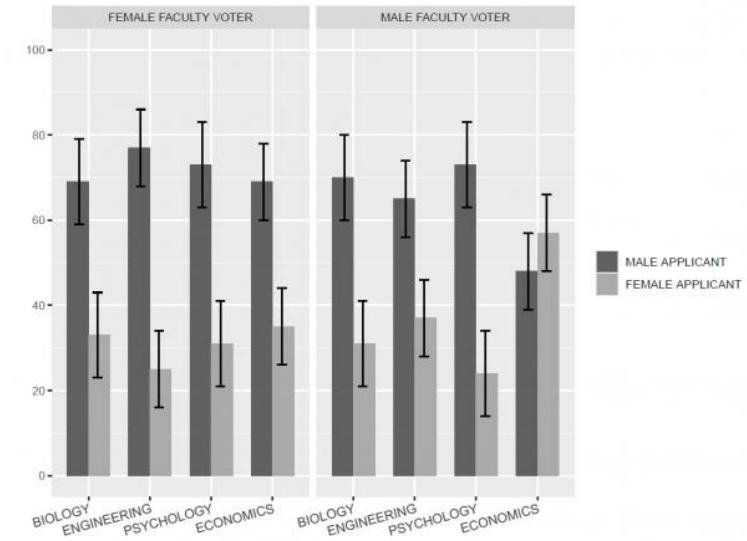
W. M. Williams and S. J. Ceci

Cornell University, Ithaca, NY 14853

Background:
In life and social sciences, women now earn the majority of doctorates, but they make up a minority of assistant professors. The underrepresentation of women in academic science is typically attributed, both in scientific literature and in the media, to sexist hiring.

Design:
In the present study, 363 faculty members (182 women, 181men) were asked to evaluate hypothetical narrative summaries describing identically qualified female and male applicants for tenure-track assistant professorships in biology, engineering, economics, and psychology. The profiles were systematically varied to disguise identical academic credentials; applicants shared the same lifestyle (e.g., single without children, married with children); and the profiles were counterbalanced by gender across faculty.

Results:



In line with prevailing assumptions, our data revealed that men and women faculty members from all four fields preferred male applicants 2:1 over identically qualified females with matching lifestyles (single, married, divorced), with the exception of male economists, who showed no gender preference.

Conclusions:
Our findings, supported by real-world academic hiring data, suggest that women encounter gender discrimination when launching careers in academic sciences.

A6. Counterfactual "bias favoring women" research summary based on the article of Moss-Rocusin (2012) used in Experiment 2

Science faculty's subtle gender biases favor female students

Moss-Racusin et al.

Yale University, New Haven, CT 06520

Background:

Despite efforts to recruit and retain more women, a stark gender disparity persists within academic science. Abundant research has demonstrated gender bias in many demographic groups, but has yet to experimentally investigate whether science faculty specifically exhibit a bias against female students that could contribute to the gender disparity in academic science.

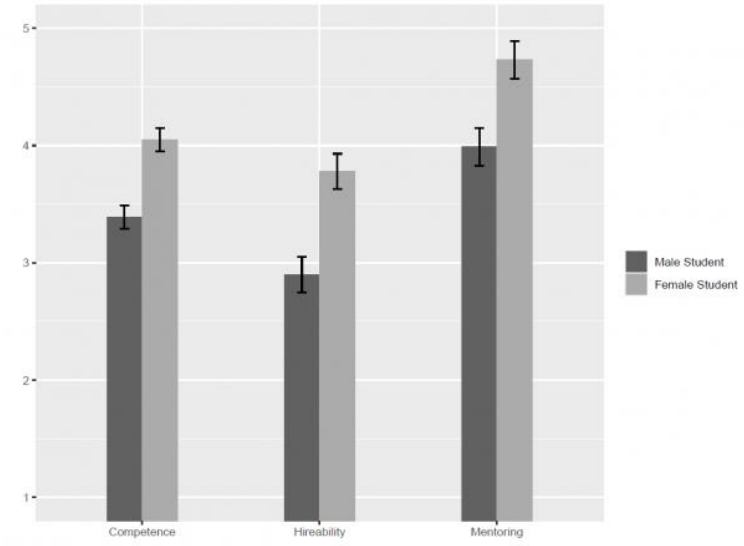
Design:

The present study asked science faculty members (N = 127) from research-intensive universities to rate the application materials of a student for a laboratory manager position (which, in scientific labs, is often occupied by a student). The student applicant was randomly given a male or female name, but was described as having exactly the same skills in both groups.

In both groups, faculty members were asked to: rate the student's perceived competence and hireability, propose a starting salary, and offer a certain amount of career mentoring to the student.

A comparison between average ratings of student's perceived competence, hireability, proposed salary and career mentoring was conducted between the male student and the female student groups.

Results:



- Science faculty members rated the female student as significantly more competent and hireable than the (identical) male student.
- Faculty members also offered more career mentoring to the female student and selected a higher starting salary.

- The gender of the faculty did not affect their ratings, such that female and male faculty were equally likely to exhibit bias in favor of the female student.

Conclusions:
These results suggest that well-documented bias against females does not always exist in academia.

A7. The stimulus study and instructions for the prediction task in the "bias against women" condition in Experiment 3

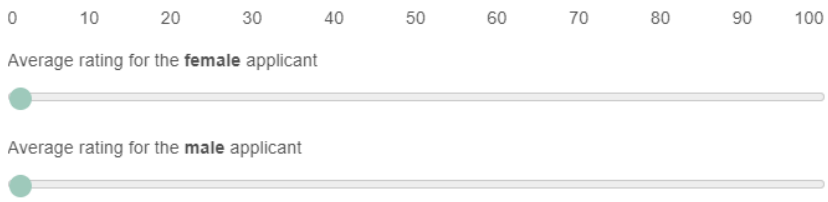
Is there a gender bias in hiring?

Background:
There are ongoing debates about gender bias in various professions. This study attempts to find out whether there exists a gender bias in hiring in academic sciences or not.

Design:
The present study will ask science faculty members (N = 127) from research-intensive universities to rate the application materials of a student for a laboratory manager position (which, in scientific labs, is often occupied by a student). The student's application materials will be shown to two different groups of professors. Half of the professors will be randomly assigned to "Group A," where they will see the candidate presented with a male name, while the other half will be randomly assigned to "Group B," where they will see the candidate presented with a female name. Crucially however, all of the other application materials will be perfectly identical. This will allow the researchers to assess whether there is any bias in the evaluation process that is solely due to the gender of the candidate.
In both groups, faculty members will be asked to rate the student's perceived competence and hireability on a scale from 0 to 100.
A comparison between average ratings of student's perceived competence and hireability will be conducted between faculty members assigned to Groups A and B (i.e. professors who saw the candidate presented as a male and those who saw the candidate presented as a female).

Now please make predictions about the likely results of the study.

What do you think will be the average ratings of the two groups?



A8. The study for the prediction task in the “bias favoring women” condition in Experiment 3

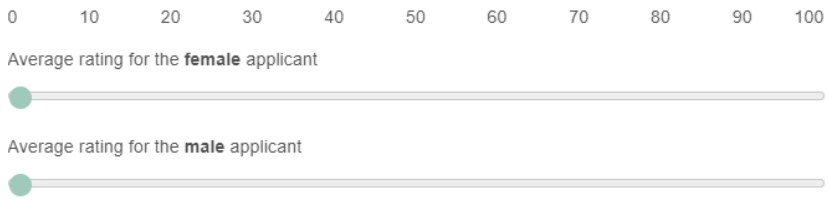
Is there a gender bias in hiring?

Background:
There are ongoing debates about gender bias in various professions. This study attempts to find out whether there exists a gender bias in hiring in the field of academic sciences or not.

Design:
In the present study, 363 science faculty members (182 women, 181 men) will be asked to evaluate hypothetical narrative summaries of applicants for tenure-track assistant professorships in biology, engineering, economics, and psychology. The applicants will be depicted as having the exact same academic credentials and lifestyles (e.g., single without children, married with children). Importantly, however, applicants will be given a female name in group A, and a male name in group B. This will allow the researchers to assess whether there is any bias in the evaluation process that is solely due to the gender of the candidate.
Each faculty member will be randomly assigned to either group A (female applicant) or group B (male applicant). In both groups, they will be asked to rate the applicant's profile on a scale from 0 to 100.
A comparison between average ratings will be conducted between faculty members assigned to groups A and B (i.e. professors who saw the candidate presented as a male and those who saw the candidate presented as a female).

Now please make predictions about the likely results of the study.

What do you think will be the average ratings of the two groups?



A9. The "bias against women" research summary based on the article of Sheltzer and Smith (2014) used in Experiment 4

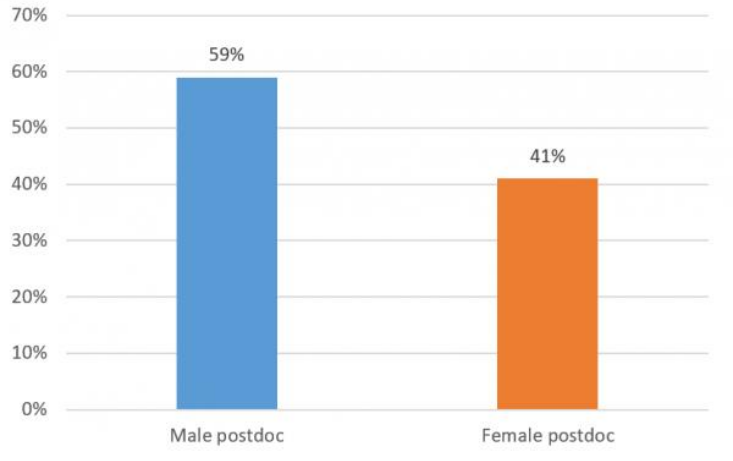
Elite faculty in the life sciences employ fewer women

Sheltzer et al.
Massachusetts Institute of Technology, Cambridge, MA 02139

Overview
Despite many good intentions and initiatives to recruit and retain more women, a wealth of evidence shows that gender inequality is still rife in the fields of science, technology, engineering, and mathematics (STEM). This study measured the proportion of male to female postdoctoral researchers (postdocs, who are research trainees under the direction of senior professors of a laboratory) in labs run by senior professors. The results showed that there was a much larger percentage of men than women postdocs in the labs investigated.
These results provide clear evidence of women being discriminated against in academia because of their sex, a property that should be irrelevant to their fitness for scientific research.

Methods
In order to examine the gender distribution of biomedical scientists in academia, we collected information on professors, and postdocs, employed in 39 departments at 24 of the highest-ranked research institutions in the United States. We focused on departments that study molecular biology, cell biology, biochemistry, and/or genetics.
In total, we obtained information on 2,062 professors and 4,904 postdocs in the life sciences. We examined the proportions of male vs. female postdocs across those labs.

Results



On average, laboratories comprised significantly fewer female postdocs than male postdocs.

Conclusion
The study provides clear evidence of women being discriminated against in academia because of their sex, a characteristic that should not matter for research work.

Appendix B. Results of statistical analyses for Experiment 3

B1. Results of linear regressions for Experiment 3

DV	IV	*β*	SE	t	*p*
Accuracy	Bias (favoring)	-0.77	0.08	-9.49	**<.0001**
	Moral commitment	0.10	0.04	2.53	**0.01**
	Priors	0.01	0.006	2.03	**0.04**
	Order (com->task)	-0.15	0.08	-1.82	0.07
	Bias * Moral commitment	-0.13	0.04	-3.31	**0.001**
	Bias * Priors	-0.02	0.01	-3.72	**<.001**
	Moral commitment * Priors	0.005	0.003	1.81	0.07
	Bias * Order	-0.19	0.08	-2.38	**0.02**
	Moral commitment * Order	0.08	0.04	1.98	**0.05**
Reliability	Bias (favoring)	-0.51	0.08	-6.27	**<.0001**
	Moral commitment	0.11	0.04	2.74	**0.006**
	Sex (male)	0.12	0.08	1.36	0.17
	Order (com -> task)	-0.19	0.08	-2.40	**0.02**
	Bias * Moral commitment	-0.10	0.04	-2.33	**0.02**
	Bias * Sex	0.16	0.08	1.92	0.06
	Bias * Order	-0.22	0.08	-2.73	**0.007**

B2. Results of mediation analysis with "prior beliefs" as the mediator for Experiment 3

DV	Statistic	Estimate	95% CI Lower	95% CI Upper	*p*
Accuracy	ACME	0.045	0.017	0.08	**<.0001**
	ADE	0.236	0.132	0.34	**0.002**
	Total Effect	0.281	0.170	0.39	**<.0001**
	Prop. Mediated	0.160	0.068	0.29	**<.0001**
Reliability	ACME	0.036	0.013	0.06	**0.002**
	ADE	0.183	0.055	0.31	**<.0001**
	Total Effect	0.219	0.083	0.34	**<.0001**
	Prop. Mediated	0.164	0.062	0.41	**0.002**

B3. Results of mediation analysis with “moral commitment” as the mediator for Experiment 3

DV	Statistic	Estimate	95% CI Lower	95% CI Upper	*p*
Accuracy	ACME	0.007	0.002	0.01	**0.002**
	ADE	0.035	0.020	0.05	**<.0001**
	Total Effect	0.043	0.028	0.06	**<.0001**
	Prop. Mediated	0.170	0.050	0.37	**0.002**
Reliability	ACME	0.006	0.0009	0.01	**0.004**
	ADE	0.029	0.015	0.04	**<.0001**
	Total Effect	0.034	0.021	0.05	**<.0001**
	Prop. Mediated	0.165	0.031	0.39	**0.004**

Chapter 5. Conclusion

The natural, linguistic and social facets of gender are intertwined, and together they play an important role in shaping the structure of languages, the social norms and role beliefs about men and women, as well as people's moral attitudes and ideologies. As a linguistic feature, grammatical gender has been argued to have consequences for the mental representations of objects and persons. A wealth of research efforts has been devoted to the question of whether the classification of nouns denoting objects into masculine and feminine gender classes influences language users' conceptualizations of objects. At the same time, the unequal treatment of the masculine and feminine grammatical gender across languages has been questioned as a source and reflection of gender stereotypes and gender role beliefs that contribute to the status quo of gender inequality. Furthermore, people's ideologies and attitudes built on gender stereotypes and social role beliefs show impacts on their perception of gender inequality and their trust in research on gender bias.

In this concluding chapter, I first summarize findings of the three empirical studies presented in previous chapters. Then I discuss the theoretical and empirical implications of the three studies. I will end the chapter by pointing out the limitations of the research and directions for future investigations.

Summary

The first study (Chapter 2) investigates the relationship between language and thought from the angle of grammatical gender, as an attempt to test the Neo-Whorfian hypothesis that language structure influences people's conceptualizations of the world. In this study, I ask if the grammatical gender system of a language has consequences for the speakers' mental

representations of objects such that asexual objects are conceived of possessing gender qualities. Put in another way, I question if the assignment of object nouns to the masculine gender class would make speakers associate these objects with male qualities and vice versa for nouns assigned to the feminine gender class. The two psycholinguistic experiments test the hypothesized language effect within French (Experiment 1) and cross-linguistically between French and German (Experiment 2). As the article in which this study is described has been submitted as a registered report for peer review, the experiments have not been conducted at the moment. Here, I mainly present a summary of the experimental design and results of two pilot experiments.

Conflicting results have been reported in previous studies using various paradigms (see Bassetti, 2007; Beller et al., 2015; Bender et al., 2011, 2016a; Boroditsky et al., 2003; Boutonnet et al., 2012; Cubelli et al., 2011; Kousta et al., 2008; Mickan et al., 2014; Sato & Athanasopoulos, 2018). Some of them employed explicit methods that might have led to participants' use of response strategies, a result of task demands (e.g. Clarke et al., 1981; Sera et al., 1994, 2002). To avoid this happening, we employ a word association paradigm built on the work of Boroditsky et al. (2003). In addition, to minimize an experimenter bias in item selection, we ask participants to create the materials for Experiment 1. Based on extensive piloting work (see Pilot Studies of Chapter 2) as well as the findings reported in Mickan et al. (2014), we assume that any influence of grammatical gender on object representations is either weak or non-existent, and thus we adopt an approach of "stack the deck" *in favor* of the original hypothesis. That is, the experiments are designed in such a way that the odds of detecting any underlying language effects were maximized. To do this, firstly, the prospective participants will be tested in their native language instead of a language with no grammatical gender as did in Boroditsky et al.

(2003). Secondly, we increase the likelihood of detecting a language effect (if there is one to be found) by showing gender marked determiners with the relevant noun items, as making grammatical gender information more salient would again serve to enhance the chance of observing any underlying effects. And lastly, to ensure that any null effect we find is not due to a lack of statistical power, we will run Bayesian analyses that will allow us to quantify our confidence in null or positive results.

Specifically, we will ask native speakers of French and German to produce adjectives for grammatically masculine and feminine nouns in their native language, and then we will ask another group of native participants to assess these adjectives in terms of how likely they can be associated with male and female qualities. In Experiment 1, we will test semantically related French nouns that are assigned opposite grammatical gender in French. Experiment 2 will extend the investigation across languages by testing French and German translation-equivalents that have opposite grammatical gender in the two languages. We will commence data collection once the article gets in-principle acceptance from an academic journal. Our previous pilot results suggested null effects of grammatical gender on how objects are conceived of by native French and German speakers.

The second study (Chapter 3) examined the influence of language forms on the prominence of women in people's mental representations. Focusing on French, we compared the three generic forms: masculine, double-gender and middot. Participants were shown a short text to read, which described the taking place of a professional gathering, and then asked to estimate the proportions of men and women present at the gathering by responding on a slider. Aside from language forms, we manipulated the gender stereotypicality of the professions: gender-neutral (Experiment 1), and male- and female-stereotyped (Experiment 2). The results showed

that compared to the masculine form, both the double-gender and middot forms increased the perceived proportions of women across the three types of professions. The two gender-fair forms did not differ from each other in terms of increasing women's presence in mental representations. The observed language effects were not moderated by the gender stereotype of professions. Consistent with previous research on English and German (Braun et al., 1998; Gastil, 1990; Hamilton, 1988; Irmen & Roßberg, 2004; Martyna, 1978), these findings provided further evidence of the impact of gender-fair language in promoting the salience of women in the minds of perceivers.

To establish the consistency of representations, we compared the %-women provided by our participants with real-world estimates obtained in Misersky et al.'s (2013) norming study for each profession. Results of the comparison showed that gender-fair language had varying effects on the consistency of the perceived proportions of women. Both double-gender and middot forms induced consistent representations for gender-balanced occupations; however, they introduced a female bias for male-dominated occupations, demonstrated by the overestimates of %-women compared to the normed data; and finally, for female-dominated professions, the gender-fair forms failed to provide enough boost such that a bias favoring male was still present in participants' estimates. In line with a previous study that investigated only gender-neutral professions in Swedish (Lindqvist et al., 2019), our results regarding the gender-balanced professions added proof of the existence of a male bias induced by the masculine generics in people's mental representations as well as the efficacy of using gender-fair language in removing such a bias.

In the last study (Chapter 4), we looked at the role of morality in people's differing trust in scientific evidence related to gender discrimination. In Experiment 1, we showed participants

a summary of Moss-Racusin et al's (2012) research article that reported discriminatory hiring practice in STEM favoring male applicants, and we asked them to evaluate the quality of the research in terms of how accurate they found the research findings and how reliable they deemed the methods to be. In addition, we asked participants to report their degree of moral commitment to the cause of gender equality, as a way to measure their moral attitudes on the issue. Results revealed a positive effect of moral attitudes such that individuals with higher commitment to gender equity tended to see the research findings as more accurate and the methods as more reliable. Contradicting the results of Handley et al. (2015) that people's sex affected their research evaluations - men rated less favorably than women scientific evidence suggesting discriminatory hiring practices against women in academia – our results did not show any effect of sex.

In Experiments 2 and 3, we extended the investigation to varied scenarios regarding the directionality of the reported hiring bias and the diversity of evidence by using as stimuli multiple research summaries that documented hiring bias *against* vs. *favoring* women. Experiment 3 additionally tested if the effect of moral attitudes could be explained by participants' pre-existing expectations about the likelihood of hiring discrimination against women happening in STEM by asking them to provide predictions of the research results (in terms of how much the hiring committee's evaluations of women are biased). The results of Experiments 2 and 3 were that, overall, the "bias favoring women" summaries received lower ratings of accuracy and reliability compared to the "bias against women" summaries. Prior beliefs and moral commitment had varying effects on people's trust in evidence of a bias *against* and *favoring* women: both prior beliefs and moral attitudes predicted the evaluations of "bias against women" summaries, while neither of the two factors showed any effect on the ratings of

the “bias favoring women” summaries Results of the mediation analyses suggested that moral attitudes and prior beliefs have distinct influences on people’s research evaluations.

Furthermore, Experiment 4 of the study addressed the question of whether moral convictions made people more prone to imprecise inferences when faced with an attitudinal conclusion. Here, we presented participants with a research summary demonstrating the status quo that female post-doctoral research trainees were outnumbered by the male counterparts in research labs of life sciences based on results of the study (Sheltzer & Smith, 2014). (despite being true, these findings cannot serve as statistical evidence of gender discrimination as other factors such as women’s career preferences could contribute to the observed gender imbalance.) However, at the end of the summary, we intendedly made an invalid conclusion that women were being discriminated against in academia because of their sex (based on the smaller proportions of female post-docs in these labs). Participants were asked to judge how much they considered the conclusion as justified by the research results shown in the summary. We found that moral commitment affected people’s judgments such that individuals with higher moralization of gender equality were more inclined to take gender imbalance as solid evidence of sex-based discrimination.

Limited Influences of Language on Mental Representations

In spite of the debates, the Neo-Whorfian hypothesis has gained empirical support from various domains (Boroditsky, 2001; Davies & Corbett, 1997; Gilbert et al., 2006; Haun et al., 2006; Loewenstein & Gentner, 2005; Lupyan et al., 2007; Majid et al., 2004; Thierry et al., 2009). Now the field has seen a shift of focus from questioning the existence of linguistic influences on thought to when it happens, how it operates and what factors account for the strength and durability of such influences (Bender et al., 2018). Earlier research on this topic seems to suggest

that thinking can be affected by language *before*, *during* and *after* language use (Wolff & Holmes, 2011). For example, the *thinking for speaking* (Slobin, 1996, 2003) mechanism was responsible for the *before* effect, as demonstrated in the eye movement differences between speakers of English, German and Greek when they were asked to watch motion events and then to describe them verbally (Gennari et al., 2002; Papafragou et al., 2008). They may sound trivial as it is not surprising that people attend to the aspects of an event for which their language provides ready-to-use expressions when they asked to complete a task of verbal reproduction. The *during* effect is explained by a mechanism of *thinking with language* (Wolff & Holmes, 2011), or actively employing language when performing a task. Here, language serves as a conceptual tool that can facilitate certain mental activities that would be difficult or impossible without language, such as numerical cognition (Dehaene et al., 1999; Gordon, 2004), the understanding of false beliefs (Milligan et al., 2007; Pyers & Senghas, 2009), and category learning (Lupyan et al., 2007). And finally, language can highlight certain aspects of the world by having them regularly encoded in the lexicon and syntax. After a long process of language learning and language use in which people repeatedly practice a certain way of categorizing the world entities, their access to these highlighted aspects are reinforced and even become unavoidable when asked to perform tasks for which language is not required. Such kind of *after* effect is shown in the differing preferences and proficiency between language groups regarding the utilization of absolute, intrinsic and relative spatial frames of reference (Haun et al., 2006; Levinson et al., 2002), and similarly in the differing sensitivity to the distinction between a loose and tight fit between objects (McDonough et al., 2003).

As the literature suggests, the structure of the language one speaks contributes to the diversity in people's conceptualizations and construal of the reality. However, we should be

aware that any linguistic working on the human mind is not irreversible, for that language does not change the fundamental cognitive machinery that human beings evolved to share in common. For example, Korean speakers may be relatively more attentive to the distinction between a loose and tight fit of objects in a containment than speakers of English. But, this is not to say that English speakers are unable to make such a distinction at all (otherwise, one would doubt the stability of all three-dimensional structures constructed by English speakers). Instead, one can view these effects of language as the impacts of habituation. After years and years of experience in a specific way of thinking/problem solving, people become more adept and efficient in analyzing the world from that perspective, being reinforced by the language structure. Thus, when faced with a new similar problem, they are more inclined to approach it from what they regard as the vantage point and find solutions by utilizing their cognitive toolkit that has been proved useful in the past.

Although previous research remains inconclusive with respect to the hypothesized influence of grammatical gender on object cognition (Bassetti, 2007; Beller et al., 2015; Bender et al., 2011, 2016a; Boroditsky et al., 2003; Boutonnet et al., 2012; Cubelli et al., 2011; Haertlé, 2017; Imai et al., 2014; Kousta et al., 2008; Mickan et al., 2014; Saalbach et al., 2012, 2012; Sato & Athanasopoulos, 2018; Sera et al., 2002), taken together, the body of literature suggests that any effects of language on the gendered conceptualizations of objects are likely to be limited. Language may not change human cognitive machinery, but it may tweak our attention and memory in a subtle way. For example, speakers of grammatical gender languages are required to mind the gender class of a noun and the behavior of the words that appear with it. By so doing, they may conditionally assign more similar qualities to nouns of the same gender class, especially when the task prompts them to make associations between these words and when

grammatical gender can be employed to solve the problem at hand (Beller et al., 2015; Clarke et al., 1981; Guiora & Sagi, 1978; Sera et al., 1994, 2002). But if the condition is taken out (e.g. task demands, cultural connotations), the effects of grammatical gender should be largely reduced or removed, as demonstrated in Beller et al. (2015).

As to the relationship between language and mental representations of persons, however, our study in Chapter 2 added consistent evidence of grammatical gender influencing the perceptions of men and women in a group. People perceived fewer women group members when the masculine form was used, where the female gender was made implicit/hidden and thus less accessible in the memory; conversely, gender-fair forms rendered the female gender salient, hence prompting more female associations in the minds of the perceivers. This speaks to the old saying "Out of sight, out of mind". When we think with language, we are likely to have our attention tweaked by grammatical gender that also affects how a certain category in our memory is accessed. Language plays an important role in shaping people's mental representations of gender groups and this influence is of great societal importance, in terms of how women are evaluated and considered for certain professions (Braun et al., 2005; Gabriel et al., 2008; Gygax et al., 2008, 2012; Gygax & Gabriel, 2008; Hansen et al., 2016; Sato et al., 2016; Stahlberg et al., 2001; Stahlberg & Sczesny, 2001; Stout & Dasgupta, 2011; Vervecken & Hannover, 2012), how women's roles are viewed in general (Koniuszaniec & Blaszkowa, 2003; Merkel et al., 2012; Mucchi-Faina, 2005), and how policies related to gender equality is treated (Tavits & Pérez, 2019). In addition to the social impact, gender-fair language also helps to improve the overall consistency of representations people have of gender distributions across professions that are otherwise male-biased when the masculine form is used. As demonstrated in Chapter 3, gender-fair forms removed a male bias in the representations of gender-balanced professional group, and

they reduced some bias favoring males for occupations where female is the majority gender. Although gender-inclusive forms induced a female bias for the male-dominated professions, overall, the perceivers formed more consistent representations after seeing gender-fair forms compared to the masculine form.

Stereotype, Ideology and Truth

Gender stereotypes contributes to the formation of people's gender role beliefs and moral attitudes toward gender equality. Stereotypes such as *women are less competent than men*, *math is not for girls*, *women are emotional and not assertive*, and *women are family-centered* constantly characterize women as intellectual inferiors with lesser career aspirations. Depending on how these gender stereotypes are treated by individuals, different ideologies such as gender essentialism and egalitarianism have been developed (Knight & Brinton, 2017). According to Knight and Brinton (2017), people who hold what is known as the gender-essentialist ideology believe in the existence of innate sex differences between men and women and that these inherent differences account for the differing accomplishments of the two sexes; those embracing a flexible egalitarianism dismiss the idea that women were born less intelligent than men but partially accept the gender stereotypes, thus attributing gender disparities in society to free or constrained personal preferences; beyond these two, there are also advocates of liberal egalitarianism who reject all the traditional gender stereotypes and aspire to a society of gender equality (Knight & Brinton, 2017).

Gender attitudes and ideology influence individuals' perception of inequality and their trust in scientific evidence of discriminatory practices that undermine the advance of gender equity. Individuals who are morally concerned about the well-being of women are more likely to detect any practices or external barriers that discourage women from pursuing academic

attainment. Consistent with the awareness of gender discrimination in society, they are more receptive of research findings confirming the existence of sex-based bias against women. On the contrary, people who do not feel morally engaged in the cause of gender equity are less sensitive to any external biases that may create obstacles in the way to women's success, and in consequence, they tend to discard any information showing the opposite. As demonstrated in Chapter 4, people holding positive attitudes on gender equality find research evidence of gender discrimination against women in STEM fields more credulous than those with negative attitudes on this issue. Accordingly, the individuals who reported moral commitment to gender equity may be advocates of the liberal egalitarian ideology; while the others who were less morally concerned may be proponents of gender essentialism or flexible egalitarianism.

The influence of ideology on individuals' trust calibration can be two-fold. First, if a person's pre-conceptions are built on the accumulation of accurate information, then assigning more trust to new, attitudinal information is not only unproblematic but also efficient, since it would cost us a large amount of time and energy to deliberate over every single piece of new information we encounter. However, if a person's prior beliefs are themselves biased, based on misinformation, as is the case for COVID-19 vaccine opponents (Roozenbeek et al., 2020) and global warming deniers (Zhou & Shen, 2021), attitude-based evaluations of new information will lead them further away from the truth.

Limitations and Directions for Future Research

To establish the cognitive effect of grammatical gender on object conceptualization, future studies can focus on developing more varied, reliable research methods as the existent literature suggests the lack of valid, reproducible paradigms. For instance, the study in Chapter 2 adopted a word association approach, but more innovative research methods would be a welcome addition.

Additionally, future studies can enrich the data on the relationship between grammatical gender and the perceived properties in objects by examining language groups outside the Indo-European family. There are more than a hundred grammatical gender languages around the world that have diversified gender structures and many of them are understudied. Most of the existent literature documented the Indo-European languages that possess two to three gender classes. Future research can extend the investigations to languages with a larger number of gender classes, like the Niger-Congo languages (e.g. Swahili), that may show different patterns of results.

Next, to allow for a more holistic assessment of the costs and benefits of language reform for the society, future studies can provide more empirical evidence that is currently absent. For instance, more empirical data are needed regarding the effects of different gender-fair language forms on the mental representations of gender groups, such as contracted forms with a slash (e.g. *étudiant/es*) or a dash (e.g. *étudiant-e-s*), and the feminine-before-masculine double-gender form (e.g. *étudiantes et étudiants*) that were not covered in our study. Each of these forms may have their own advantages and drawbacks that should be validated with scientific methods.

Additionally, by investigating professions of more varied gender stereotypicality (since our study only tested occupations that are gender-balanced, or extremely male- and female-stereotyped), futures studies can shed more light on the varying influence of language form and its relations with gender stereotype. Another issue that has been recurrently brought up as counterargument for gender-fair language is that it makes language learning more difficult (than it already is with masculine forms) for people with special conditions (e.g. dyslexia). Future studies can help answer this question by investigating language learners with such conditions. In addition to the cognitive effects of language forms, future efforts can also consider the social

impact of gender-fair language in France, such as how women's roles are viewed when gender-fair language is used, how language reform affects the evaluation of women in a hiring process and whether people will be more aware of gender inequality problems.

Finally, about the impact of morality on trust calibration, prospective studies can examine the underlying factors that are accountable for the differing gender attitudes, say one's personality, cultural and socioeconomic background, and political ideology. Our study only revealed that individuals' moral attitudes affected their evaluation of evidence, but we still need to address the question of what leads some individuals but not others to be morally concerned about gender inequality and have strong attitudes on it. Furthermore, as the framing of information may have influences on how it is processed and interpreted by the recipients which in consequence leads to differing reactions, future studies are welcome to investigate what effects the framing of evidence can have on the level of credibility it induces in individuals.

References

Abbou, J. (2011). Double gender marking in French: A linguistic practice of antisexism. *Current Issues in Language Planning*, *12*(1), 55–75.

Aikhenvald, A. Y. (2016). *How gender shapes the world*. Oxford University Press.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141.

Bassetti, B. (2007). Bilingualism and thought: Grammatical gender and concepts of objects in Italian-German bilingual children. *International Journal of Bilingualism*, *11*(3), 251–273.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1).

Begeny, C. T., Ryan, M. K., Moss-Racusin, C. A., & Ravetz, G. (2020). In some professions, women have become well represented, yet gender bias persists—Perpetuated by those who think it is not happening. *Science Advances*, *6*(26), eaba7814.

Beller, S., Brattebø, K. F., Lavik, K. O., Reigstad, R. D., & Bender, A. (2015). Culture or language: What drives effects of grammatical gender? *Cognitive Linguistics*, *26*(2), 331–359.

Bem, S. L. (1974). The Measurement of Psychological Androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162.

Bem, S. L., & Bem, D. J. (1973). Does Sex-biased Job Advertising "Aid and Abet" Sex Discrimination? 1. *Journal of Applied Social Psychology*, *3*(1), 6–18.

Bender, A., Beller, S., & Klauer, K. C. (2011). Grammatical gender in German: A case for linguistic relativity? *The Quarterly Journal of Experimental Psychology*, *64*(9), 1821–1835.

Bender, A., Beller, S., & Klauer, K. C. (2016a). Crossing grammar and biology for gender categorisations: Investigating the gender congruency effect in generic nouns for animates. *Journal of Cognitive Psychology*, *28*(5), 530–558.

Bender, A., Beller, S., & Klauer, K. C. (2016b). Lady Liberty and Godfather Death as candidates for linguistic relativity? Scrutinizing the gender congruency effect on personified allegories with explicit and implicit measures. *The Quarterly Journal of Experimental Psychology*, *69*(1), 48–64.

Bender, A., Beller, S., & Klauer, K. C. (2018). Gender congruency from a neutral point of view: The roles of gender classes and conceptual connotations. *0278-7393*. https://doi.org/10.1037/xlm0000534

Ben-Shachar, M. S., Makowski, D., & Lüdecke, D. (2020). *Compute and interpret indices of effect size*.

Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review*, *55*(1), 3–21.

Blaubergs, M. S. (1980). An analysis of classic arguments against changing sexist language. *Women's Studies International Quarterly*, *3*(2–3), 135–147.

Bodine, A. (1975). Androcentrism in prescriptive grammar: Singular 'they', sex-indefinite 'he', and 'he or she'1. *Language in Society*, *4*(2), 129–146.

Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. *Language in Mind: Advances in the Study of Language and Thought*, 61–79.

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research*, *1479*, 72–79.

Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience*, *35*(25), 9329–9335.

Brauer, M., & Landry, M. (2008). Un ministre peut-il tomber enceinte? L'impact du générique masculin sur les représentations mentales. *L'Année psychologique, 108, 243–272*.

Braun, F., Gottburgsen, A., Sczesny, S., & Stahlberg, D. (1998). Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift Für Germanistische Linguistik*, *26*(3), 177–195.

Braun, F., Sczesny, S., & Stahlberg, D. (2005). Cognitive effects of masculine generics in German: An overview of empirical findings. *Communications*, *30*, 121.

Breda, T., & Hillion, M. (2016). Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science*, *353*(6298), 474–478. https://doi.org/10.1126/science.aaf4372

Breda, T., Jouini, E., Napp, C., & Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences*, *117*(49), 31063–31069.

Breda, T., & Ly, S. T. (2015). Professors in core science fields are not always biased against women: Evidence from France. *American Economic Journal: Applied Economics*, *7*(4), 53–75.

Breda, T., & Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences*, *116*(31), 15435–15440.

Budziszewska, M., Hansen, K., & Bilewicz, M. (2014). Backlash over gender-fair language: The impact of feminine job titles on men's and women's perception of women. *Journal of Language and Social Psychology*, *33*(6), 681–691.

Burnett, H., & Pozniak, C. (2020). Political dimensions of écriture inclusive in Parisian universities. *Manuscript. Université de Paris, LLF, CNRS*.

Bußmann, H., & Hellinger, M. (2003). Engendering female visibility in German. In *Gender across languages* (pp. 141–174). John Benjamins.

Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in Italian role nouns. *Applied Psycholinguistics*, *28*(2), 277–293.

Campbell, T. H., & Kay, A. C. (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology*, *107*(5), 809–824.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In *Eye movements in reading* (pp. 275–307). Elsevier.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A*, *49*(3), 639–663.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, *108*(8), 3157–3162.

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

Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, *114*(4), 924–976.

Charlesworth, T. E., & Banaji, M. R. (2019). Gender in science, technology, engineering, and mathematics: Issues, causes, solutions. *Journal of Neuroscience*, *39*(37), 7228–7243.

Charlesworth, T. E. S., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender Stereotypes in Natural Language: Word Embeddings Show Robust Consistency Across Child and Adult Language Corpora of More Than 65 Million Words. *Psychological Science*, 095679762096361. https://doi.org/10.1177/0956797620963619

Chatard, A., Guimont, S., & Martinot, D. (2005). Impact de la féminisation lexicale des professions sur l'auto-efficacité des élèves: Une remise en cause de l'universalisme masculin? *L'année Psychologique*, *105*(2), 249–272.

Clarke, M. A., Losoff, A., McCracken, M. D., & Still, J. (1981). Gender perception in Arabic and English. *Language Learning*, *31*(1), 159–169.

Corbett, G. G. (1991). *Gender*. Cambridge University Press.

Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The Effect of Grammatical Gender on Object Categorization. *Learning, Memory*, *37*(2), 449–460.

Danbold, F., & Huo, Y. J. (2017). Men's defense of their prototypicality undermines the success of women in STEM initiatives. *Journal of Experimental Social Psychology*, *72*, 57–66.

Davies, I. R., & Corbett, G. G. (1997). A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British Journal of Psychology*, *88*(3), 493–517.

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review*, *94*(3), 369–389.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*(5416), 970–974.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568–584.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, *15*(4), 543–558.

Eagly, A. H., & Wood, W. (2016). Social Role Theory of Sex Differences. In *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The Developmental Social Psychology of Gender*, *12*, 174.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*(1), 5–24. https://doi.org/10.1037/0022-3514.71.1.5

Elmiger, D. (2008). *La féminisation de la langue en français et en allemand: Querelle entre spécialistes et réception par le grand public* (Vol. 30). Honoré Champion.

Flaherty, M. (2001). How a language gender system creeps into perception. *Journal of Cross-Cultural Psychology*, *32*(1), 18–31.

Formanowicz, M., Bedynska, S., Cis\lak, A., Braun, F., & Sczesny, S. (2013). Side effects of gender-fair language: How feminine job titles influence the evaluation of female applicants. *European Journal of Social Psychology*, *43*(1), 62–71.

Formanowicz, M. M., Cis\lak, A., Horvath, L. K., & Sczesny, S. (2015). Capturing socially motivated linguistic change: How the use of gender-fair language affects support for social initiatives in Austria and Poland. *Frontiers in Psychology*, *6*, 1617–1617.

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression* (Third). Sage.

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods*, *40*(1), 206–212.

Gabriel, U., & Mellenberger, F. (2004). Exchanging the generic masculine for gender-balanced forms-The impact of context valence. *Swiss Journal of Psychology*, *63*(4), 273–278.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Garnham, A., Doehren, S., & Gygax, P. (2015). True gender ratios and stereotype rating norms. *Frontiers in Psychology*, *6*, 1023–1023.

Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*, *30*(3), 439–446.

Gastil, J. (1990). Generic pronouns and sexist language: The oxymoronic character of masculine generics. *Sex Roles*, *23*(11), 629–643.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*(1), 49–79.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494.

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.

Glick, P., Fiske, S. T., Mladinic, A., Saiz, J. L., Abrams, D., Masser, B., Adetoun, B., Osagie, J. E., Akande, A., & Alao, A. (2000). Beyond Prejudice as Simple Antipathy: Hostile and Benevolent Sexism Across Cultures. *Journal of Personality and Social Psychology*, *79*(5), 763–775.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*(5695), 496–499.

Greene, K., & Rubin, D. L. (1991). Effects of gender inclusive/exclusive language in religious discourse. *Journal of Language and Social Psychology*, *10*(2), 81–98.

Guiora, A. Z., & Sagi, A. (1978). A Cross-Cultural Study of Symbolic Meaning-Developmental Aspects. *Language Learning*, *28*(2), 381–386.

Gustafsson Sendén, M., Bäck, E. A., & Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: The influence of time on attitudes and behavior. *Frontiers in Psychology*, *6*, 893.

Gygax, P., Elmiger, D., Zufferey, S., Garnham, A., Sczesny, S., von Stockhausen, L., Braun, F., & Oakhill, J. (2019). A Language Index of Grammatical Gender Dimensions to Study the Impact of Grammatical Gender on the Way We Perceive Women and Men. *Frontiers in Psychology*, *10*, 1604–1604.

Gygax, P., & Gabriel, U. (2008). Can a group of musicians be composed of women? Generic interpretation of French masculine role names in the absence and presence of feminine forms. *Swiss Journal of Psychology*, *67*(3), 143–151.

Gygax, P., Gabriel, U., Lévy, A., Pool, E., Grivel, M., & Pedrazzini, E. (2012). The masculine form and its competing interpretations in French: When linking grammatically masculine role names to female referents is difficult. *Journal of Cognitive Psychology*, *24*(4), 395–408.

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, *23*(3), 464–485.

Gygax, P., & Gesto, N. (2007). Féminisation et lourdeur de texte. *L'Année Psychologique*, *107*(2), 239–255.

Haertlé, I. (2017). Does grammatical gender influence perception? A study of Polish and French speakers. *Psychology of Language and Communication*, *21*(1), 386–407.

Hamilton, M. C. (1988). Using masculine generics: Does generic he increase male bias in the user's imagery? *Sex Roles*, *19*(11–12), 785–799.

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, K., Littwitz, C., & Sczesny, S. (2016). The Social Perception of Heroes and Murderers: Effects of Gender-Inclusive Language in Media Reports. *Frontiers in Psychology*, *7*, 369–369.

Haun, D. B., Rapold, C. J., Call, J., Janzen, G., & Levinson, S. C. (2006). Cognitive cladistics and cultural override in Hominid spatial cognition. *Proceedings of the National Academy of Sciences*, *103*(46), 17568–17573.

Healy, A. F. (1994). Letter detection: A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin & Review*, *1*(3), 333–344.

Hegarty, P., Watson, N., Fletcher, L., & McQueen, G. (2011). When gentlemen are first and ladies are last: Effects of gender stereotypes on the order of romantic partners' names. *British Journal of Social Psychology*, *50*(1), 21–35.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*(4), 657–674.

Heilman, M. E., Block, C. J., & Martell, R. F. (1995). Gender in the workplace. Sex stereotypes: Do they influence perceptions of managers? *Journal of Social Behavior and Personality*, *10*, 237–252.

Hogg, M. A., & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes.* Taylor & Frances/Routledge.

Horvath, L. K., Merkel, E. F., Maass, A., & Sczesny, S. (2016a). Does gender-fair language pay off? The social perception of professions from a cross-linguistic perspective. *Frontiers in Psychology*.

Horvath, L. K., Merkel, E. F., Maass, A., & Sczesny, S. (2016b). Does gender-fair language pay off? The social perception of professions from a cross-linguistic perspective. *Frontiers in Psychology*, *6*, 1–12.

Horvath, L. K., & Sczesny, S. (2016). Reducing women's lack of fit with leadership positions? Effects of the wording of job advertisements. *European Journal of Work and Organizational Psychology*, *25*(2), 316–328.

Hyde, J. (1984). Children's understanding of sexist language. *Developmental Psychology*, *20*(4), 697–706.

Imai, M., Schalk, L., Saalbach, H., & Okada, H. (2014). All giraffes have female-specific properties: Influence of grammatical gender on deductive reasoning about sex-specific properties in german speakers. *Cognitive Science*, *38*(3), 514–536.

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, *36*(6), 431–456.

Irmen, L., & Roßberg, N. (2004). Gender markedness of language: The impact of grammatical and nonlinguistic information on the mental representation of person information. *Journal of Language and Social Psychology*, *23*(3), 272–307.

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.

Kennison, S. M., & Trofe, J. L. (2003). Comprehending Pronouns: A Role for Word-Specific Gender Stereotype Information. *Journal of Psycholinguistic Research*, *32*(3), 355–378.

Kesebir, S. (2017). Word Order Denotes Relevance Differences: The Case of Conjoined Phrases With Lexical Gender. *Journal of Personality and Social Psychology*, *113*(2), 262–279.

Knight, C. R., & Brinton, M. C. (2017). One egalitarianism or several? Two decades of gender-role attitude change in Europe. *American Journal of Sociology*, *122*(5), 1485–1532.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*(1), 128–161.

Kollmayer, M., Pfaffel, A., Schober, B., & Brandt, L. (2018). Breaking away from the male stereotype of a specialist: Gendered language affects performance in a thinking task. *Frontiers in Psychology*, *9*, 1–10.

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, *22*(5), 519–534.

Koniuszaniec, G., & Blaszkowa, H. (2003). Language and gender in Polish. In *Gender across languages* (Vol. 3, pp. 259–285). Benjamins.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2008). Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 843.

Kruppa, A., Fenn, J., & Ferstl, E. (2021). *Does the Asterisk in Gender-fair Word Forms in German Impede Readability? Evidence from a Lexical Decision Task.* Albert-Ludwigs University. https://amlap2021.github.io/program/90.pdf

Kunda, Z. (1987). Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories. *Journal of Personality and Social Psychology*, *53*(4), 636–647.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Kurinski, E., & Sera, M. D. (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism: Language and Cognition*, *14*(2), 203–220.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.4.4) [Computer software]. https://CRAN.R-project.org/package=emmeans

Levinson, S. C. (2003). Space in language and cognition: Explorations in cognitive diversity. *Language, Culture and Cognition; 5*.

Levinson, S. C., Kita, S., Haun, D. B., & Rasch, B. H. (2002). Returning the tables: Language affects spatial reasoning. *Cognition*, *84*(2), 155–188.

Lewandowsky, S., & Oberauer, K. (2021). Worldview-motivated rejection of science and the norms of science. *Cognition*, *215*, 104820.

Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, 1–8.

Li, P., Abarbanell, L., Gleitman, L., & Papafragou, A. (2011). Spatial reasoning in tenejapan mayans. *Cognition*, *120*(1), 33–53.

Liberman, A., & Chaiken, S. (1992). Defensive processing of personally relevant health messages. *Personality and Social Psychology Bulletin*, *18*(6), 669–679.

Lindqvist, A., Renström, E. A., & Sendén, M. G. (2019). Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies. *Sex Roles*, *81*(1), 109–117.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). *Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence*. 12.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083.

Maciuszek, J., Polak, M., & Świa, T. N. (2019). Grammatical Gender Influences Semantic Categorization and Implicit Cognition in Polish. *Frontiers in Psychology*, *10*, 2208.

MacKay, D. (1980). Psychology, prescriptive grammar, and the pronoun problem. *American Psychologist*, *35*(5), 444–449.

Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *TRENDS in Cognitive Sciences*, *8*(3).

Martyna, W. (1978). What does ‘he’mean? Use of the generic masculine. *Journal of Communication*, *28*(1), 131–138.

Martyna, W. (1980). Beyond the" he/man" approach: The case for nonsexist language. *Signs: Journal of Women in Culture and Society*, *5*(3), 482–493.

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, *46*(3), 229–259.

Menegatti, M., & Rubini, M. (2018). Gender bias and sexism in language. In *The Oxford Encyclopedia of Intergroup Communication*. Oxford University Press.

Merkel, E., Maass, A., & Frommelt, L. (2012). Shielding women against status loss: The masculine form and its alternatives in the Italian language. *Journal of Language and Social Psychology*, *31*(3), 311–320.

Mickan, A., Schiefke, M., & Stefanowitsch, A. (2014). Key is a llave is a Schlüssel: A failure to replicate an experiment from Boroditsky et al. 2003. *Yearbook of the German Cognitive Linguistics Association*, *2*(1), 39–50.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646.

Mills, A. E. (1986). *The acquisition of gender: A study of English and German* (Vol. 20). Springer Science & Business Media.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T., Englund, K., Hanulikova, A., & Oettl, A. (2013). Norms on the gender perception of role nouns in Czech,

English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, *46*(3), 841–871.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012a). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012b). 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.

Moulton, J., Robinson, G., & Elias, C. (1978). Sex bias in language use:"Neutral" pronouns that aren't. *American Psychologist*, *33*(11), 1032–1036.

Mucchi-Faina, A. (2005). Visible or influential? Language reforms and gender (in) equality. *Social Science Information*, *44*(1), 189–215.

Mullen, M. K. (1990). Children's classifications of nature and artifact pictures into female and male categories. *Sex Roles*, *23*(9), 577–587.

Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition*, *81*(3), 171–208.

Napier, J. L., Thorisdottir, H., & Jost, J. T. (2010). The joy of sexism? A multinational investigation of hostile and benevolent justifications for gender inequality and their relations to subjective well-being. *Sex Roles*, *62*(7–8), 405–419.

Ng, S. H. (2007). Language-based discrimination: Blatant and subtle forms. *Journal of Language and Social Psychology*, *26*(2), 106–122.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.

Nisbet, E. C., Cooper, K. E., & Garrett, R. K. (2015). The partisan brain: How dissonant science messages lead conservatives and liberals to (dis) trust science. *The ANNALS of the American Academy of Political and Social Science*, *658*(1), 36–66.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155–184.

Parks, J. B., & Roberton, M. A. (1998). Contemporary arguments against nonsexist language: Blaubergs (1980) revisited. *Sex Roles*, *39*(5), 445–461.

Pennycook, G., McPhetres, J., Bago, B., & Rand, D. G. (2021). *Beliefs about COVID-19 in Canada, the UK, and the USA: A novel test of political polarization and motivated reasoning*.

Prewitt-Freilino, J. L., Caswell, T. A., & Laakso, E. K. (2012). The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles*, *66*(3), 268–281.

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological Science*, *20*(7), 805–812.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramos, S., & Roberson, D. (2011). What constrains grammatical gender effects on semantic judgements? Evidence from Portuguese. *Journal of Cognitive Psychology*, *23*(1), 102–111.

Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., & Huguet, P. (2019a). Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature Human Behaviour*, *3*(11), 1171–1179.

Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., & Huguet, P. (2019b). Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature Human Behaviour*, *3*(11), 1171–1179.

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.

Reynolds, D. J., Garnham, A., & Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *The Quarterly Journal of Experimental Psychology*, *59*(05), 886–903.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., Van Der Bles, A. M., & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10), 201199.

RStudio Team. (2020). *RStudio: Integrated Development Environment for R* (1.3.959) [Computer software]. RStudio, PBC. http://www.rstudio.com/

Rubin, D. L., Greene, K., & Schneider, D. (1994). Adopting gender-inclusive language reforms: Diachronic and synchronic variation. *Journal of Language and Social Psychology*, *13*(2), 91–114.

Rutjens, B. T., Sutton, R. M., & van der Lee, R. (2018). Not all skepticism is equal: Exploring the ideological antecedents of science acceptance and rejection. *Personality and Social Psychology Bulletin*, *44*(3), 384–405.

Saalbach, H., Imai, M., & Schalk, L. (2012). Grammatical Gender and Inferences About Biological Properties in German-Speaking Children. *Cognitive Science*, *36*, 1251–1267.

Saint-Aubin, J., & Poirier, M. (1997). The influence of word function in the missing-letter effect: Further evidence from French. *Memory & Cognition*, *25*(5), 666–676.

Sarrasin, O., Gabriel, U., & Gygax, P. (2012). Sexism and attitudes toward gender-neutral language. *Swiss Journal of Psychology*.

Sato, S., & Athanasopoulos, P. (2018). Grammatical gender affects gender perception: Evidence for the structural-feedback hypothesis. *Cognition*, *176*, 220–231.

Sato, S., Gabriel, U., & Gygax, P. M. (2016). Altering male-dominant representations: A study on nominalized adjectives and participles in first and second language German. *Journal of Language and Social Psychology*, *35*(6), 667–685.

Schindler, R. M. (1981). Error in proofreading: Evidence of syntactic control of letter processing? *Journal of Experimental Psychology: Human Perception and Performance*, *7*(3), 573–579.

Sczesny, S., Formanowicz, M. M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, *7*(25), 1–11.

Sera, M. D., Berge, C. A., & del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*(3), 261–292.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, *131*(3), 377–397. https://doi.org/10.1037//0096-3445.131.3.377

Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences*, *111*(28), 10107–10112.

Shen, H. (2013). Inequality Quantified: Mind the Gender Gap. *Nature*, *495*(7439), 22–24.

Skitka, L. J. (2010). The Psychology of Moral Conviction. *Social and Personality Psychology Compass*, *4*(4), 267–281. https://doi.org/10.1111/j.1751-9004.2010.00254.x

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral Conviction: Another Contributor to Attitude Strength or Something More? *Journal of Personality and Social Psychology*, *88*(6), 895–917. https://doi.org/10.1037/0022-3514.88.6.895

Slobin, D. I. (1996). *From" thought and language" to" thinking for speaking."* In Gumperz & Levinson (Eds.), Rethinking linguistic relativity (pp. 70 -96). Cambridge University Press.

Slobin, D. I. (2003). "*Language and thought online: Cognitive consequences of linguistic relativity." Language in mind: Advances in the study of language and thought.*

Soarea, R., Bartkiewicz, M. J., Mulligan-Ferry, L., Fendler, E., & Kun, E. W. C. (2013). *2013 Catalyst census Fortune 500 women executive officers and top earners.* New York, NY: Catalyst.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. *Social Communication*, 163–187.

Stahlberg, D., & Sczesny, S. (2001). Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen. *Psychologische Rundschau*, *52*(3), 131–140.

Stahlberg, D., Sczesny, S., & Braun, F. (2001). Name your favorite musician: Effects of masculine generics and of their alternatives in German. *Journal of Language and Social Psychology*, *20*(4), 464–469.

Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, *26*(1), 340–346.

Stout, J. G., & Dasgupta, N. (2011). When he doesn't mean you: Gender-exclusive language as ostracism. *Personality and Social Psychology Bulletin*, *37*(6), 757–769.

Strickland, B., & Suben, A. (2012). Experimenter Philosophy: The Problem of Experimenter Bias in Experimental Philosophy. *Review of Philosophy and Psychology*, *3*(3), 457–467.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769.

Tachihara, K., & Goldberg, A. E. (2020). Cognitive accessibility predicts word order of couples' names in English and Japanese. *Cognitive Linguistics*, *31*(2), 231–249.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science*, *1*(S1), 173–191.

Tavits, M., & Pérez, E. O. (2019). Language influences mass opinion toward gender and LGBT equality. *Proceedings of the National Academy of Sciences*, *116*(34), 16781–16786.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*(11), 4567–4570.

van der Lee, R., & Ellemers, N. (2018). Perceptions of gender inequality in academia: Reluctance to let go of individual merit ideology. In *Belief Systems and the Perception of Reality* (pp. 63–78). Routledge.

Vervecken, D., Gygax, P. M., Gabriel, U., Guillod, M., & Hannover, B. (2015). Warm-hearted businessmen, competitive housewives? Effects of gender-fair language on adolescents' perceptions of occupations. *Frontiers in Psychology*, *6*, 1437–1437.

Vervecken, D., & Hannover, B. (2012). Ambassadors of gender equality? How use of pair forms versus masculines as generics impacts perception of the speaker. *European Journal of Social Psychology*, *42*(6), 754–762.

Vervecken, D., & Hannover, B. (2015). Yes I Can!: Effects of Gender Fair Job Descriptions on Children's Perceptions of Job Status, Job Difficulty, and Vocational Self-Efficacy. *Social Psychology*, *46*(2), 76–92.

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General*, *134*(4), 501.

Washburn, A. N., & Skitka, L. J. (2018). Science denial across the political divide: Liberals and conservatives are similarly motivated to deny attitude-inconsistent science. *Social Psychological and Personality Science*, *9*(8), 972–980.

Whorf, B. L. (1956). *Language, thought and reality. Selected writings of Benjamin Lee Whorf*. MIT Press.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences*, *112*(17), 5360–5365.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265.

Zhou, Y., & Shen, L. (2021). Confirmation Bias and the Persistence of Misinformation on Climate Change. *Communication Research*, 00936502211028049.

Zlatev, J., & Blomberg, J. (2015). Language may indeed influence thought. *Frontiers in Psychology*, *6*, 1631–1631.

RÉSUMÉ

Les multiples aspects du genre jouent un rôle important dans le façonnement des différentes cultures. Les humains sont classés en hommes ou en femmes selon leur sexe biologique ; les langues humaines diffèrent quant à la manière dont le genre est codé dans la structure du langage ; et dans la société, il existe différentes idéologies de genre concernant les rôles et les positions que les hommes et les femmes devraient occuper. Les relations entre ces différentes facettes sont souvent entrelacées. Dans cette thèse, j'étudie d'abord la relation entre le langage et les représentations mentales du genre (chapitres 2 et 3). En particulier, je cherche à savoir si l'attribution d'un genre grammatical, masculin ou féminin, aux noms représentant des objets inanimés amènerait les locuteurs natifs à considérer ces objets comme ayant des qualités " masculines " ou " féminines ". Un tel effet est postulé par l'hypothèse Néo-Whorfienne selon laquelle les catégories linguistiques affectent la représentation des entités du monde par les humains. Des travaux pilotes approfondis sur ce sujet suggèrent l'absence d'effets du genre grammatical sur la conceptualisation des objets par les locuteurs. Contrairement aux noms représentant des objets, le genre grammatical des noms liés aux personnes est significatif en ce sens qu'il a un fondement sémantique (c'est-à-dire homme – masculin ; femme - féminin). J'examine les influences du genre grammatical sur la perception humaine de la répartition hommes-femmes dans diverses professions. On constate que l'utilisation de différentes formes linguistiques induit des associations hommes-femmes distinctes, dont certaines sont cohérentes et d'autres biaisées. Enfin, j'explore la relation entre l'attitude morale des individus en matière d'égalité des sexes - la mesure dans laquelle l'égalité des sexes est considérée comme un impératif moral - et leur confiance dans les preuves scientifiques concluant à l'existence de préjugés sexistes défavorisant les femmes dans le milieu universitaire (chapitre 4). Six études expérimentales montrent que les personnes ayant un plus grand engagement moral envers l'égalité des sexes sont plus réceptives aux recherches révélant une partialité à l'embauche à l'encontre des femmes. Dans l'ensemble, cette thèse démontre que l'encodage du genre dans le langage a un impact sur les représentations mentales des groupes de personnes mais probablement pas sur celles des objets inanimés, et que l'attitude morale des individus à propos du genre influence leurs réactions sur la recherche concernant préjugé sexiste.

MOTS CLÉS

Genre, Langage, Représentation mentale, Préjugé sexiste, Attitudes morales, Confiance en science

ABSTRACT

The various facets of gender play an important role in shaping our cultures. People are categorized into males or females based on their biological sex; human languages differ in how gender is encoded in the language structure; and in society, different gender ideologies exist concerning what roles and positions men and women should occupy. The relationships between these facets are often intertwined. In this dissertation, I first investigate the relationship between language and people's mental representations of gender (Chapters 2 and 3). In particular, I ask if assigning grammatical masculine or feminine gender to nouns denoting inanimate objects would make native speakers think of these objects as having "male" or "female" qualities, a language effect as postulated by the Neo-Whorfian hypothesis that linguistic categories affect people's construal of the world entities. Extensive piloting work on this topic suggests null effects of grammatical gender on speakers' conceptualization of objects. Unlike object nouns, the grammatical gender of person nouns is meaningful in that it has a semantic underpinning (i.e. male – masculine; female - feminine). I then examine the influences of grammatical gender on people's perceptions of male-female distributions across various professions in two experiments, and found that different language forms induce differential male and female associations, some of which are consistent, others biased. Finally, I explore the relationship between individuals' moral attitudes on gender equality – the extent to which gender equality is deemed to be a moral imperative – and their trust in written scientific evidence of hiring bias disfavoring women in academia (Chapter 4). Six experiments show that people of greater moral commitment to gender equality are more receptive of research revealing a hiring bias against females. Overall, the dissertation demonstrates that the encoding of gender in language has impacts on the mental representations of gender groups but likely not on those of inanimate objects, and that individuals' gender attitudes influence their reactions to research on gender bias.

KEYWORDS

Gender, Language, Mental representation, Gender bias, Moral attitudes, Trust in science