





ARTICLE

The masculine bias in fully gendered languages and ways to avoid it: A study on gender neutral forms in Québec and Swiss French

Jonathan Kim^{1*}, Sarah Angst², Pascal Gygax³, Ute Gabriel⁴ and Sandrine Zufferey⁵

¹Department of Psychology, Norwegian University of Science and Technology, Norway, ²Department of Psychology, University of Fribourg, Switzerland, ³Department of Psychology, University of Fribourg, Switzerland, ⁴Department of Psychology, Norwegian University of Science and Technology, Norway and ⁵Institut de Langue et de Littérature françaises, Université de Berne, Switzerland

*Corresponding author: jonathan.kim@ntnu.no.

(Received 02 September 2021; revised 06 April 2022; accepted 11 April 2022; first published online 08 June 2022)

ABSTRACT

The extent to which gender neutral and gendered nouns impact differently upon native French speakers' gender representations was examined through a yes-no forced choice task. Swiss (Experiment 1) and Québec (Experiment 2) French-speaking participants were presented with word pairs composed of a gendered first name (e.g., Thomas) and a role (e.g., doctor), and tasked to indicate whether they believed that [first name] could be one of the [role]. Roles varied according to gender stereotypicality (feminine, masculine, non-stereotyped), and were either in a plural masculine (interpretable as generic) or gender neutral (epicenes and group nouns) form. The results indicated that the use of gender neutral forms of roles avoided a strong male bias found for the masculine forms, and that both gender neutral and masculine forms used equal cognitive resources. Further, stereotype effects associated with both gender-neutral and grammatically masculine forms were quite small (<1%). These results were highly reliable across both Swiss French and Québec speakers. Our study suggests that gender neutral forms are strong alternatives to the use of the masculine form as default value.

1. INTRODUCTION

Language plays an instrumental role in the way that we connect with the world. The linguistic factors inherent in a language, such as its grammatical rules, provide a relatively rigid framework by which we perceive and navigate our social landscapes (Boroditsky, 2011; Fausey & Boroditsky, 2011; Krauss & Chiu, 1997). In grammatical gender languages including French, a common grammatical rule

Author Note: The authors report no known conflicts of interest. The authors received no financial support for the research, authorship, and/or publication of this article.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

is that most words used to refer to generic members of a role are in a grammatically masculine form, regardless of the gender make-up of those holding that role (Gygax *et al.*, 2008). As a result, speakers of grammatical gender languages have been found to be subject to a ‘male bias’, where repeated exposure to the grammatically masculine form, even if intended as ‘generic’, results in all kinds of roles being perceived as more appropriate for men than for woman (Gabriel & Gygax, 2016). This has been found to have societal consequences by, for example, biasing hiring practices in such a manner as to favour men (e.g., Rudinger *et al.*, 2018; Zhao *et al.*, 2018). Motivated by efforts to alleviate such biases, alternatives to the generic use of masculine forms have been suggested. While some studies in German (e.g., Irmen & Roßberg, 2004) have examined potential solutions, research as to whether, and to what extent, some of these alternatives resolve the problems caused by masculine grammatical forms is still very scarce for French. As such, the present study seeks to examine how one alternative, the use of gender-neutral word forms, affects French readers’ gendered associations. More specifically, we aim to empirically evaluate whether the use of *epicenes* and *group nouns*, as gender-neutral word forms in French, may mitigate the male bias associated with the masculine form.

2. MALE BIAS

In French, a grammatical gender language, all nouns and pronouns (even those referring to objects) are assigned a gender and given associated grammatical gender markings. For human-related nouns and pronouns these markings are essential elements of referential gender, and occur in keeping with human physical gender (i.e., female and male). Nouns used to refer to an individual or a mixed gender group without *intending* to reference physical gender are the exact same as nouns used to refer specifically to a male individual or male individuals (i.e., the masculine form).¹ This marking of the *masculine* as *generic* is theorised to be the cause of the male bias (e.g., Gygax & Gabriel, 2008; Lévy *et al.*, 2014; Gygax *et al.*, 2019), with the grammatically masculine form presenting a special case of lexical ambiguity as the appropriate meaning (male specific vs. generic) depends entirely on context. This can be understood through the Activation-Selection Model (Gorfein, 2001; Gorfein & Bubka, 1989; Gorfein *et al.*, 2007).

The Activation-Selection Model postulates that when we hear or read a word, an activation process occurs that selects a set of attributes associated with the word to inform our understanding of the meaning of the word. Each time an attribute is selected to inform understanding of a word, it is theorised to gain ‘weight’, increasing the likelihood of it being automatically selected in the future, *even when it is incorrect to do so*. As roles in their masculine grammatical form are associated with male referents independent of whether it is intended specifically (masculine specific) or generically (masculine-as-generic), Lévy *et al.* (2014)

¹Note that some role nouns in the feminine forms may also act as generics, especially when those role nouns are highly stereotypical. For example, some people may use *les infirmières* [female nurses] to include every person in this profession. However, this genericity is only speculative and informal, and has not yet received any scientific attention.

argued that the ‘male’ attribute linked to a grammatical masculine form over time gains sufficient weight so that the specific meaning is eventually *ceteris paribus* going to reach the threshold of automatic activation. Further, they argued that a mitigation strategy would be to increase the weight of the ‘female’ attribute linked to grammatical masculine forms. To test this idea, they ran an experiment that introduced a semantic bias. In their study, French speaking participants were presented with gendered kinship terms (such as *sœur* [sister] or *père* [father]) paired with grammatically masculine role nouns (e.g. *musiciens*-MSC [musicians]) and instructed to state whether they believed a person represented by the kinship term could be part of the group represented by the roles (i.e., *can a sister be part of a group of musicians?*). In the course of the experiment, the ratio between the number of female kinship-role pairs and male kinship-role pairs was increasingly shifted towards female kinship-role pairs. This increasingly biased presentation lead to participants more easily accepting that a person represented by a female kinship term could be part of the group represented by the grammatically masculine roles, yet without entirely erasing the male bias introduced by the use of the masculine form.

3. ALTERNATIVES TO TRADITIONAL GENDERED LANGUAGE

There are a variety of approaches available when considering alternatives to traditional gendered language. These can be broadly separated into ‘visibility by feminisation’ strategies and ‘degenderisation by neutralisation’ strategies (e.g., Gabriel & Gyax, 2016). Feminisation strategies (sometimes referred to as differentiation strategies) focus on *increasing* the visibility of feminine (and, in some cases, specific non-binary) forms through using gender-specific terms, while degenderisation strategies focus on *decreasing* the visibility of gender-specific forms through using gender-non-specific terms. Although some studies in French have addressed (re)feminisation strategies (i.e., explicitly stating that a group is composed of both women and men; e.g., Chatard et al., 2005; Vervecken et al., 2015), we are not aware of any studies in French that have focused on degenderisation (one study looked at *avoiding gender cues* though, see below). As such, this article seeks to add to the body of knowledge on this topic by examining the potential for degenderisation to mitigate gender effects for French.

In their review of feminisation and degenderisation strategies, Gabriel et al. (2018) called attention to the fact that degenderisation strategies may be susceptible to gender-stereotypical expectations. This is based on the notion that roles can – independent of whether they are presented in a grammatically marked or unmarked form – be semantically associated with a particular gender. Gender-related information conveyed by grammatically marked forms may interfere with gender-stereotypical associations (e.g. *la mécanicienne*-FEM undoubtedly indicates a female mechanic, hence overriding a possible expectation of mechanics as male), whereas for grammatically unmarked forms, gender representations might be solely based on stereotypes. While there is strong evidence from research in non-gendered languages for gender-stereotype

influence (e.g., Pyykkönen *et al.*, 2010, for Finnish), the situation is less clear for gender-neutral alternative forms used in fully gendered languages.

While research on degenderisation in French is relatively lacking, this is a topic that has been more frequently examined in languages such as German (e.g., Irmen, 2007; Irmen & Roßberg, 2004; Sato *et al.*, 2016; Steiger-Loerbroks & von Stockhausen, 2014). For example Irmen (2007), testing unmarked role nouns, found that neutral forms could at least *partly* alleviate the male bias induced by the masculine form. Similarly, Irmen and Roßberg (2004, Exp. 2) found that nominalised forms only partly alleviated the male bias induced by the masculine form, and argued that this sustained male bias was likely due to the idea that *people = male* (Hamilton, 1991; Silveira, 1980).

Other studies examining nominalised forms through different paradigms (e.g., Sato *et al.*, 2016; Steiger-Loerbroks & von Stockhausen, 2014) did show *full* alleviation of the male bias, supporting the possibility that nominalised form are more inclusive. However, while nominalised forms (at the minimum) alleviate male bias in German, they are not available for use in French. As such, other alternatives must be explored, such as the use of feminisation and (of importance to this article) degenderising by neutralisation strategies.

While no known research has explored degenderisation strategies for French, one study (Richy & Burnett, 2021) *did* more generally examine strategies to avoid gender cues. Richy and Burnett (2021) asked participants to judge – in terms of how likely it was that the person presented would be a man or a woman – noun phrases where the grammatical gender was marked (i.e., *le*-MSC *violoniste*-NON *talentueux*-MSC [the talented male violonist]) and noun-phrases where the gender was neutralized by using unmarked adjectives starting with a vowel/(i.e., *l'*-NON *unique*-NON *violoniste*-NON [the unique violonist]). The results showed that, while the nouns' grammatical gender marking drove participants' gender representations when present, in their absence participants' gender representations were driven by the nouns' stereotypical representations. This is very interesting as it supports the idea that gender cues can be mitigated. The strategy used for this specific research cannot be applied in general, as the use of an adjective starting with a vowel between the determiner and noun is clever experimentally, but of minimal use in naturalistic conversation or written text where such adjectives are not always acceptable or available. However, it does offer support for the potential use of degendered forms for alleviating gender bias in French.

Two forms that are both degendered and usable in naturalistic conversations (and therefore are of particular interest for the present study) are (1) the use of *epicenes*, and (2) the use of *group* nouns instead of titles referring to a group's constituents. Both forms are very commonly supported in inclusive language guides (e.g., Viennot, 2018), but have received, to the best of our knowledge, little attention. Epicenes indifferently refer to women, men, and anyone identifying outside of these genders (e.g., in French: *une personne*-FEM [a person]). The use of group nouns refers to the practice of using nouns that avoid gender while indicating an entire group, instead of the traditional gendered nouns that refer to (at least some of the) constituents in the group. For example, instead of using *les informaticiennes*-FEM *et les informaticiens*-MSC [IT specialists], one could say *le service*-MSC *informatique*-MSC [the computing

department], or instead of using *les chanteuses*-FEM *et les chanteurs*-MSC [singers], one could say – when relevant – *la chorale*-FEM [the choir]. The meaning of roles and alternative expressions may not be exactly the same as the former, yet, as pointed out by Gabriel et al. (2018), there is often sufficient semantic overlap to justify the replacement. Both forms, epicene nouns and group nouns, are grammatically gender marked, but in contrast to role nouns their grammatical gender cannot be linked to the actual gender of people they refer to. As such, based on the arguments postulated by Lévy et al. (2014), the grammatical information of these alternative forms should not contribute to ambiguity, meaning that no gender should receive a systematic advantage based on grammatical information.

4. THE PRESENT STUDY

To summarise, previous research in French has repeatedly found evidence for a male bias associated with the generic use of grammatically masculine forms. As a response, alternatives, such as the use of gender-neutral forms, have been suggested, but little is known about whether these alternatives indeed counteract a male bias. The current study seeks to address this knowledge gap by examining the influence of traditional (masculine gendered) and gender-neutral (epicene; group) French word forms on gender representation while choosing a design that takes gender stereotypes as a possible additional biasing source into account.

In line with previous research (e.g., Richy & Burnett, 2021), we expect a male bias if participants are presented with masculine – intended to be interpreted as generic – forms. As the grammatical gender of epicenes and group nouns is not informative of the gender of people referred to by them, Hypothesis 1 is that participants who are presented with epicenes and group nouns (gender-neutral language) will display a weaker male bias than participants who are presented with roles in the masculine form (gendered language).

In line with this diminished effect of grammatical gender, and based on previous research conducted in languages that are not fully gendered (Finnish: Pyykkönen et al., 2010; Norwegian: Gabriel & Gygax, 2008; English: Gygax et al., 2012) Hypothesis 2 is that participants who are presented with roles in epicene and group forms will respond more strongly in keeping with gender stereotypes than participants who are presented with masculine forms.

In this study we used a forced choice task suggested by Kim et al. (2019) in which participants are presented with female and male first names paired with roles and are instructed to press the yes- or the no-button depending on whether they agree or not agree that someone with this name could hold the role shown. Participants' responses (i.e. *yes* or *no*) and response times were recorded. Both measures were intended to evaluate the effort needed to integrate the information provided by the first name and the role, with difficulty heightening the likelihood for negative responses and slower responses, and vice versa for ease of integration heightening the likelihood for positive responses and faster responses. With reference to our first hypothesis, a male bias would show in more positive and

faster responses to role-name pairs with a male first name than with a female first name. With reference to our second hypothesis, a stereotype effect would show in more positive and faster responses to role-name pairs for which the gender-association of the role matches the gender of the first name than for those that mismatch with gender of the first name. We tested our hypotheses across two samples, a Swiss French speaking sample (Experiment 1) and a Canadian French speaking sample (Experiment 2).

EXPERIMENT 1

5. METHOD

5.1. Participants

A total of 121 French-speaking Swiss participants (mean age 21 years [$SD = 2.4$ years]) took part in this experiment. This sample was entirely composed of Psychology students at the University of Fribourg, Switzerland, who received course credit for their participation. In terms of gender distribution, 105 women, 14 men, one non-binary individual and one person who did not wish to state their gender took part. Those who agreed to participate in this research were given a web link to the experiment. General ethical approval was given by University of Fribourg, Switzerland.

5.2. Research Design

The experiment used a 2 (Version: gendered vs. gender-neutral) by 3 (Stereotype: feminine vs. non-stereotyped vs. masculine) by 2 (Name Gender: female vs. male) design, with Version as between-participant factor and Stereotypicality and Name Gender as within-participant factors. The dependent variables were *response* (yes/no) and *positive response times* in a two-alternative (yes-no) forced choice task, based on Kim *et al.* (2019). In this task, participants were instructed to answer, as quickly as possible, whether they believed an individual called [first name] could be part of a group of [role]. These pairings were always presented in the form '[first name] – [role]'. Presentation order was randomised by participant.

5.3. Materials

For both versions of the task (i.e., masculine vs. gender-neutral), stimuli were composed of six first names paired with 36 roles and 36 filler items. Over the course of the experiment, participants in both versions were presented with 360 first name/role pairings, composed of 216 experimental pairings and 144 filler pairings.

Names. Three female (Cloé, Léa, and Sarah) and three male (David, Samuel, and Thomas) first names were selected based on the most frequent and common names given from 1997 to 2002 (the years participants of our samples were born) in French-speaking Switzerland. The data were obtained from the *Office fédéral de la statistique* [Federal office of statistics]. Each first name was paired with all roles, for a total of 216 experimental pairings. The 36 filler items were gender-marked kinship

terms or definitional gender terms (e.g., Fathers, Sisters, Kings; 18 female gender marked, 18 male gender marked) that were selected to prevent the development of a strategy of always answering positively. These items were paired with first names that were both gender congruent (e.g. *David – Kings*) and gender incongruent (e.g., *Léa – Brothers*), to prevent the adoption of a strategy where participants respond positively to roles but always negatively to kinship terms. Each first name was paired with all incongruent filler items, for a total of 108 incongruent first name/filler item pairings, and was paired with six of the congruent filler items, for a total of 36 congruent first name/filler item pairings.

Role Themes. The term ‘role theme’ is used in this article to refer to the *meaning* of a given role. Specifically, each experimental role used in this experiment was presented in two forms (gendered, gender-neutral) with a shared meaning, which can be conceptualised as the ‘theme’ for that role. These forms can also differ greatly in how many individual words are required to convey this meaning. As an example of a role theme, the theme ‘beauticians’ has the gendered form ‘esthéticiens’ and the gender-neutral form ‘les spécialistes des soins de beauté’.

A total of 36 experimental roles (12 feminine stereotyped, 12 masculine stereotyped, and 12 non-stereotyped) were selected based on both Misersky et al. (2014), and on whether the roles selected had a shared role theme with an epicene form or a group noun (see Table 1). Misersky et al. produced stereotypicality ratings between 0 and 1, with 0 representing roles perceived as fully masculine, 1 representing roles perceived as fully feminine, and 0.50 representing roles perceived as non-stereotyped. The French results from Misersky’s study were obtained in Switzerland. For this study, the masculine roles selected had a mean rating of 0.27, while the feminine roles had a mean rating of 0.74, and the non-stereotyped roles had a mean rating of 0.50.

The specific forms used differed between the gendered and gender-neutral versions of the experiment. In the gendered version of the task, the roles were presented in the masculine plural form (e.g. *ingénieurs* [engineers]). In the gender-neutral version, several choices were made to avoid too much repetition in the neutralizing strategy (and risking participants’ attention drawn towards form rather than meaning) and to make sure the forms chosen were quite natural. For *group* nouns, the roles ($N = 20$) were presented in a gender-neutral singular form (e.g., *un groupe-MSc de danse* [a dance group]). Although group nouns are grammaticalized, they do not explicitly refer to a specific gender. For *epicenes*, some roles ($N = 10$) were presented in a gender-neutral plural form, with a noun that could be both masculine or feminine in the singular form (e.g., *une-FEM ou un-MSc spécialiste* [a female or male specialist]), but neutral in the plural form (e.g. *les spécialistes-EPCN en ingénierie* [specialists in engineering]). Corbett (1991) referred to them as *common gender* nouns. The remaining epicene roles ($N = 6$) were presented in a gender-neutral plural form (e.g., *les personnes-FEM* [persons]).² Although those epicenes are grammaticalized, they do not explicitly refer to a specific gender.

²Note that Brauer and Landry (2008) did show that “une personne” (a person) was considered generic, at least more than “un individu” (an individual).

Table 1. Stereotypicality scores for roles, as determined from the findings of Misersky *et al.* (2014)

English Translation	Gendered Title	Gender-Neutral Title	Score
<i>FEMININE</i>			
Beauticians	Esthéticiens	Les spécialistes des soins de beauté	.86
Birth attendants	Assistants maternels	Les spécialistes de l'assistance maternelle	.82
Cashiers	Caissiers	Le personnel de caisse	.75
Dancers	Danseurs	Le groupe de danse	.68
Dieticians	Diététiciens	Les spécialistes en diététique	.75
Dressmakers	Couturiers	Le personnel de couture	.73
Fortune tellers	Diseurs de bonne aventure	Les personnes qui prédisent l'avenir	.76
Hairdressers	Coiffeurs	Le personnel du salon de coiffure	.75
Nurses	Infirmiers	Le corps infirmier	.72
Psychology students	Étudiants en psychologie	La population étudiante en psychologie	.76
Salespersons	Vendeurs	Le personnel de vente	.65
Social workers	Assistants sociaux	Le service social	.69
<i>Mean</i>			.74
<i>MASCULINE</i>			
Bosses	Patrons	La direction	.29
Computer specialists	Informaticiens	Le service informatique	.26
Engineers	Ingénieurs	Les spécialistes en ingénierie	.22
Golfers	Golfeurs	Les membres du club de golf	.24
Physics students	Étudiants en physique	La population étudiante en physique	.23
Pilots	Aviateurs	Les pilotes d'avion	.24
Police officers	Policiers	La police	.28
Politicians	Politiciens	Les personnalités politiques	.31
Spies	Espions	Le service de renseignement	.29
Statisticians	Statisticiens	Les analystes en statistiques	.32
Surgeons	Chirurgiens	Les spécialistes en chirurgie	.30
Technicians	Techniciens	Le service technique	.30
<i>Mean</i>			.27
<i>NON-STEREOTYPED</i>			
Cinema goers	Spectateurs de cinéma	Le public de cinéma	.53
Concert listeners	Auditeurs de concert	Le public d'un concert	.52 [†]
Musicians	Musiciens	L'orchestre	.47
Neighbours	Voisins	Le voisinage	.51

(Continued)

Table 1. (Continued.)

English Translation	Gendered Title	Gender-Neutral Title	Score
Singers	Chanteurs	La chorale	.58
Skiers	Skieurs	Les personnes pratiquant le ski	.43
Spectators	Spectateurs	Le public	.51
Students	Écoliers	Les élèves	.58
Swimmers	Nageurs	Les personnes pratiquant la natation	.44
Teenagers	Adolescents	Les personnes adolescentes	.52
Tennis players	Joueurs de tennis	Les membres du club de tennis	.44
Walkers/Hikers	Promeneurs	Les personnes qui se promènent	.50 [†]
<i>Mean</i>			.50

Note: †From Gabriel et al. (2008).

5.4. Procedure

All text shown to the participant as part of this experiment was presented in French. To ensure that we could test a wider sample, the internet-based instrument PsyToolkit was utilised for data collection (Stoet, 2010; Stoet, 2017). This instrument was selected as it has been found to have a high level of replicability compared to laboratory-based studies (Kim et al., 2019). Participants gave informed consent, answered questions on age, gender, location, handedness and first language, and stated whether they were currently enrolled as university students, before experimental onset. Participants were then randomly assigned to a word form version. Participants were therefore presented with pairs of terms composed of a first name and a role in either the grammatically masculine plural form or in a gender-neutral form.

Participants were instructed to rest their hands on their keyboard so that their index fingers were on the ‘e’ and ‘i’ keys and their thumbs were on the spacebar. They were then instructed to press, as fast as possible, ‘e’ if they did *not* agree that the individual could be a member of the group, or ‘i’ if they *did* agree. For each stimulus presentation, a fixation cross was presented for 200ms before the *first name/role* pairing appeared on screen. After each answer was given, the pairing was replaced with a blank screen for 500ms before the next pairing began. Response time for each stimulus presentation was recorded from the moment that the *first name/role* pairing appeared on screen until the moment the participant pressed either ‘e’ or ‘i’. If a participant failed to respond within 5000ms it was recorded as a non-response, as slower responses may result from overly conscious processes (Cat et al., 2015; Kim et al., 2019), or can be considered as signs of distracted attention (Harjunen et al., 2018). Participants undertook a five-item training phase before undertaking the main experimental phase. For either word form condition, the experiment took between 20 and 30 minutes to complete. After the experiment was finished, participants were asked to guess the purpose of the experiment.

5.5. Data Preparation

Both item-by-participant (i.e., the removal of individual data points from the final dataset on a per participant basis) and by-participant (i.e., the removal of all data points relating to a specific participant from the final dataset) data screening were used, with by-participant screening undertaken prior to item-by-participant screening. By-participant item screening was composed of excluding non-native French speakers and excluding participants with a high error rate. Error rate was calculated based on the percentage of incorrect answers participants gave to the filler items, based on the assumption that the correct answer for congruent name/filler item pairings is 'yes', and for incongruent name/filler item pairings is 'no'. The error rates of the filler items eliciting 'yes' and 'no' responses were examined separately, with deselection following a two-step process. In the first step, participants with error rates that were at or above 50% for either set of filler items were removed (14 participants). In the second step, participants whose error rates were two standard deviations above the mean were removed (zero participants). The decision to require low error rates from the results of the 'yes' and 'no' filler items separately was taken due to the 'no' filler items appearing far more frequently; as such, a participant always answering 'no' to all items would receive an error rate below 50% if all filler items were examined together, but of 100% for the 'yes' filler items while they are examined separately. Thirty participants were deselected due to not being Swiss native French speakers and 14 participants were deselected due to high error rates. Following deselection, the final sample was composed of 77 participants, with a mean age of 20.9 years ($SD = 8.5$ years). This included 67 women, eight men, one non-binary individual, and one individual who did not state their gender. A total of 39 participants undertook the gender-neutral version, while 38 undertook the gendered version. Item-by-participant deselection was conducted based on response times, and in keeping with standard procedures (e.g., Schubert *et al.*, 2013). Responses that were faster than 300ms, or which hit the maximum of 5000ms, were removed (0.63% of the data).

Prior to data analysis, participants' attempts to guess the aim of the experiment were examined. Since participants only responded to one version of the experiment, participants were not expected to include reference to differences between gendered and gender-neutral in their guesses. As such, responses were coded as *correct* if the participant had responded that it was examining gender stereotypes, but *incorrect* if 1) they had responded with other interpretations, or 2) if they did not respond to this question. A total of 75 of the participants in the final sample – which is rather high – did so correctly (38 in the gender-neutral version, 37 in the gendered version), with two participants failing to correctly guess the aim of the study. This indicates that participants were equally likely to guess the aim of the experiment regardless of the word form used.

Data were examined through two forms of linear mixed-effects modelling. Firstly, participants' responses (yes/no) per item were analysed. Secondly, response time for *positive* responses were analysed. Participants' yes/no responses were modelled through generalized linear mixed effects regression, while participants' response times for positive responses were modelled through

linear mixed effects regression. Analysis was conducted through the *lmer* and *glmer* functions of the *lme4* package (Version 1.1-12, Bates et al., 2015) in R (Version 3.3.3). To identify the best model, we tested an initial model that included all fixed factors and their interactions, and had a minimal random structure composed of the best fitting random intercept; this was determined by examining the AIC values of competing models where only one random intercept was included, with the model with the lowest AIC (and therefore best fit) selected as the initial model. We then refined this initial model to find the model of best fit in keeping with Baayen (2008) and Baayen and Milin (2010). Namely, the fixed effects structure was back-fitted (i.e., main and interaction effects found not to improve the model were removed), then the random effects structure was forward-fitted (i.e., random intercepts and slopes found to improve the model were added to it), and finally the fixed effects structure was re-back-fitted (i.e., removing any main or interaction effects that, with the finalisation of the random effects structure, no longer improved the model). As such, all potential main and interaction effects, and all potential random intercepts and slopes, are automatically evaluated and then included or discarded. This was done automatically through the *bfFixefLMER_F*, *ffRanefLMER*, and *fitLMER.fnc* functions of the *lme4* package (version 1.1-27.1). The initial models for both analyses were composed of fixed effects of the experimental factors (Version [neutral vs. gendered], Stereotypicality [feminine vs. non-stereotyped vs. masculine], and Name Gender [female vs. male]) and their interactions, as well as fixed effects of Age and Handedness and, for the analyses of positive response times, Trial Number. The minimal random structure for all models was found to be a random intercept of Participant. The random intercepts of Role Theme, First Name, and random slopes of Name Gender and Stereotypicality by Participant, Role Theme, and First Name, as well as random slopes of Version by Role Theme and First Name, were tested during forward fitting. The potential slope of Version by Participant was excluded from consideration due to Version being a between-participant variable. All other potential fixed effect by random effects slopes were tested. Post-hoc analysis for main effects and interaction effects was done through the *effects()* function of the *effects* package (version 4.2-0; Fox, 2003).

Local effect sizes were estimated through partial omega squared (ω_p^2), obtained through the *omega_sq()* function of the *sjstats* package (Version 0.18.1; Lüdtke, 2018). This is in line with previous research (e.g., Kim et al., 2019; Nakagawa & Schielzeth, 2010; Wagner-Egger & Gygas, 2017).

6. RESULTS

6.1. Yes/No choice responses

The model that best fit participants' responses contained Version, Name Gender, and Stereotypicality as fixed effects, with the random structure composed of a random intercept of Participant and a random slope of Stereotypicality by Participant. The results indicated a medium sized significant main effects of Version, Wald $X^2(1, N = 77) = 5.89, p = 0.015, \omega_p^2 = 0.08$ and a small main effect

Table 2. Three-way interaction between Version, Name Gender, and Stereotypicality for Choice in Experiment 1

Version	Name Gender	Stereotypicality		
		Feminine	Non-Stereotyped	Masculine
Gender-Neutral	Female	99.92[99.77, 99.98]	99.66[99.16, 99.86]	99.41[98.60, 99.75]
	Male	98.91[97.72, 99.48]	99.64[99.11, 99.85]	99.89[99.67, 99.96]
Gendered	Female	93.87[89.08, 96.63]	94.56[89.43, 97.28]	93.59[87.76, 96.75]
	Male	98.79[97.63, 99.38]	99.67[99.25, 99.86]	99.80[99.52, 99.91]

Note. Numbers are presented in percentages. Numbers in brackets are the 95% confidence intervals.

of Name Gender, Wald $X^2(1, N = 77) = 282.06, p < 0.001, \omega_p^2 = 0.03$. These were qualified by a small significant two-way interaction between Version and Name Gender, Wald $X^2(1, N = 77) = 135.35, p < 0.001, \omega_p^2 = 0.03$, which, along with a very small but significant two-way interaction between Stereotypicality and Name Gender, Wald $X^2(2, N = 77) = 79.87, p < 0.001, \omega_p^2 < 0.01$, were qualified by a very small but significant three-way interaction between Version, Stereotypicality, and Name Gender, Wald $X^2(2, N = 77) = 16.12, p < 0.001, \omega_p^2 < 0.01$.

The three-way interaction between Version, Stereotypicality, and Name Gender (Table 2, Figure 1) indicated a male bias in the gendered version that was partially modulated by stereotype such that the difference was highest for gender-congruent, followed by neutral, and finally gender-incongruent name-role pairs, and indicated a potential, but very weak, female bias in the gender-neutral version that was modulated by stereotype in the same manner. More specifically, for the gender-neutral version, participants responded significantly more positively to feminine roles paired with female compared to male names ($MDIFF = 1.02\%$, 95%CI [0.29%, 2.26%]), tended to respond more positively to masculine roles paired with male compared to female names ($MDIFF = 0.48\%$, 95%CI [-0.08%, 1.36%]), and responded no differently for non-stereotyped roles paired with female compared to male names ($MDIFF = 0.02\%$, 95%CI [-0.69%, 0.75%]). For the gendered version, participants responded significantly more positively male compared to female names paired with masculine ($MDIFF = 6.20\%$, 95%CI [2.77%, 12.16%]), non-stereotyped ($MDIFF = 5.11\%$, 95%CI [1.97%, 10.43%]) and feminine roles ($MDIFF = 4.92\%$, 95%CI [1.00%, 10.30%]). As gender stereotypicality was found to have some modicum of effect in both versions of the experiment, to further examine these effects we compared the 'slopes' of the stereotype effect between each version (focusing only on the stereotyped roles). Steeper slopes between stereotypically congruent and incongruent pairings are indicative of a stronger effect of stereotype. The results indicated equivalent slopes for male names (gender-neutral = 0.98, gendered = 1.01) but, for female names, a larger slope for the gender-neutral (0.51) compared to the gendered (0.28) version.

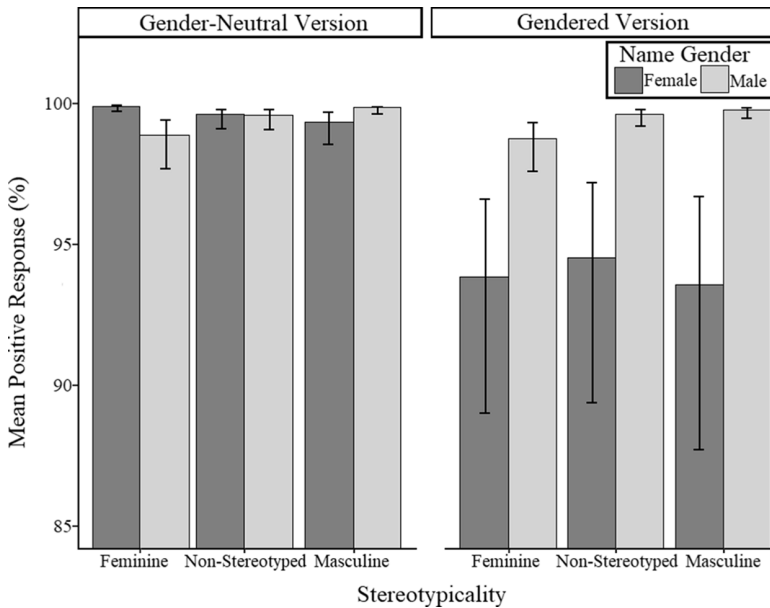


Figure 1. The effect of the interaction between Version, Stereotypicality, and Name Gender on Choice in Experiment 1. Error bars indicate the 95% confidence interval.

6.2. Positive response times

The final model contained fixed effects of Trial Number, Version, Stereotypicality, Name Gender, and Age, and the random structure was composed of random intercepts for Participant and Role Theme, and random slopes of Trial Number and Stereotypicality by Participant. Trial Number was found to have a large significant effect, Wald $X^2(1, N = 77) = 331.52, p < 0.001, \omega_p^2 = 0.81$, with participants responding increasingly quickly over the course of the experiment. Age was found to have a medium sized and significant effect, Wald $X^2(1, N = 77) = 6.28, p = 0.012, \omega_p^2 = 0.07$, with responses getting increasingly slower with age. The results indicated a large and significant main effect of Stereotypicality, Wald $X^2(2, N = 77) = 12.06, p = 0.002, \omega_p^2 = 0.21$, as well as a very small but significant two-way interaction between Version and Name Gender, Wald $X^2(1, N = 77) = 16.73, p < 0.001, \omega_p^2 < 0.01$.

The main effect of Stereotypicality indicated significant differences between the categories overall, but no significant differences between any two given categories. Participants tended to respond fastest to masculine role themes ($MRT = 1102\text{ms}$, 95%CI[1025ms, 1180ms]), slightly slower to non-stereotyped role themes ($MRT = 1124\text{ms}$, 95%CI[1047ms, 1202ms]), and most slowly to feminine role themes ($MRT = 1199\text{ms}$, 95%CI[1119ms, 1278ms]).

The two-way interaction between Version and Name Gender (Table 3, Figure 2) indicated significant differences between the categories overall, but no significant differences between any two given categories. For the gendered version, participants tended to respond more quickly to male compared to female names ($MDIFF = 36\text{ms}$, 95%CI[-158ms, 240ms]). For the gender-neutral version,

Table 3. Two-way interaction between Version and Name Gender for positive response times in Experiment 1

Version	Name Gender	
	Female	Male
Gender-Neutral	1162 [1066, 1257]	1174 [1079, 1269]
Gendered	1131 [1034, 1228]	1095 [998, 1192]

Note. Numbers are presented in milliseconds. Numbers in brackets are the 95% confidence intervals.

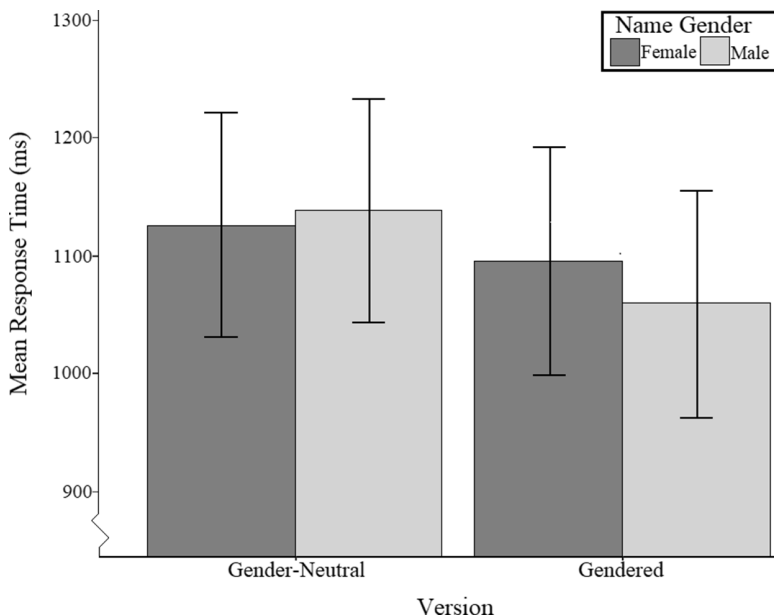


Figure 2. The effect of the interaction between Version and Name Gender on positive response times in Experiment 1. Error bars indicate the 95% confidence interval.

participants tended to respond slightly quicker to female compared to male names ($M_{DIFF} = 12\text{ms}$, $95\%CI[-178\text{ms}, 203\text{ms}]$). Further, responses in the gendered version (especially for male names) were given faster than responses in the gender-neutral version.

7. DISCUSSION

We hypothesized that participants who were presented with roles in a gender-neutral form would display a weaker ‘male bias’ (H1), but would respond more strongly in keeping with gender stereotypes than participants who were presented with roles in the masculine form (H2). The results indicated a masculine bias, and (based on the slopes) a slightly weaker stereotype effect, in

the gendered version of the experiment, offering support for our first hypothesis, and partial support for our second hypothesis.

Since our study was one of the first attempts to evaluate gender-neutral forms in French and their propensity to influence gender representations, and before extensively discussing our results, we decided to replicate the study using a different sample. To avoid a cohort geographical effect, and as the second author was visiting Québec for their University training, we decided to run the exact same experiment using a convenience sample of Québec speakers, drawing from those geographically close to the second author during their visit. To the best of our knowledge, the effect of the masculine form – as well as gender-neutral forms – has never been experimentally tested in Québec, and the discussion on gender inclusion and language has been very similar in Québec and in Switzerland. In fact, as far as we know, it has been very similar in most French-speaking regions or countries. For example, as in Switzerland, official government positions on the matter in other French-speaking countries or regions are rather conservative, such as in Québec (Samson, 2019), where official government texts are still written in the masculine form only.

EXPERIMENT 2

8. METHOD

8.1. Participants

The sample consisted of 79 French-speaking participants from Québec (mean age = 27.6 years [$SD = 8.8$ years]; 51 women and 28 men) with 42 of them having been selected through the internet-based participant recruitment pool Prolific. This sample was also a mix of students (32) and non-students (47). Participants received either a small financial reward for their time (equivalent to £2.50; Prolific users) or no reward (non-Prolific users). Those who agreed to participate in this research were given a web link to the experiment. As in Experiment 1, general ethical approval was given by the University of Fribourg, Switzerland.

8.2. Research design

The research design was the same as for Experiment 1.

8.3. Materials

The materials used were the same as for Experiment 1. To ensure appropriate comparisons of experiments, several checks were made on the stimuli, as detailed below.

Names. The first names were double-checked against the most frequent and common names given from 1997 to 2002, as set by the *Direction de la statistique et de l'analyse quantitative de Retraite Québec* [The statistics and quantitative analysis direction of Retraite Québec], and were found appropriate.

Role Themes. The stereotype norms from Misersky et al. (2014) were double-checked by two judges in Québec to ensure that no major difference would be apparent (note that in Misersky et al., 2014, the correlations across six countries were very high, ranging from 0.86 to 0.96).

8.4. Data Preparation

Deselection occurred in the same manner as with Experiment 1. On the by-participant level, one participant was deselected for being a non-native Canadian French speaker, while for error rate six participants were deselected at step one (i.e., for having an error rate at or above 50%) and zero participants were deselected at step two (i.e., for having an error rate two standard deviations above the mean). This final sample was composed of 48 women and 24 men. A total of 36 participants undertook the gender-neutral version, while 36 participants undertook the gendered version. This left 72 participants after deselection (36 Prolific), with a mean age of 27.1 years ($SD = 8.5$ years). On the item-by-participant level, responses faster than 300ms, or which hit the maximum of 5000ms, were removed (0.45% of the data).

In regards to guessing the aim of the experiment, a total of 53 of these participants – which is rather high – did so correctly (27 in the gender-neutral version, 26 in the gendered version), with 16 participants failing to correctly guess the aim of the study (eight in the gender-neutral version, eight in the gendered version), and with three participants not responding (one in the gender-neutral version, two in the gendered version). As in Experiment 1, this indicates that participants were equally likely to guess the aim of the experiment regardless of the word form used. Compared to Experiment 1, Québec participants were a little less likely than Swiss participants to guess the aim of the experiment.

Data analysis and effect size estimations (partial Omega squared) were conducted in keeping with those used in Experiment 1, with the addition of fixed effects of Student Status [student vs. non-student] and Prolific Recruitment [recruited through Prolific vs. not recruited through Prolific], and the removal of Handedness. Handedness was removed as only one ambidextrous and one left-handed individual were included in the final dataset, meaning that inclusion of handedness would result in these two individual participants' results to unduly influence the final model.

9. RESULTS

9.1. Yes/No choice responses

The model that best fit participants' responses contained Version, Name Gender, and Stereotypicality as fixed effects, with the random structure composed of a random intercept of Participant. The results indicated medium significant main effects of Version, Wald $X^2(1, N = 72) = 7.96, p = 0.005, \omega_p^2 = 0.09$, and Name Gender, Wald $X^2(1, N = 72) = 310.66, p < 0.001, \omega_p^2 = 0.06$, as well as a very small significant main effect of Stereotypicality, Wald $X^2(1, N = 72) = 15.54, p < 0.001, \omega_p^2 < 0.01$. These were

Table 4. Three-way interaction between Version, Name Gender, and Stereotypicality for Choice in Experiment 2

Version	Name Gender	Stereotypicality		
		Feminine	Non-Stereotyped	Masculine
Gender-Neutral	Female	93.36[85.59, 97.08]	93.75[86.36, 97.26]	93.98[86.82, 97.37]
	Male	99.87[99.66, 99.95]	99.96[99.88, 99.99]	99.94[99.84, 99.98]
Gendered	Female	99.90[99.69, 99.97]	99.94[99.81, 99.98]	99.66[99.02, 99.88]
	Male	99.48[98.54, 99.82]	99.88[99.64, 99.96]	99.97[99.88, 99.99]

Note. Numbers are presented in percentages. Numbers in brackets are the 95% confidence intervals.

qualified by a medium two-way interaction between Version and Name Gender, Wald $X^2(1, N = 72) = 277.93, p < 0.001, \omega_p^2 = 0.06$, and very small significant two-way interactions between Stereotypicality and Name Gender, Wald $X^2(2, N = 72) = 29.19, p < 0.001, \omega_p^2 < 0.01$, and between Version and Stereotypicality, Wald $X^2(2, N = 72) = 8.92, p = 0.012, \omega_p^2 < 0.01$, which in turn were qualified by a very small but significant three-way interaction between Version, Stereotypicality, and Name Gender, Wald $X^2(2, N = 72) = 30.57, p < 0.001, \omega_p^2 < 0.01$.

The three-way interaction between Version, Stereotypicality, and Name Gender (Table 4, Figure 3) indicated a male bias in the gendered version that was not modulated by stereotype such that the difference was highest for gender-congruent, followed by neutral, and finally gender-incongruent name-role pairs, and indicated a potential, but very weak, female bias in the gender-neutral version that was modulated by stereotype in the same manner. More specifically, for the gender-neutral version, participants tended to respond more positively to feminine roles paired with female compared to male names ($MDIFF = 0.42\%$, 95%CI [-0.13%, 1.43%]), tended to respond more positively to masculine roles paired with male compared to female names ($MDIFF = 0.31\%$, 95%CI [0.00%, 0.97%]), and responded no differently for non-stereotyped roles paired with female compared to male names ($MDIFF = 0.06\%$, 95%CI [-0.15%, 0.34%]). For the gendered version, participants responded significantly more positively male compared to female names paired with masculine ($MDIFF = 5.96\%$, 95%CI [2.47%, 13.16%]), non-stereotyped ($MDIFF = 6.21\%$, 95%CI [2.62%, 13.63%]) and feminine roles ($MDIFF = 6.51\%$, 95%CI [2.58%, 13.16%]).

As in Experiment 1, we compared the 'slopes' of the stereotype effect between each version (focusing only on the stereotyped roles). The results indicated that the female slope was generally larger for the gendered version (0.62) than for the gender-neutral version (0.24), but that the male slope was generally larger for the gender-neutral version (0.49) than for the gendered version (0.07).

9.2. Positive response times

The final model contained fixed effects of Trial Number, Version, Stereotypicality, Name Gender, Age, and Student Status as fixed effects, and the random structure was composed of random intercepts for Participant, Role Theme, and Name, as well

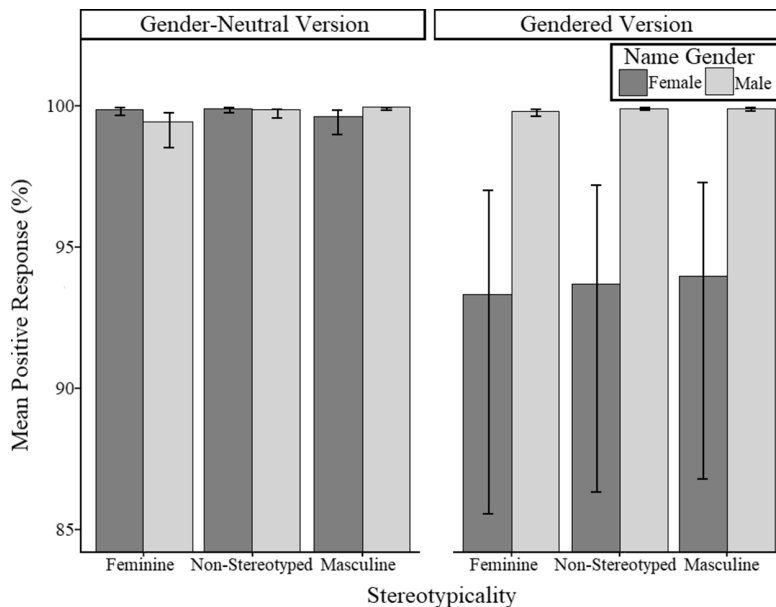


Figure 3. The effect of the interaction between Version, Name Gender, and Stereotypicality for Choice in Experiment 2. Error bars indicate the 95% confidence interval.

as random slopes of Stereotypicality by Participant and Version by Role Theme. Trial Number was found to have a large and significant effect, Wald $X^2(1, N = 72) = 2695.68, p < 0.001, \omega_p^2 = 0.16$, with participants responding increasingly quickly over the course of the experiment. Age was found to have a medium sized and significant effect, Wald $X^2(1, N = 72) = 10.98, p < 0.001, \omega_p^2 = 0.12$, with responses getting increasingly slower with age. The results indicated a large and significant main effect of Stereotypicality, Wald $X^2(2, N = 72) = 13.01, p = 0.002, \omega_p^2 = 0.20$, as well as a very small significant two-way interaction between Version and Name Gender, Wald $X^2(1, N = 72) = 20.47, p < 0.001, \omega_p^2 < 0.01$.

The main effect of Stereotypicality indicated significant differences between the categories overall, but no significant differences between any two given categories. Participants tended to respond fastest to masculine role themes ($MRT = 1061\text{ms}$, 95%CI[1000ms, 1122ms]), slightly slower to non-stereotyped role themes ($MRT = 1086\text{ms}$, 95%CI[1027ms, 1146ms]), and most slowly to feminine role themes ($MRT = 1164\text{ms}$, 95%CI[1098ms, 1229ms]).

The two-way interaction between Version and Name Gender (Table 5, Figure 4) indicated significant differences between the categories overall, but no significant differences between any two given categories. For the gendered version, participants tended to respond more quickly to male compared to female names ($MDIFF = 41\text{ms}$, 95%CI[-100ms, 181ms]). For the gender-neutral version, participants tended to respond slightly quicker to female compared to male names ($MDIFF = 9\text{ms}$, 95%CI[-135ms, 154ms]).

Table 5. Two-way interaction between Version and Name Gender for positive response times in Experiment 2

Version	Name Gender	
	Female	Male
Gender-Neutral	1116 [1043, 1188]	1125 [1053, 1197]
Gendered	1106 [1035, 1176]	1065 [995, 1135]

Note. Numbers are presented in milliseconds. Numbers in brackets are the 95% confidence intervals.

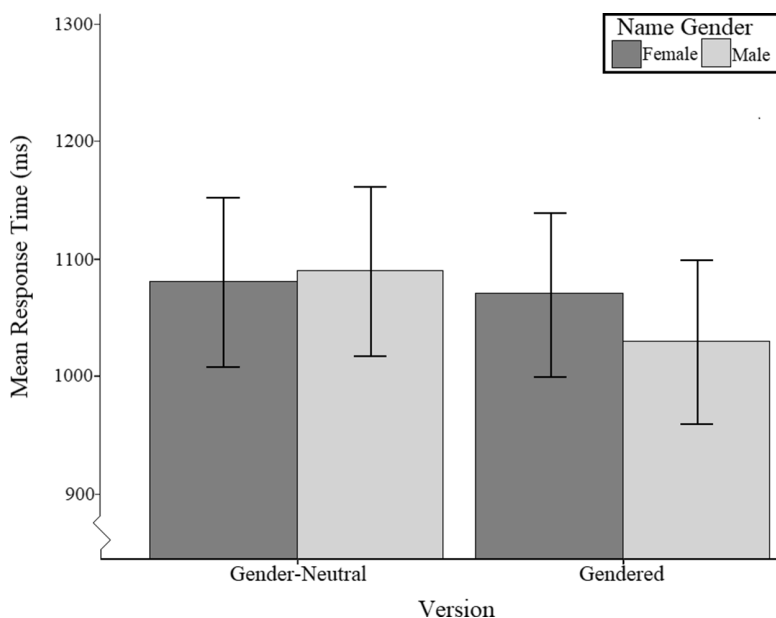


Figure 4. The effect of the interaction between Version and Name Gender on positive response times in Experiment 2. Error bars indicate the 95% confidence interval.

10. REPLICATION EVALUATION AND DATA SENSITIVITY

As the results of Experiments 1 and 2 are very similar, Bayes factors were calculated³ to determine the level to which the results found in Experiment 2 were truly indicating a male bias when the masculine form was used, replicating that of Experiment 1, as advocated for by Dienes (2014), Dienes et al. (2018), and Verhagen and Wagenmakers (2014). Bayes factors quantify the strength of the evidence that the data provides for the existence ($H1$) or absence ($H0$), given the results found in a previous experiment (or as predicted by the theory). As such,

³R code and detailed explanation of Bayes Factors provided on Zoltan Dienes web page: http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm

Bayes factors are an effective manner of determining whether the results from Experiment 1 hold in Experiment 2.

The calculation of Bayes factors was done through utilising the effects found in Experiment 1 (i.e., the estimated effects as predicted by the models) as priors, with a half-normal distribution. In other terms, we statistically verify whether the effects of Experiment 2 were truly present (H1), considering the effects of Experiment 1 as the baseline. This simply means that we tested the strength of evidence of Experiment 2 to support the effects found in Experiment 1. We were particularly interested in the two-way interaction between Version and Name Gender in both outcome measures.

All Bayes factors were above 3 (Choice: $B=2.81e+48$; positive response times: $B=5.57e+3$). Using the conventional cut-offs suggested by Jeffreys (1961), Bayes factors less than 1/3 would indicate that the interaction effect of Experiment 1 was absent in Experiment 2. Bayes factors greater than 3 would show substantial evidence for an interaction effect, as that of Experiment 1. In all, as hinted by the statistical and numerical similarities between the two experiments, the resulting Bayesian analysis showed strong evidence for the results of Experiment 2 being very similar to those of Experiment 1.

11. GENERAL DISCUSSION

In this article, we hypothesized that participants who were presented with roles in a gender-neutral form would display a weaker ‘male bias’ (H1), but would respond more strongly in keeping with gender stereotypes than participants who were presented with roles in the masculine form (H2). The results offered support for our first hypothesis, and partial support for our second hypothesis.

In line with previous research, a masculine bias was found for both choice and positive response time in the gendered version of the experiment, with French speakers from *both* experiments responding “yes” more often, and more quickly, to male first names than to female first names. However when epicenes and group nouns (i.e., gender-neutral form) were used, this male bias disappeared, in both outcome variables. Further, there was some evidence of a slight ‘female bias’ observed when roles were presented in a gender-neutral form (i.e., epicenes and group nouns), although not overcoming the gender stereotype effect, with participants responding “yes” slightly more often, and faster, to female compared to male first names. Overall these results are largely in line with H1 and previous research (e.g., Sato *et al.*, 2016), yet somehow differ from studies where traces of male bias were still apparent even with gender unmarked role nouns (e.g., Irmen & Roßberg, 2004). These findings suggest that using gender-neutral forms as alternatives to the masculine form prevents a well-documented and problematic male bias, although it raises some questions about the nature of the potential feminine bias.

The slope calculations for choice indicated that, in keeping with H2, gender stereotypicality was more often important in guiding responses to the gender-neutral form than the gendered form, with larger slopes found for the gender-neutral form for female names in the Swiss sample and male names in the Canadian sample. These findings were in keeping with previous research (e.g.,

Gygax et al., 2008; Richy & Burnett, 2021). Further, the general results for choice indicated that gender stereotypicality modulated responses to the gendered form for Swiss participants, but not for Canadian participants. Conversely, the stereotypicality of male names was found to be equal for the Swiss sample, and more important for the gendered compared to gender-neutral form for female names in the Canadian sample. These results were not in keeping with H2.

The finding that gender stereotypes informed participants' responses in the Swiss gendered version is in keeping with research suggesting that grammatical gender and stereotypes do interact (e.g., Irmen, 2007; Irmen & Roßberg, 2004; Vervecken et al., 2015), while the finding that it did not modulate participants' responses in the Canadian gendered version is keeping with other research that found the male bias to completely override stereotype effects (e.g., Gabriel & Gygax, 2008; Gabriel et al., 2008; Garnham et al., 2012; Sato et al., 2016). We believe that the paradigm used in the present study (i.e., using first names) may potentially increase French speakers' sensitivity to gender information more generally, and in potentially different manners depending on specific cultural context. The exact mechanism underlying such a heightened sensitivity would need further research. Still, as stereotype effects were always very small (slopes of less than 1% in all cases), this suggests that gender-neutral forms may be particularly adequate when one wishes to avoid any asymmetric activation of gendered information. As such, our results indicate that the use of the gender-neutral forms (epicenes and group nouns) may reduce social bias caused by grammatical gender in French.

The results of the positive response times indicated that, aside from the interaction between grammatical form and Name Gender, there was no global difference between the gender-neutral and masculine forms in the difficulty to process them in both experiments. This suggests that the same amount of cognitive resources are required for processing role titles in both gender-neutral and masculine forms. Interestingly, a main effect of Stereotypicality was found, that indicated that participants responded most slowly to feminine stereotyped roles. It is likely that this was due to feminine stereotyped roles being longer in terms of both words and characters than non-stereotyped and masculine roles for both versions of the experiment.

A few final issues need to be discussed before reaching a conclusion. Firstly, in contrast to previous research employing similar tasks (e.g. Gygax & Gabriel, 2008), the proportions of positive responses to *female name – role* pairs were generally high, and this is striking as previous studies using kinship terms (“une sœur” [a sister]) instead of names found rather lower proportions of positive responses to these pairs (e.g., between 28–70 in Gygax & Gabriel, 2008). The existence of such a difference is perhaps not surprising, as the interaction between stereotypical and grammatical gender has been found to be sensitive to the specific stimulus materials used in an experiment (Esaulova & Stockhausen, 2022). One explanation could be that, familial roles, such as “a sister”, activate primarily gender and age based expectations, while names, such as “Léa”, may activate other (non-gender) expectations, such as social status. While our name selection criteria was based on common gender typical names, we did not specifically control for social status. In order to mitigate the potential for social

status bias, future research might wish to utilise a wider range of names. Also, “une sœur” is both semantically and grammatically gendered in French, which might increase its gender salience compared to names. Yet it could be argued that some names might also be frequently preceded by a potential determiner (i.e., “la Léa”). It could be interesting to test the present experiment in a language where names can also carry more explicit grammatical gender marks, such as Italian (e.g., Roberto vs. Roberta). The effects may be closer to that with “une sœur”. In all, it might be interesting to further our understanding of the different features activated when reading first names. Secondly, as the experiments undertaken in this article were conducted through the internet-based instrument PsyToolkit, there are several issues that arise that do not exist in laboratory-based experiments (Reips, 2002; Reips et al., 2015). In terms of response time noise, often seen as the largest barrier to internet-based experimentation, PsyToolkit has been examined through a replication study (Kim et al., 2019) which found that results obtained for both choice and positive response time were in line with results obtained through the laboratory based instrument E-Prime 3.0, indicating that PsyToolkit can be utilised for delicate choice response tasks. Most of the other issues Reips (2002) and Reips et al. (2015) raise are also addressed through specific decisions made in how the experiment would be structured, in how participants would be recruited, and in how data would be analysed. However, one issue that was impossible to address is that it is not possible to ensure that everyone who undertook the experiment was fully truthful about their demographic information. We have not discussed this information nor addressed it directly, yet it might be an issue to keep in mind for future studies. As some demographic information was included in the final models for choice and positive response time, it is possible that, if people have misrepresented themselves in their demographic responses, that these results would change slightly. Thirdly, we have used different neutralizing strategies (i.e., group nouns, grammaticalized epicenes, epicenes that are gender specific in the singular form) that may prove to generate different representations if examined in detail. Future research may counterbalance these strategies to get a clearer and more accurate picture. Fourthly, as a large percentage of participants were able to guess that this experiment focused on gender stereotypicality, it is possible that their responses were affected by social desirability bias. If this is the case, participants might have sought to mitigate the effect of gender stereotypicality by responding more positively to all items. As such, the stereotype effects might naturalistically be larger for both gendered and gender-neutral language than was found in this article. One potential way to deal with this might be to increase the number of filler items with counter-stereotypical items, or to include non-stereotyped filler items (e.g., gender neutral roles, nonsense strings). Fifthly, it is possible that the frequency with which the terms used in both versions of these experiments are common in language might have affected participants’ responses, with participants responding in a more stereotypical manner to terms that they are more familiar with. As such, future research should control for potential frequency effects. Sixthly, while the use of occupations allowed for the examination of non-stereotyped roles, the use of gender-typical names means that the results focus

on a binary view of gender (female and male). As such, future research could replace the names with a different source of gender information that allows for, at the least, a three-category examination (female, non-binary, and male). This would also allow for an examination of whether non-stereotyped occupations are viewed as more appropriate for non-binary individuals, or whether there is some underlying bias that leads to non-binary individuals being perceived in the same manner as female or male individuals. Finally, very few non-binary individuals took part in the experiments conducted in this article. Future research might wish to specifically recruit equal numbers of women, non-binary, and male individuals in order to examine whether the responses from non-binary individuals are in line with the general population, or whether their experiences with gender have led to a different level of acceptance of the societal gender beliefs underpinning male bias and stereotype effects than the general population.

12. CONCLUSION

Gender-neutral role titles were found to avoid the male bias associated with the grammatically masculine form. Gender-neutral role titles were also found to require equal cognitive resources to the grammatically masculine form to process, indicating that both forms are equally cognitively easy to activate. Further, slope examination for choice indicated that the stereotype effects associated with both gender-neutral and grammatically masculine forms were quite small (<1%). These effects were quite robust across two different geographically distinct socio-cultural environments, namely the French-speaking parts of Switzerland and Québec. Taken as a whole, our results therefore suggest that switching from the masculine form to using gender-neutral nouns may reduce masculine biases while not producing large stereotype effects and not requiring more cognitive resources to process; as such, the results support the use of epicenes and group nouns as a strategy to reduce biases from the use of grammatical gender markers in grammatical gender languages.

Future studies into this topic may wish to take steps to carefully monitor and control for social status effects, social desirability bias, and frequency effects. Future studies might also consider counterbalancing neutralisation strategies to get a clearer picture of the different representations each provide, replacing names with a source of gender information that allows for non-binary representation, or purposefully recruiting equal numbers of female, male, and non-binary individuals.

In all, our study documents the propensity for different forms of inclusive language (*écriture inclusive* in French) to generate more gender-balanced representations, at least when compared to the use of the masculine form, even if meant as a generic one.

References

- Baayen, R. H., (2008) *Analysing linguistic data: a practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

- Baayen, R. H., & Milin, P.**, (2010) Analysing reaction times. *International Journal of Psychological Research*, 3, 12–28
- Bates, D., Maechler, M., Bolker, B., & Walker, S.**, (2015) Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, 67, 1–48
- Boroditsky, L.**, (2011) How language shapes thought: The languages we speak affect our perceptions of the world. *Scientific American*, 2, 63–65
- Brauer, M., & Landry, M.** (2008). Un ministre peut-il tomber enceinte? L'impact du générique masculin sur les représentations mentales. *Année psychologique*, 108, 243–272
- Cat, C. D., Klepousniotou, E., & Baayen, R. H.**, (2015) Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. *Frontiers in Psychology*, DOI: 10.3389/fpsyg.2015.00077
- Chatard, A., Guimond, S., & Martinot, D.**, (2005) Impact de la féminisation lexicale des professions sur l'auto-efficacité des élevés: une remise en cause de l'universalisme masculin? [Occupational self-efficacy as a function of grammatical gender in French] *L'Année psychologique*, 105, 249-272
- Corbett, G. G.** (1991). *Gender*. Cambridge: Cambridge University Press.
- Dienes, Z.**, (2014) Using bayes to get the most out of non-significant results. *Frontiers in Psychology*. DOI: 10.3389/fpsyg.2014.00781
- Dienes, Z., Coulton, S., & Heather, N.**, (2018) Using bayes factors to evaluate evidence for no effect: examples from the SIPS project. *Addiction*, 113(2): 240–246
- Dienes, Z.**, Making the most of your data with Bayes. *School of Life Sciences at the University of Sussex – Staff Portfolio of Professor Zoltan Dienes*. Retrieved on the 28th of January 2022 from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm
- Esaulova, Y., & von Stockhausen, L.**, (2022) Empirical evidence for subtle gender biases in language. In D. Ayoun (Ed.), *The acquisition of gender: crosslinguistic perspectives* (pp. 49–70). <https://doi.org/10.1075/sibil.63.03esa>
- Fausey, C. M., & Boroditsky, L.**, (2011) Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, 18, 150–157
- Fox, J.**, (2003) Effect displays in R for generalized linear models. *Journal of Statistical Software*, 8, 1–27
- Gabriel, U., & Gygax, P.**, (2008) Can societal language amendments change gender representation? The case of Norway. *Scandinavian Journal of Psychology*, 49(5): 451–457
- Gabriel, U., & Gygax, P.**, (2016) Gender and Linguistic Sexism. In H. Giles & A. Maass (Eds.), *Advances in intergroup communication* (pp. 177–192). Peter Lang Publishing.
- Gabriel, U., Gygax, P. M., & Kuhn, E. A.** (2018). Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations*, 21, 844–858. <https://doi.org/10.1177/1368430218771742>
- Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A. & Oakhill, J.** (2008). Au-pairs are rarely male: Role names' gender stereotype information across three languages. *Behavior Research Methods Instruments and Computers*, 40, 206–212.
- Garnham, A., Gabriel, U., Sarrasin, O., & Gygax, P.**, (2012) Gender representation in different languages and grammatical marking on pronouns: when beauticians, musicians, and mechanics remain men. *Discourse Processes*, 49(6): 481–500
- Gorfein, D. S., & Bubka, A.**, (1989) *Resolving semantic ambiguity*. New York, NY: Springer.
- Gorfein, D. S.**, (Ed.) (2001) *On the consequences of meaning selection: perspectives on resolving lexical ambiguity*. Washington, DC: American Psychological Association.
- Gorfein, D. S., Brown, V. R., & DeBiasi, C.**, (2007) The activation-selection model of meaning: explaining why the son comes out after the sun. *Memory & Cognition*, 35, 1986–2000
- 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
- Gygax, P., & Gabriel, U.**, (2008) Can a group of musicians be composed of women? Generic interpretation of French masculine role names in 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, 395–408

- Gygax, P., Gabriel, U., Sarrasin, O., Garnham, A., & Oakhill, J., (2008) Some grammatical rules are more difficult than others: the case of the generic interpretation of the masculine. *European Journal of Psychology of Education*, 24, 235–246
- Gygax, P., Gabriel, U., & Zufferey, S., (2019) Le masculin et ses multiples sens : Un problème pour notre cerveau...et notre société. [The masculine and its multiple meanings: A problem for our brain... and our society] *Savoirs En Prisme*. Retrieved on the first of march 2019 from <https://savoirsenprisme.com/>
- Hamilton, M. C., (1991). Masculine bias in the attribution of personhood. *Psychology of Women Quarterly*, 15, 393–402. <http://doi.org/10.1111/j.1471-6402.1991.tb00415.x>
- Harjunen, V., Spape, M., Ahmed, I., Jacucci, G., & Ravaja, N., (2018) Persuaded by the machine: the effect of virtual nonverbal cues and individual differences on compliance in economic bargaining. *Computers in Human Behaviour*, 87, 384–394.
- Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of psycholinguistic research*, 36, 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, 272–307. doi: 10.1177/0261927X04266810
- Jeffreys, H., (1961) *Theory of Probability* (3rd edition). Oxford: Oxford University Press.
- Kim, J. D., Gabriel, U., & Gygax, P., (2019) Testing the effectiveness of the Internet-based instrument PsyToolkit: a comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLoS One*, 14, e0221802
- Krauss, R. M., & Chiu, C.-y., (1997) *Language and social behavior*. In Gilbert, D., Fiske, S., & Lindsey, G., (Eds.) *Handbook of social psychology*, vol. 2. (pp. 41–88) Boston: McGraw-Hill
- Lévy, A., Gygax, P., & Gabriel, U., (2014) Fostering the generic interpretation of grammatically masculine forms: when my aunt could be one of the mechanics. *Journal of Cognitive Psychology*, 26, 27–38
- Lüdtke, D. (2018) *Sjstats: statistical functions for regression models*. <https://CRAN.R-project.org/package=sjstats>
- Misersky, J., Gygax, P., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T., Englund, K., Hanulíková, A., Öttl, A., Valdová, J., Von Stockhausen, L., & Sczesny, S., (2014) Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavioural Research Methods*, 46, 841–871
- Nakagawa, S., & Schielzeth, H., (2010) A general and simple method for obtaining R^2 from generalised linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133–142
- Pyykkönen, P., Hyönä, J., & van Gompel, R. P. G., (2010) Activating gender stereotypes during online spoken language processing: evidence from visual world eye tracking. *Experimental Psychology*, 57(2): 126–133
- Reips, U. -D., (2002) Standards for internet-based experimenting. *Experimental Psychology*, 49, 243–256
- Reips, U. -D., Buchanan, T., Krantz, J. H., & McGraw, K., (2015) Methodological challenges in the use of the internet for scientific research: ten solutions and recommendations. *Studia Psychologica. Advance online publication*. Doi : 10.21697/sp.2015.14.2.09
- Richy, C., & Burnett, H. (2021). Démêler les effets des stéréotypes et le genre grammatical dans le biais masculin : Une approche expérimentale. *GLAD! Revue sur le langage, le genre, les sexualités*, 10, <https://doi.org/10.4000/glad.2839>
- Rudinger, R., Naradowsky, J., Leonard, B., & Van Durme, B., (2018) Gender bias in coreference resolution. *Proceedings of NAACL-HLT*, North American Chapter of the Association for Computational Linguistics: Human Language Technology, New Orleans, 2018. Michigan: ACL Anthology
- Samson, V. (2019). Féminisation de la langue et rédaction inclusive : le Québec tantôt précurseur, tantôt conservateur. Flückiger, A. (ed.). *La rédaction administrative et législative inclusive : La francophonie entre impulsions et résistances*. Stämpfli Editions : Berne.
- Sato, S., Gabriel, U., & Gygax, P., (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): 1–19
- Schubert, T. W., Murteira, C., Collins, E. C., & Lopes, D., (2013) ScriptingRT: a software library for collecting response latencies in online studies of cognition. *PLoS One*, 8, e67769

- Silveira, J.** (1980). Generic masculine words and thinking. *Women's Studies International Quarterly*, *3*, 165–178. [http://doi.org/10.1016/S0148-0685\(80\)92113-2](http://doi.org/10.1016/S0148-0685(80)92113-2)
- Steiger-Loerbroks, V., & von Stockhausen, L.**, (2014) Mental representations of gender-fair nouns in German legal language: an eye-movement and questionnaire based study. *Linguistische Berichte*, *2014*(247): 57–80
- Stoet, G.**, (2010) PsyToolkit – a software package for programming psychological experiments using Linux. *Behaviour Research Methods*, *42*, 1096–1104
- Stoet, G.**, (2017) PsyToolkit: a novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*, 24–31
- Verhagen, J., & Wagenmakers, E.-J.**, (2014) Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, *143*(4): 1457–1475
- Vervecken, D., Gygax, P., 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*. <https://doi.org/10.3389/fpsyg.2015.01437>
- Viennot, E.** (2018). *Le langage inclusif : Pourquoi, comment* [Inclusive language : why, how]. Éditions iXe. [<https://www.editions-ixe.fr/catalogue/le-langage-inclusif-pourquoi-comment/>]
- Wagner-Egger, P., & Gygax, P.**, (2017) Diana was not involved in the 9/11 terrorist attacks! Or was she? Newspaper headings and the boomerang effect. *Swiss Journal of Psychology*, *77*, 15–22
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., Chang, K.-W.**, (2018) Gender bias in coreferent resolution: evaluation and debiasing methods. *Proceedings of NAACL-HLT*, North American Chapter of the Association for Computational Linguistics: Human Language Technology, New Orleans, 2018. Michigan: ACL Anthology

Cite this article: Kim J, Angst S, Gygax P, Gabriel U, and Zufferey S (2023). The masculine bias in fully gendered languages and ways to avoid it: A study on gender neutral forms in Québec and Swiss French. *Journal of French Language Studies* *33*, 1–26. <https://doi.org/10.1017/S095926952200014X>