Alternatives to the masculine generic in two countries: A comparative study of mental representations of gender in the Netherlands and Belgium

Boudewijn, Julia

Universiteit Leiden
Psychologie
Faculteit der Sociale Wetenschappen

Alternatives to the masculine generic in two countries

A comparative study of mental representations of gender in the Netherlands and Belgium

Name: Julia Boudewijn
Student number: s2284715
Date: August 17th 2023
Supervisor: Juan Olvido Perea-García
Second reader: Fabiola Diana
Word count: 11518
Cognitive Psychology
Thesis Msci Applied Cognitive Psychology

Abstract

Feminist language critics have argued that the use of a masculine generic, i.e., a masculine noun in reference to people irrespective of their gender, contributes to the under-representation of women (Backer & Cuypere, 2012). There are two alternatives available: feminization, in which both the masculine and feminine form are used, and neutralization, in which a gender neutral form is used. Given the differences between different languages regarding grammatical gender, there might be differences in which alternative is most effective to counteract the under-representation of women in each language. In this study we examined these differences between Dutch as spoken in the Netherlands and Dutch as spoken in Belgium, as the former only uses two grammatical genders and the latter seems to use three. The purpose of this study was to examine which alternative has the most potential to balance the mental representation of gender in each country, and to compare the countries to each other. We hypothesized that neutralizations would be a more effective alternative in the Netherlands compared to Belgium and for feminizations vice versa. Furthermore, we hypothesized that both alternatives would generally lead to a more balanced mental representation of gender than masculine generics. To examine this mental representation of each type of noun, a modified version of the EAST was used. The final sample consisted of 31 participants: 29 from the Netherlands and 2 from Belgium. The results showed no significant differences between the countries or the different types of nouns, which means that none of the hypotheses were supported. For the comparisons between the two countries, this was likely due to the small Belgian sample. Therefore, we could not draw any valid conclusion about whether there is a difference between the Netherlands and Belgium. Future research should be able to draw a more reliable conclusion about that. Additionally, future research could focus on the influence of stereotypicality on the effectiveness of alternatives to the masculine generic, or on increasing gender neutrality in languages.

Keywords: masculine generic, feminization, neutralization, Dutch, the Netherlands, Belgium, EAST, mental representation

Alternatives to the masculine generic in two countries

Languages have many ways to refer to groups of people. One of those ways is using a masculine generic. The masculine generic is a masculine noun in reference to people irrespective of their natural gender. This means that it could be used to refer to a group consisting of both men and women. An example would be the Dutch word ‘winnaars’ (‘winners’). This is a masculine noun, but you could use it in reference to groups that include women. This does, however, create a problem. Feminist language critics have argued that using these masculine generics contributes to the under-representation of women (Backer & Cuypere, 2012).

In the Dutch language, the first critiques date back to 1975, when Annie Romein-Verschoor published an article on gender bias in Dutch, referring to generic nouns with masculine pronouns, and gender-specific terms for professions (Gerritsen, 2002). She argued that male bias causes the language to lag behind the actual relations between men and women, which consequently slows down any progress in these relations. For example, when people hear the word ‘arts’ (‘doctor’), they tend to think of a man, even though it is a neutral term. This leaves a woman invisible and the idea of a female doctor out of the picture (Romein-Verschoor, 1975). Romein-Verschoor published another article in 1977 on gender bias in languages that led the executive committee of the *Vereniging voor Vrouwen met een Wetenschappelijke Opleiding* (VVAO; ‘Society for Women with an Academic Education’) to request linguists to find solutions to this gender bias in Dutch. As a response, both linguists and non-linguists set to find out how Dutch could be made more gender fair (Gerritsen, 2002).

One way to decrease gender bias in Dutch is to introduce gender-fair alternatives to the masculine generic. Two alternatives are available: neutralization and feminization (or differentiation). Neutralization refers to the use of one gender neutral word for both men and women, e.g., ‘leidinggevende’ (‘manager’) instead of ‘baas’ or ‘bazin’ (respectively the masculine and feminine version of ‘boss’). Feminization refers to the use of both the masculine and the feminine version of a word, e.g., ‘bazen en bazinnen’ (‘male and female bosses’; De Cocq & Redl, 2021).

The question is which of these two alternatives is preferred. Throughout the 1980s and 1990s there has been a lot of discussion about this in both the Netherlands and Belgium. In the Netherlands it was initially recommended by a committee set up by the Ministry of Social Affairs that neutral terms, like ‘dokter’ (‘doctor’) or ‘professor’ (‘professor’) should be used.

Their reasoning behind this was that it was unimportant to know the gender of a person practicing a profession, which was in line with Romein-Verschoor's 1975 article. Furthermore, using neutral terms would eliminate gender differentiation, allowing women to practice professions that were formerly mainly practiced by men. As a result, the tendency to mainly think of men when talking about said professions would disappear as well. Additionally, gender differentiation in occupational terms might also lead to functional differentiation, in which the terms referring to women could gain a lower social status than those referring to men. However, these recommendations resulted in critique. First of all, for masculine and feminine occupational terms that had the words for 'man' and 'woman' in the word, new neutral terms were created. For example, the neutral term for 'timmerman' ('male carpenter') and 'timmervrouw' ('female carpenter') became 'timmer' ('male or female carpenter'). These new words were considered ridiculous and thought to never gain popularity. Second, the new terms were not considered neutral at all, as they seemingly only referred to men. Based on this, it was argued that there should be gender specific occupational terms. In the end, no conclusion could be drawn from this discussion and none of the proposed guidelines – from the committee or their opponents – became official guidelines (Gerritsen, 2002).

In Belgium there was a similar discussion on the feminization of occupational terms. According to the linguist Willy van Langendonck creating new feminine terms would be unwanted and cause confusion (Gerritsen, 2002). Other arguments against these gender-specific occupational terms were that creating feminine terms would cause the original terms to lose their gender neutrality and become exclusively masculine in meaning. These neutral terms are useful when describing a group of people. Additionally, for some terms no feminine version is possible, whereas for other terms multiple feminine versions are possible. A lot of the newly created feminine terms do not correspond with Dutch grammar, making them not very intuitive for Dutch speakers. Lastly, it seemed optimistic to think that new words and grammar rules can be introduced into a language that easily (Cohen, 1997). Eventually it was decided that no conclusion could be drawn about the choice between neutralization and feminization in either the Netherlands or Belgium (Gerritsen, 2002).

Nowadays, it seems to be the case that people are inclined to use less feminine and more neutral terms. This is in line with the evolution within Dutch, in which grammatical gender of nouns seems to matter less and slowly dies out (Mortelmans, 2008). Because of that, feminization would be more difficult to convincingly introduce. Nonetheless, there are

situations in which feminine terms are still used, for example for a profession that is mainly practiced by women, like ‘lerares’ (‘teacher’). Although, in these situations neutral words are also used. However, in professions in which prestige is important, people tend to not use either neutralization or feminization, but often opt for masculine terms to refer to women, like for example ‘advocaat’ (‘lawyer’), ′kok′ (‘chef’), or ‘journalist’ (‘journalist’). This trend also showed from a sample from hits on Google Search and a sample from a women’s magazine called *Flair* that were analyzed to see which occupational terms were used in reference to women. The latter sample can be divided into a subsample from the Netherlands and one from Flanders. This division disclosed a difference between the two subsamples. In the subsample from Flanders there were a lot less masculine terms compared to that of the Netherlands. This seemed to imply that in Flanders people tend to use feminization more, whereas in the Netherlands they use more neutralization (Mortelmans, 2008).

So far, the discussion on which alternative is best has been mostly from a social point of view and not so much on the cognitive effects of each of the alternatives. That is, a lot has been talked about what the end-result should be, but little attention has been paid to the cognitive systems through which change should be instantiated. To truly find a solution to the linguistic under-representation of women, it is important to look at the effects that these alternatives have on the mental representation of gender and to find out which alternative actually has the most potential to balance this mental representation. This likely depends on the kind of language. In some languages neutralization is not possible, as in those languages nouns have a grammatical gender. These languages, like for example German, are called grammatical gender languages. The presence of grammatical gender markings like articles makes it hard to neutralize a noun. In these languages feminization is possible. Neutralization does occur in natural gender languages. Gender markings in these languages refer almost always to the actual gender of a person, like for example in English (De Cocq & Redl, 2021).

A study by Nissen (2013) indeed seemed to show that feminization increases visibility of women in a grammatical gender language. The study replicated another study by Nissen (1997) that was conducted a decade prior. In both studies three questionnaires were filled out by native Spanish speakers. These questionnaires consisted of sentences containing either masculine generics, feminizations, or neutralizations. In each sentence there were two blank spaces, which the participants had to fill in with two first names. The names that the participants filled in would subsequently show whether the different forms exhibit different

associations with men or women. The results showed that only feminizations exhibit a bias toward an association with women (Nissen, 2013).

Furthermore, a review by Gabriel et al. (2018) comparing feminization and neutralization in different languages advised that feminizations should be used in gendered contexts and neutralizations in contexts that are non-gendered, to keep the context gender neutral. This conclusion was based on the findings that using gender neutral words contributed to a less biased gender representation in the absence of other gender cues. However, when gender cues, such as stereotypical expectations, were present, this facilitated other types of biases related to these stereotypes. Feminizations, on the other hand, were not susceptible to stereotypical expectations (Gabriel et al., 2018). In fact, when this strategy was used gender stereotypical occupations were seen as less stereotyped (Vervecken et al., 2015). Similarly, having a grammatical gender system creates a gendered context. In addition to that, Gabriel et al. (2018) discussed an approach by Slobin (2003) that assumes that different languages have different options to grammatically encode certain characteristics of objects. Therefore, the way these characteristics are attended to will vary between different languages. This means that speakers of a language that grammatically encodes gender will think of the concept of gender and its communicative significance when speaking the language (Gabriel et al., 2018). This effect was shown by a study by Chen and Su (2011), which examined the performance in gender and non-gender related questions of speakers of Chinese, a language that does not have gendered third-person pronouns, and English, a language that does have gendered pronouns. The results showed that English speakers were faster and more accurate in answering gender related questions than Chinese speakers (Chen & Su, 2011). Following this argument, speakers of a grammatical gender language might be more inclined to use feminizations, which explicitly refers to both men and women, because the nature of their language makes communicating a referent's gender significant. Because of this, neutralizations might elicit a gender association as well, negating their intended gender neutrality.

The Dutch language finds itself between a natural and a grammatical gender language, with two grammatical genders: neuter and common. References in gender that still exist are led by the actual gender of the referent, and therefore, Dutch tends more towards a natural gender language in the practical sense (De Cocq & Redl, 2021). However, Dutch used to have three grammatical genders. In fact, this three-gender system is still found in most dialects spoken in Belgium, as Belgium may be more conservative as regards grammatical gender (De

Vogelaer, 2010). Furthermore, these differences could be due to the fact that Dutch as spoken in the Netherlands has been influenced by English (a natural gender language) and Dutch as spoken in Belgium by French (a grammatical gender language; Donaldson, 1983).

Dutch as spoken in the Netherlands tends more towards a natural gender language, whereas Dutch as spoken in Belgium seems to be more of a grammatical gender language. Given the differences in possibilities in neutralization and feminization in these different kinds of languages, it stands to reason that there are differences in the most effective alternatives to the masculine generic in the Netherlands versus Belgium. In this study, we investigated these differences; to see whether there actually is a difference and if so, to identify what this difference is. Furthermore, we investigated how the different types of gender fair alternatives fared compared to the masculine generic to balance the mental representation of gender. In sum, the purpose of this study was to compare the mental representation of gender induced by male generic nouns and gender fair alternatives to the male generic between people from the Netherlands and Dutch-speaking people from Belgium. Furthermore, we sought to determine which terms have the most potential to balance the mental representation for each country.

In order to create a gender fair language in which women are as represented as men, the type of noun people associate with both men and women should be taken into account. If for one country one kind of alternative creates a more balanced mental representation of gender and for another country another alternative, it would not make sense if both countries used the same alternative, even if in both countries the same language is spoken. There is not a lot of empirical research on this topic for Dutch; a lot of literature has focused on a more theoretical perspective (Backer & Cuypere, 2012). Therefore, with this study we aimed to provide more insight from an empirical perspective. To investigate the participants' attitudes towards the different types of nouns, we used a modified version of the extrinsic affective Simon task (EAST). The EAST is a test used as an indirect measure of attitudes. In the EAST, there are trials in which white words are presented. In these trials the participant is asked to respond to the meaning of the stimulus with one of two keys on a keyboard (De Houwer, 2003). In this study the white words were Dutch first names and the required response was whether the name is male or female. Doing so causes the keys used for the responses to become associated with either men or women. Besides trials with white words, there will also be trials with colored words, in which the participant has to indicate whether the word is green or blue, using the same keys as in the trials with white words (De Houwer, 2003). In this

study, the stimuli for the colored word trials were the different types of nouns. Since an association between the response keys and men or women was created, it is expected that responses to colored words will be faster and more accurate when participants' attitudes related to the stimulus correspond with this association. So, if the participant has to respond to the color of a word with the same key as for female names, and they associate the word with women, then the response will be faster and more likely to be correct. Therefore, the reaction times and success rate of the female response-associated trials form an indication of the mental representation of women associated with the different types of nouns.

Hypotheses

In Dutch, both neutralization and feminization are possible. However, for feminization this is not always the case in the Netherlands. For some words, a feminine version does not exist, e.g., the word 'arts' ('doctor') does not have a feminine counterpart. As such, the Dutch language as spoken in the Netherlands tends more towards neutralization (De Cocq & Redl, 2021). On the other hand, in Dutch as spoken in Belgium, feminization does often seem to be possible. Whereas in the Netherlands no feminine counterpart exists for, for example, 'arts', in Belgium the word 'artse' is very common (Cohen, 1997). Given these possibilities in the respective countries, this would mean that people in the Netherlands are more likely to use neutralization as a strategy and people in Belgium feminization. In practice this also seems to be the case (Mortelmans, 2008). As such, we expected that participants from the Netherlands associate neutralization more with women than participants from Belgium, and the reverse for feminization. Therefore, we hypothesized that

1. female response-associated trials with a neutralization as the stimulus lead to shorter reaction times in participants from the Netherlands than participants from Belgium.
2. female response-associated trials with a neutralization as the stimulus lead to a higher success rate in participants from the Netherlands than participants from Belgium.

This means that we hypothesized that participants from the Netherlands respond faster and more accurately to trials with a stimulus like 'leerkrachten' ('teachers') than participants from Belgium. The reaction times and success rates were the means of the trials with neutralizations as the stimulus, when the same response key was used as for female names. Additionally, we hypothesized that

3. female response-associated trials with a feminization as the stimulus lead to shorter reaction times in participants from Belgium than participants from the Netherlands.

4. female response-associated trials with a feminization as the stimulus lead to a higher success rate in participants from Belgium than participants from the Netherlands.

This means that we hypothesized that participants from Belgium respond faster and more accurately to trials with a stimulus like 'leraren en leraressen' (‘male and female teachers’) than participants from the Netherlands. The reaction times and success rates were the means of the trials with feminization as the stimulus, when the same response key was used as for female names.

Furthermore, in Dutch a masculine noun is often seen as gender neutral. However, factually these nouns exclude everyone except men (De Cocq & Redl, 2021). Therefore, we hypothesized that

5. female response-associated trials with the masculine generic as the stimulus lead to longer reaction times and a lower success rate than female response-associated trials with neutralizations or feminizations as the stimulus in participants from both countries.

This means that we hypothesized that participants respond faster and more accurately to trials with a stimulus like ‘leerkrachten’ or ‘leraren en leraressen’, than trials with a stimulus like ‘leraren’ (‘male teachers’). The reaction times and success rates were the means of the trials with either masculine generics, neutralizations or feminizations as the stimulus, when the same response key was used as for female names.

Methods

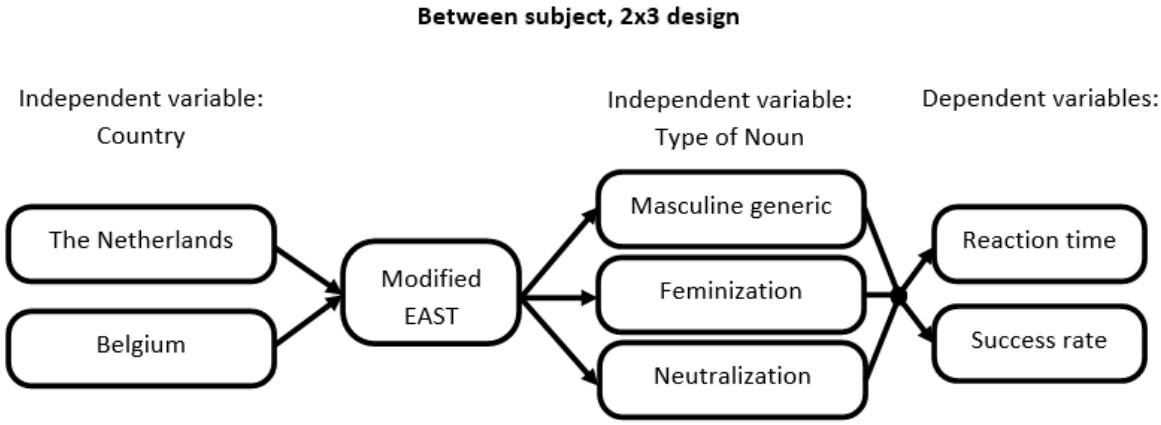
Design

This study followed a between-subject, 2x3 design (see figure 1). There were two groups: the first group consisted of participants from Belgium and the second consisted of participants from the Netherlands. All participants performed the modified EAST. The independent variables were the country the participant is from (nominal, 2 levels: Belgium, the Netherlands), and the type of noun that is used (nominal, 3 levels: masculine generic, neutralization, feminization). The three different types of nouns all appeared for each participant in the modified EAST. The dependent variables were the reaction time and the success rate of the colored word trials as measured by the modified EAST (both interval). We controlled for alternative explanations by assessing the participants’ sexism levels and their

political preference through inventories. The order of the trials in the EAST were counterbalanced to prevent order effects; half of the participants was presented with the stimuli in one order, while the other half was presented with the stimuli in the reverse order.

Figure 1

Study design



Participants

Dutch and Belgian people whose first language is Dutch were recruited to participate in this study via SONA, via personal contacts, and via social media. Each group was to consist of approximately 25 to 30 students, for a total of 50 to 60 participants. This is similar to other studies using a version of the EAST (De Houwer, 2003; De Houwer & De Bruycker, 2007; Huijding & De Jong, 2005). Half of the participants were to be male and half were to be female. Exclusion criteria were color blindness or other uncorrected visual problems, problems in hand movements, and (a history of) attentional problems or neurological disorders or impairments. Participants were compensated for their efforts with 1 credit. This study was approved by the Leiden Psychology Research Ethics Committee under application number 2022-06-24-J.O. Perea García-V2-4039. We conducted the study in accordance with the ethical guidelines.

In total, there were 42 people who participated in this study. This sample consisted of 39 Dutch participants and 3 Belgian participants. Furthermore, 3 of them were male and 39 were female. The minimum age was 18 and the maximum age was 47 (M = 20.05, SD = 4.633). Observations were excluded from the data when they departed more than 2 standard deviations below or above the 25th or 75th percentiles respectively on one or more of the variables. This resulted in the exclusion of data from 11 participants, leaving a final sample of

N = 31, that consisted of 29 Dutch participants and 2 Belgian participants. The total sample consisted of 2 male and 29 female participants. The minimum age was 18 and the maximum age was 47 (*M* = 20.39, *SD* = 5.277).

Measures

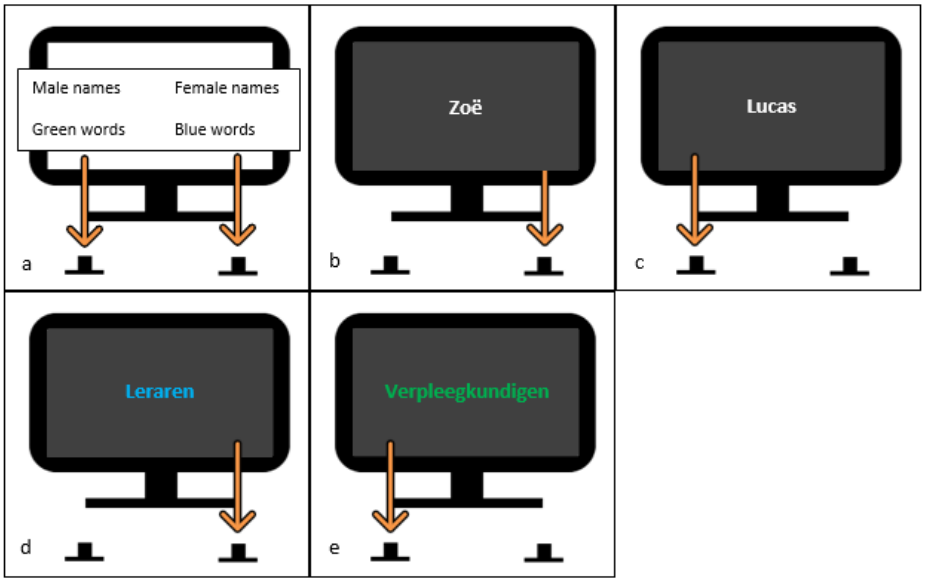
The modified EAST and the questionnaires were completed on an iiyama Vision Master™ Pro 454 monitor and Dell OptiPlex 3070 desktop. The modified EAST was run in E-Prime 3. The questionnaires were completed in Qualtrics.

Modified extrinsic affective Simon task (EAST)

To measure the reaction times and the success rates of the colored word trials, we used a modified version of the extrinsic affective Simon task or EAST (De Houwer, 2003). The EAST is an indirect measure of attitudes in which participants are presented with words on a black screen. When a white word is presented, the participant is instructed to respond to the valence of the word using a left or right key (e.g., press left if the word is positive and right is the word is negative). When the word is colored (blue or green), the participant is instructed to respond to the color of the word (e.g., press left for blue and right for green). Since in the trials with white words the keys are assigned to either positive or negative, these keys become extrinsically associated with positive or negative valence. Thus, on colored word trials, when a word is positive, participants are faster and more accurate when they have to respond with the same key as for positive white words. Similarly, when a word is negative, reaction time and accuracy improve when the participants have to respond with the same key as for negative white words (De Houwer, 2003). In the modified EAST (see figure 2) instead of positive and negative white words, we used male and female Dutch names. In these trials, the participants had to respond to whether the name was male or female. For the colored words, we used the different types of nouns (masculine generic, neutralization, feminization). In these trials the participants had to respond to the color of the word. In the colored word trials, trials were congruent when a noun was associated with the gender linked to the response key. In other words, if the noun was a blue colored neutralization, and the response key for blue nouns was the same key as for female names, the reaction time and accuracy would improve if the participant associated the noun with women. The stimuli that were used in the modified EAST are found in Appendix A.

Figure 2

Modified EAST



Note. In (a) the response options are shown. (b) (c) (d) and (e) show example stimuli and their correct responses.

Ambivalent Sexism Inventory (ASI)

To assess sexism levels, we used the Ambivalent Sexism Inventory or ASI (Glick & Fiske, 1996). The ASI is used to measure two components of sexism that each represent different orientations toward women. The first component is hostile sexism, which refers to an antipathy toward women. The second component is benevolent sexism, which refers to attitudes of viewing women in a stereotypical way and in restricted roles, but with a seemingly positive tone. An example of this is seeing women as caregivers. The ASI consists of 22 items, each item consisting of a statement relating to the roles of men and women in society. For each item, the participants were instructed to indicate the degree to which they agree or disagree with the statement using a scale ranging from 0 (disagree strongly) to 5 (agree strongly; Glick & Fiske, 1996). To calculate the total score, the scores for all the items were averaged, after reversing the score for some of them (items 3, 6, 7, 13, 18, and 21). For this study, we translated all items to Dutch.

Political preference inventory

To assess the political preference of the participants, we used an inventory of political statements. The inventory was based on the one used in the paper by Brenner and Inbar

(2015). For each statement, the participants were asked to indicate whether they agree with it using a scale ranging from 1 (totally disagree) to 5 (totally agree; Brenner & Inbar, 2015). The statements used in this study were those that were similar to currently relevant issues in both the Netherlands and Belgium, based on the statements presented on voter-information websites. In total, 10 statements were chosen, of which 5 were conservative statements and 5 were progressive. We translated the statements to Dutch. These statements are found in Appendix B. To calculate the total score, the scores for all the items were averaged, after reversing the scores for the conservative items.

Procedure

Participants were asked to come to the FSW lab. Before the experiment started, participants were provided with an information letter to inform them about the study. The participants were reminded of the exclusion criteria and they were informed about the different questionnaires and tasks they were about to perform during the experiment, as well as the duration of the entire experiment. Furthermore, the participants were informed on the compensation for their participation and confidentiality of the data that would be collected, and they were provided with the contact information of the investigators for further questions. In the information letter the participants were, however, not informed about the actual purpose of the study, as to make sure there would not be any influences from knowing this. Next, the participants were asked for their consent to participate in the study.

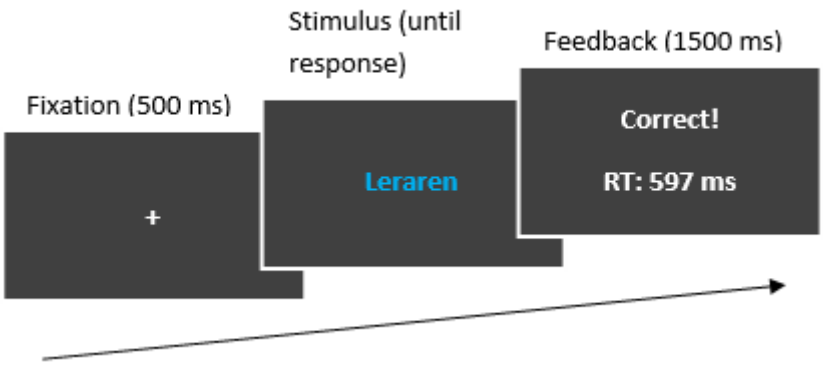
After that, the experiment started. The participants were sat behind a computer in the lab. The test leader opened the Qualtrics and the participants were asked to answer some demographic questions like age, gender, and which country they are from. This part of the study lasted for less than a minute. After finishing these questions, the participants notified the test leader. Next, the test leader opened E-prime and the participants were asked to perform the EAST.

The EAST started with some general instructions. Then, by clicking the space bar on the keyboard, the participants proceeded to the different blocks. First, there were two practice blocks, both consisting of 24 trials. In the first practice block, the participants were presented with six male names and six female names, each presented twice in white. The participants responded to each name by indicating whether the name was male or female. This was done by clicking either the Q-key on the keyboard for male names, or the P-key on the keyboard for female names. In the second practice block, the participants were presented with 12

word(group)s. Four of these were masculine generics, four were feminizations, and four were neutralizations. Each word(group) was presented once in green, and once in blue. When it was presented in green, the participants responded by clicking the Q-key, and when it was presented in blue, the participants responded by clicking the P-key. After the practice blocks, four test blocks followed. Each test block consisted of 36 trials. Each of the names from practice block 1 were presented once in white, and each of the word(group)s from practice block 2 were presented once in each color. The responses were the same as in the respective practice blocks. The order of the trials in each block was randomized, with the restriction that the same word(group) could not appear on two consecutive trials and that the required response could not be the same on four or more consecutive trials. Each test block started with three warm up trials; across the four test blocks, all of the 12 names appeared in one of these warm up trials. Furthermore, the order of the trials were counterbalanced. During the practice blocks, the order of the trials was reversed for half of the participants. During the test blocks, the order of the warm up trials and the order of the rest of the trials were both reversed separately for half of the participants. A trial started with a fixation cross which lasted 500 ms. Then a word(group) was presented until a response was given. Lastly, the participants were presented with feedback which lasted 1500 ms. See figure 3 for a screen-by-screen demonstration of an example trial. After finishing all the blocks, the participants were presented with a goodbye slide. This part of the study lasted for about 10 to 12 minutes.

Figure 3

Screen-by-screen demonstration of an example trial



Next, the participants notified the test leader that they completed the EAST. The test leader closed E-prime and reopened the Qualtrics, where the participants were asked to fill in

the ASI and the political preference inventory. After completing these, the participants had reached the end of the experiment. This part of the study lasted for about 5 to 7 minutes.

After completing the experiment, the participants were debriefed. The purpose of the study was explained to them, as well as why this purpose had previously been withheld. They received a compensation for their efforts in the form of 1 credit. The duration of the study was about 25 minutes.

Analysis

There were not enough participants in the Belgian group to properly test the hypotheses comparing participants from Belgium with participants from the Netherlands. However, the analyses are still described as they were originally intended.

All the E-Prime data were merged into one file, after which the reaction times and accuracy data were copied to Excel. There they were organized and for the participants who followed the counterbalanced design, the data were reversed to the original order. The organized data were then imported into SPSS. Mean reaction time and mean success rate scores were calculated for each type of noun and for whether the same response key was used as for male or female names, creating a total of 12 variables. The mean success rate scores were transformed into count variables. The data from Qualtrics were registered in an Excel document, where the score for the ASI and the political preference inventory scores were calculated. For the ASI, the scores for item 3, 6, 7, 13, 18 and 21 were reversed, after which all the scores were averaged. For the political preference inventory, the scores for the first five items were reversed, after which all the scores were averaged. Subsequently, these scores, and the demographic data were imported into SPSS as well.

To start, boxplots were made of the ASI score, political preference inventory score and all the mean reaction time scores and mean success rate scores to test for outliers. If any outliers were detected, they were deleted from the sample.

To test whether female response-associated trials with a neutralization as the stimulus led to shorter reaction times in participants from the Netherlands than participants from Belgium, we conducted an ANCOVA. In this ANCOVA, we tested whether there were significant differences between two unrelated groups – in this case people from Belgium compared to the Netherlands (independent variable) – on a dependent variable – in this case the mean reaction time of female response-associated trials with a neutralization as the

stimulus (MeanRT_Nfemale). To control for other variables that might confound the results, we added the ASI score and political preference inventory score as covariates. As a result of adding these covariates, the output of the ANCOVA presented the differences in adjusted means for the two countries. Similarly, to test whether female response-associated trials with a feminization as the stimulus led to shorter reaction times in participants from Belgium than participants from The Netherlands, we conducted a second ANCOVA. The independent variable was the country the participant is from and the dependent variable was the mean reaction time of female response-associated trials with a feminization as the stimulus (MeanRT_Ffemale). We added the ASI score and political preference inventory score as covariates. In both analyses, the dependent variable and the covariates were measured on a continuous scale. For each analysis, we conducted the Shapiro-Wilk test of normality to check whether the residuals were normally distributed for each category of the independent variable. Furthermore, to test whether the covariate was linearly related to the dependent variable at each level of the independent variable, we plotted two grouped scatterplots of one of the covariates and the dependent variable, with the independent variable as a grouping variable. We tested homogeneity of regression slopes through calculating the interaction between the covariates and the independent variable. To test for homoscedasticity, we saved the standardized residuals and the unstandardized predicted values. After that, we plotted them against each other in a scatterplot. Lastly, we used Levene's test for homogeneity of variances to test homogeneity of variance.

To test whether female response-associated trials with a neutralization as the stimulus led to a higher success rate in participants from the Netherlands than participants from Belgium, we conducted a generalized linear model. In this analysis, we used a Poisson loglinear model, as the dependent variable contains count data. In this Poisson generalized linear model, we tested whether the country the participant was from (independent variable) and whether the ASI score and political preference inventory score (covariates) affected the number of correct responses participants had to female response-associated trials with a neutralization as the stimulus (Er_Nfemale; dependent variable). To test whether female response-associated trials with a feminization as the stimulus led to a higher success rate in participants from Belgium than participants from the Netherlands, we conducted a second Poisson generalized linear model. The independent variable was the country the participant is from and the dependent variable was the success rate of female response-associated trials with a feminization as the stimulus (Er_Ffemale). We added the ASI score and political preference

inventory score as covariates. In both analyses, the independent variables were measured on either a nominal, ordinal or continuous scale and all the observations were independent. For each analysis, we plotted a histogram to test whether the distribution of the dependent variables followed a Poisson distribution. Furthermore, we calculated mean and variance statistics to check whether they were similar. Lastly, dispersion of the data was checked by looking at the Value/df data.

To test whether female response-associated trials with a masculine generic as the stimulus led to longer reaction times than female response-associated trials with either neutralizations or feminizations as the stimulus, we conducted a repeated measures ANOVA. In this repeated measures ANOVA, we tested whether there were differences between three related groups, in this case the three different types of nouns. We created the factor 'TypeOfNoun' with three levels: masculine generic (MS), feminization (F) and neutralization (N). The dependent variables were the reaction times for the trials with masculine generics, neutralizations, or feminizations as the stimulus, when the same response key was used as for female names (MeanRT_MSfemale, MeanRT_Ffemale, and (MeanRT_Nfemale). We added the ASI score and political preference inventory score as covariates. The dependent variables and covariates were measured on a continuous scale. To test for normality of the distribution of the dependent variable in the related groups, we ran the Shapiro-Wilk test of normality. Furthermore, we ran Mauchly's test of sphericity to check for sphericity.

To test whether female response-associated trials with a masculine generic as the stimulus led to lower success rates than female response-associated trials with either neutralizations or feminizations as the stimulus, we conducted a generalized linear mixed model. In this analysis, we used a Poisson model, as the dependent variable contains count data. In this Poisson generalized linear mixed model, we tested whether there were differences between three related groups, in this case the three different types of nouns. Before conducting the analysis, we restructured the data, to create the variable 'TypeOfNoun', consisting of each of the three repeated measures. Consequently, the variable 'Value' was created, consisting of the corresponding values of the 'TypeOfNoun' variable. We added 'TypeOfNoun', the ASI score, and the political preference inventory score as fixed effects. We specified 'Value' as the target variable. The ASI score and political preference inventory score were measured on a continuous scale. To test whether the continuous variables were normally distributed, we made Q-Q plots for each of them. Furthermore, we calculated mean and variance statistics to check whether they were similar.

Results

After deleting observations that departed more than 2 standard deviations below or above the 25th or 75th percentiles, the final sample consisted of 31 participants, 29 of whom were from the Netherlands and 2 of whom were from Belgium. Table 1 shows a summary of the data. Table 2 shows a summary of the data separated per country. The Belgian sample was not big enough to properly test the hypotheses that compare participants from Belgium with the Netherlands. However, we still conducted the analyses as originally intended.

Table 1

Data summary (N = 31)

Variable	Minimum	Maximum	Mean	Std. Deviation
Mean reaction time neutralization trials	420.13	749.75	604.21	75.062
Mean reaction time feminization trials	483.06	752.88	592.07	74.760
Mean reaction time masculine generic trials	446.38	811.81	594.75	90.945
Mean success rate neutralization trials	14.00	16.00	15.42	0.672
Mean success rate feminization trials	13.00	16.00	15.06	0.892
Mean success rate masculine generic trials	12.00	16.00	15.23	1.087
ASI score	0.64	2.50	1.70	0.516
PPI score	2.60	4.70	3.67	0.409

Note. The reaction time and success rate variables were the means of trials in which the same response key was used as for female names. The PPI score shows the political preference inventory score.

Table 2

Data summary (N = 31) separated per country

Variable	The Netherlands				Belgium			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Mean reaction time neutralization trials	420.13	749.75	605.04	74.892	515.81	668.63	592.22	108.055
Mean reaction time feminization trials	483.06	752.88	592.18	73.706	502.25	678.63	590.44	124.716
Mean reaction time masculine generic trials	446.38	811.81	599.42	91.161	473.94	580.13	527.03	75.086
Mean success rate neutralization trials	14.00	16.00	15.41	0.682	15.00	16.00	15.50	0.707
Mean success rate feminization trials	13.00	16.00	15.07	0.923	15.00	15.00	15.00	0.000
Mean success rate masculine generic trials	12.00	16.00	15.24	1.091	14.00	16.00	15.00	1.414
ASI score	0.64	2.50	1.70	0.534	1.55	1.64	1.59	0.064
PPI score	2.60	4.70	3.66	0.415	3.80	4.10	3.95	0.212

Note. The reaction time and success rate variables were the means of trials in which the same response key was used as for female names. The PPI score shows the political preference inventory score.

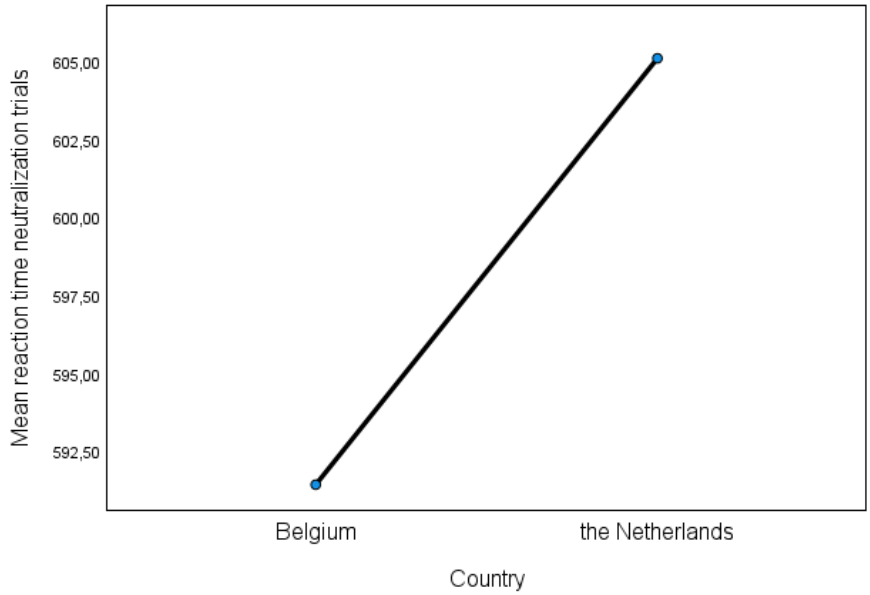
Hypothesis 1: female response-associated trials with a neutralization as the stimulus lead to shorter reaction times in participants from the Netherlands than participants from Belgium.

To test the first hypothesis, we ran an ANCOVA. To test the assumptions of an ANCOVA, we ran several tests. The Shapiro-Wilk test of normality was not significant for the Netherlands ($W(29) = .975$, $p = .691$, *ns*), therefore meeting the assumption. For Belgium there was no statistic, as the sample consisted of only two people. The interactions between the covariates and the independent variable were not significant (Country * ASI_score: $F(2,26) = .998$, $p = .382$, *ns*; Country * PPI_score: $F(2,26) = .969$, $p = .393$, *ns*), therefore meeting the assumption. Levene's test for homogeneity of variances was not significant ($F(1,29) = .212$, $p = .648$, *ns*), therefore meeting the assumption. The assumption of

homoscedasticity and linearity between the covariate and the dependent variable at each level of the independent variable were met as well. The model was not significant ($F(3,27) = .043$, $p = .988$, *ns*, partial $\eta^2 = .005$). For a plot of the estimated marginals means, see figure 4. None of the individual variables were significant either (Country: $F(1,27) = .054$, $p = .818$, *ns*, partial $\eta^2 = .002$; ASI_score: $F(1,27) = .074$, $p = .788$, *ns*, partial $\eta^2 = .003$; PPI_score: $F(1,27) = .024$, $p = .877$, *ns*, partial $\eta^2 = .001$). These results suggest that the first hypothesis was not supported.

Figure 4

Mean reaction time neutralization trials per country



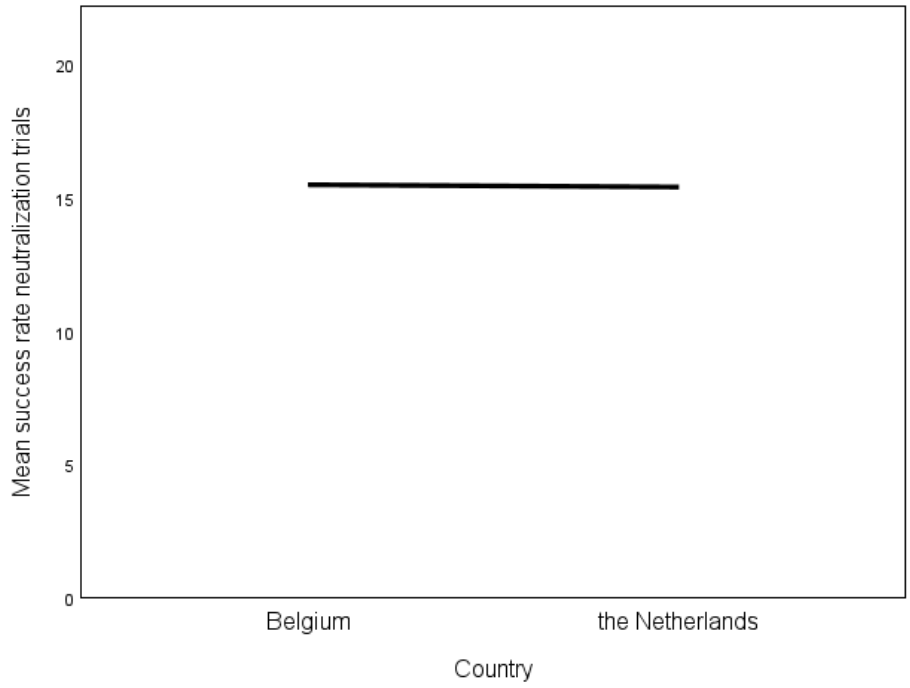
Hypothesis 2: female response-associated trials with a neutralization as the stimulus lead to a higher success rate in participants from the Netherlands than participants from Belgium.

To test the second hypothesis, we ran a Poisson generalized linear model. None of the assumptions of a Poisson generalized linear model were met. The first assumption states that the dependent variable needs to follow a Poisson distribution. We checked this assumption by making a histogram of the dependent variable. This did not follow a Poisson distribution. The second assumption stated that mean and variance statistics should be similar. This was not the case either ($M = 15.42$, variance = .452). The violation of the first two assumptions meant that the data did not follow a Poisson distribution. Lastly, the data showed under dispersion, as the Value/df was .032, instead of around 1. The model was not significant (Country: Wald Chi-Square = .003, $p = .959$, *ns*; ASI_score: Wald Chi-Square = .002, $p = .968$, *ns*; PPI_score:

Wald Chi-Square = .015, p = .902, *ns*). For a plot of the means per country, see figure 5. The exp(ß) for country was 1.010, which meant that Belgium showed an increase of 1.0% in success rate with regard to the Netherlands. These results suggest that the second hypothesis was not supported.

Figure 5

Mean success rate neutralization trials per country



Hypothesis 3: female response-associated trials with a feminization as the stimulus lead to shorter reaction times in participants from Belgium than participants from the Netherlands.

To test the third hypothesis, we ran an ANCOVA. One of the assumptions for an ANCOVA that assumes homoscedasticity was violated (see figure 6). The model was run again, this time only including ASI score as a covariate. In the new model the assumption of homoscedasticity was not violated (see figure 7). Furthermore, the Shapiro-Wilk test of normality was not significant for the Netherlands (W(29) = .941, p = .107, *ns*), therefore meeting the assumption. For Belgium there was no statistic, as the sample consisted of only two people. The interactions between the covariate and the independent variable were not significant (Country * ASI_score: F(2,28) = .161, p = .852, *ns*), therefore meeting the assumption. Levene's test for homogeneity of variances was not significant (F(1,29) = .893, p = .352, *ns*), therefore meeting the assumption. The assumption of linearity between the

covariate and the dependent variable at each level of the independent variable was met as well. The other assumptions were met as well. The model, however, was not significant ($F(2,28) = .160$, $p = .853$, *ns*, partial $\eta^2 = .011$). For a plot of the estimated marginals means, see figure 8. None of the individual variables were significant either (Country: $F(1,28) < .001$, $p = 1.000$, *ns*, partial $\eta^2 < .001$; ASI_score: $F(1,28) = .319$, $p = .577$, *ns*, partial $\eta^2 = .011$). These results suggest that the third hypothesis was not supported.

Figure 6

Scatterplot of the predicted values against the standardized residuals with both covariates included

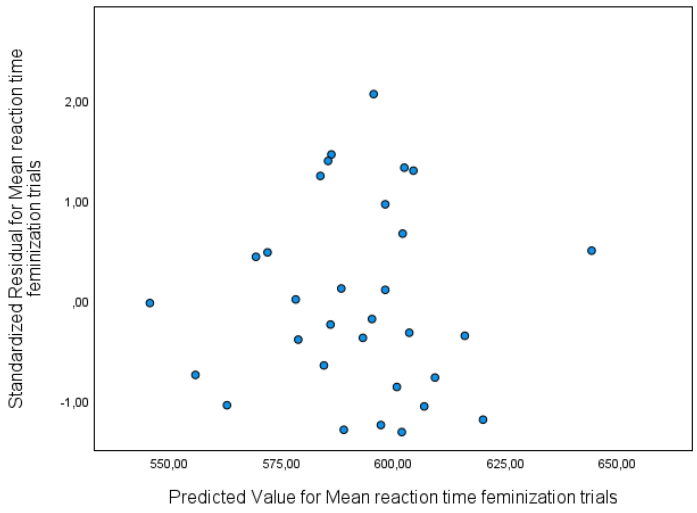


Figure 7

Scatterplot of the predicted values against the standardized residuals with only ASI score included as a covariate

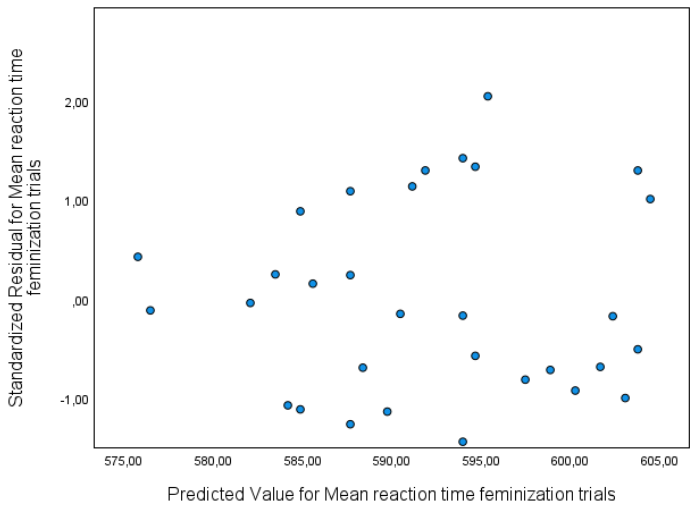
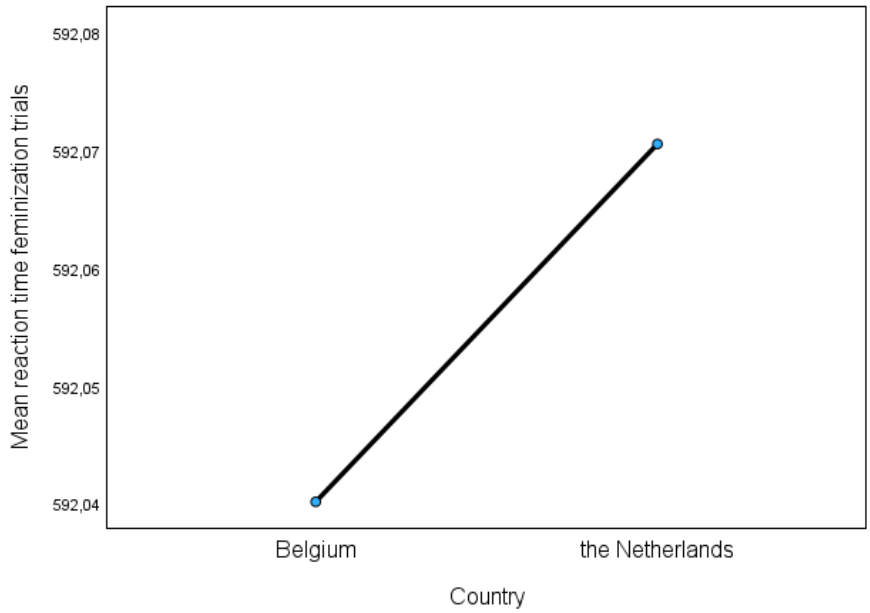


Figure 8

Mean reaction time feminization trials per country

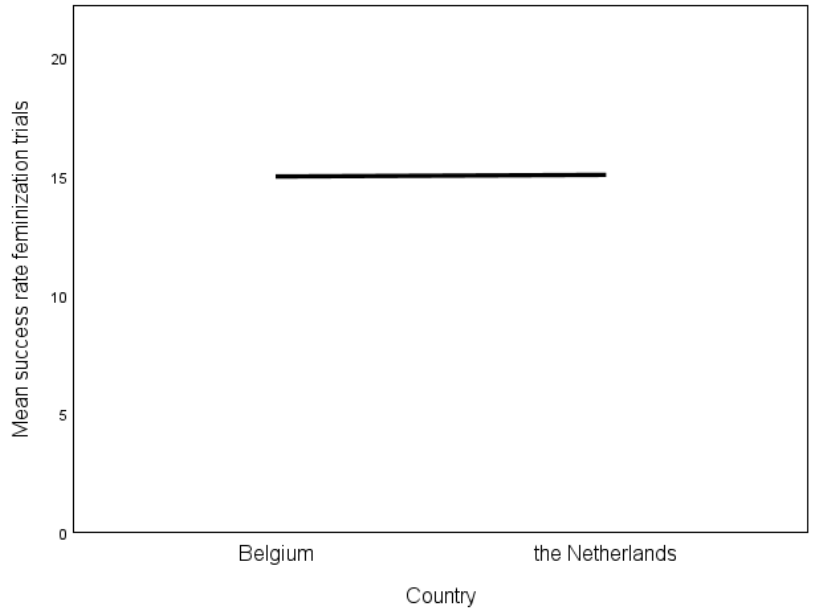


Hypothesis 4: female response-associated trials with a feminization as the stimulus lead to a higher success rate in participants from Belgium than participants from the Netherlands.

To test the fourth hypothesis, we ran a Poisson generalized linear model. None of the assumptions of a Poisson generalized linear model were met. The first assumption states that the dependent variable needs to follow a Poisson distribution. We checked this assumption by making a histogram of the dependent variable. This did not follow a Poisson distribution. The second assumption stated that mean and variance statistics should be similar. This was not the case either (*M* = 15.06, variance = .796). The violation of the first two assumptions meant that the data did not follow a Poisson distribution. Lastly, the data showed under dispersion, as the Value/df was .057, instead of around 1. The model was not significant (Country: Wald Chi-Square = .003, *p* = .959, *ns*; ASI_score: Wald Chi-Square = .036, *p* = .850, *ns*; PPI_score: Wald Chi-Square = .041, *p* = .839, *ns*). For a plot of the means per country, see figure 9. The exp(ß) for country was 0.990, which meant that Belgium showed a decrease of 1.0% in success rate with regard to the Netherlands. These results suggest that the fourth hypothesis was not supported.

Figure 9

Mean success rate feminization trials per country

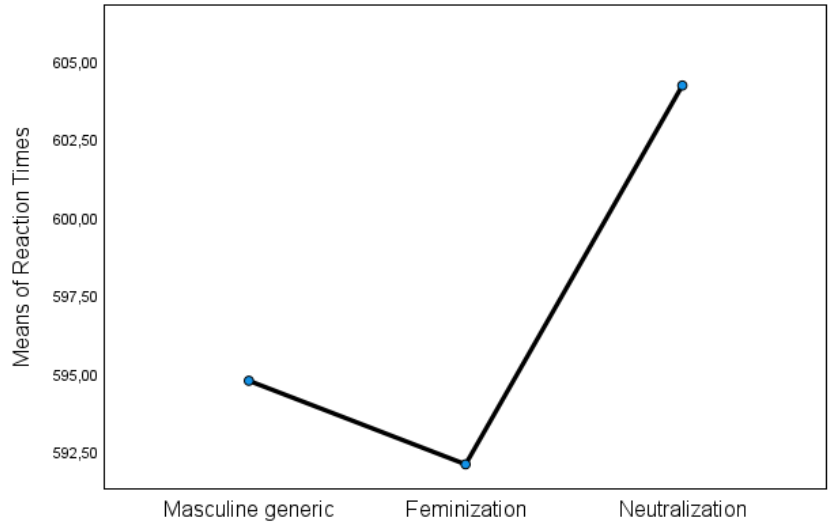


Hypothesis 5: female response-associated trials with the masculine generic as the stimulus lead to longer reaction times and a lower success rate than female response-associated trials with neutralizations or feminizations as the stimulus in participants from both countries.

To test the fifth hypothesis, we ran a repeated measures ANOVA and a Poisson generalized linear mixed model. To test the assumptions of a repeated measures ANOVA, we ran two tests. The Shapiro-Wilk test of normality was not significant on all levels (MeanRT_MSfemale: $W(31) = .950$, $p = .152$, *ns*; MeanRT_Ffemale: $W(31) = .939$, $p = .077$, *ns*; MeanRT_Nfemale: $W(31) = .975$, $p = .671$, *ns*), therefore meeting the assumption. Mauchly's Test of Sphericity was not significant either ($W(2) = .918$, $p = .313$, *ns*), therefore meeting the assumption. Thus, in the repeated measures ANOVA the statistic that assumes Sphericity was interpreted. This was not significant ($F(2,56) = .990$, $p = .378$, *ns*, partial $\eta^2 = .034$). The interaction between TypeOfNoun and the covariates were not significant either (TypeOfNoun * ASI_score: $F(2,56) = .012$, $p = .988$, *ns*, partial $\eta^2 < .001$; TypeOfNoun * PPI_score: $F(2,56) = 1.376$, $p = .261$, *ns*, partial $\eta^2 = .047$). For the estimated marginal mean reaction times of each different type of noun, see figure 10.

Figure 10

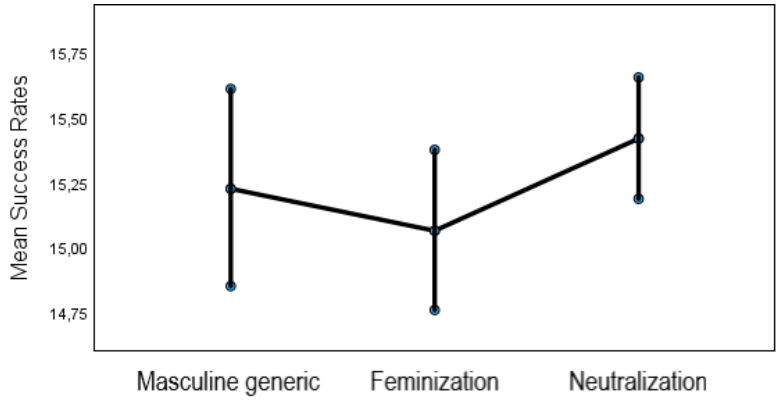
Mean reaction times for the three different types of noun



As for the second analysis, one of the assumptions was violated. Mean and variance statistics should be similar. This was not the case (M = 15.24, variance = .813). The assumption of normally distributed continuous variables was met. Since one of the assumptions was not met, we used a robust estimation of the model. This model was not significant (F(8,84) = .615, p = .763, *ns*). None of the individual variables or interactions were significant either (TypeOfNoun: F(2,84) = 1.381, p = .257, *ns*; ASI_score: F(1,84) = .316, p = .576, *ns*; PPI_score: F(1,84) = .264, p = .609, *ns*; TypeOfNoun * ASI_score: F(2,84) = .527, p = .592, *ns*; TypeOfNoun * PPI_score: F(2,84) = .997, p = .373, *ns*). For the mean success rates of each different type of noun corrected with the standard error, see figure 11. These results suggest that the fifth hypothesis was not supported.

Figure 11

Mean success rates for the three different types of noun corrected with the standard error



Discussion

The aim of this study was to investigate whether there is a difference in which gender fair alternative to the masculine generic is most effective in the Netherlands versus Belgium. Furthermore, the aim was to investigate how these different types of alternatives compared to the masculine generic to balance the mental representation of gender.

The data showed no significant results to support any of the hypotheses. An important reason for that was the nature of the sample. Four out of five hypotheses compared the Netherlands with Belgium. However, the final sample consisted only of two Belgian participants, which is not enough for a valid analysis. Furthermore, the Dutch sample was very homogeneous and consisted of almost only female participants (with the exception of one male). In addition to that, the age range was 18 to 23 years old. This homogeneous nature of the Dutch sample is a problem, as it does not accurately represent the Dutch population. It could be possible that a more variable sample would have shown differences in reaction times or success rates between genders or ages. The final Belgian sample consisted of one female and one male participant, and the age range was 25 to 47 years old. In terms of variability that is a more ideal situation. However, if the same variability was present in a bigger sample, it still might have been a problem when compared to the homogeneous Dutch sample, as the different natures of both samples pose alternative explanations for internal differences.

Despite not being significant, there were differences found between the Netherlands and Belgium, and between the three types of nouns. However, these insignificant differences only follow the predictions of one of the hypotheses. As predicted, reaction times were shorter for female response-associated trials with a feminization as the stimulus in Belgium compared to the Netherlands. The results showed the same pattern for neutralization trials, even though it was predicted that participants from the Netherlands would yield shorter reaction times. This might mean that the participants from Belgium were generally faster than the participants from the Netherlands in the task. Participants from both countries were generally faster on feminization trials compared to neutralization trials, although this difference was smaller in Belgium compared to the Netherlands. This seems to imply that for Belgian participants it does not matter which alternative to the masculine generic is used, whereas in the Netherlands feminizations would be a better alternative. The difference between the two countries was also smaller for feminization trials, so perhaps feminizations are less dependent on what kind of language it is than neutralizations. For success rate, the results showed opposite patterns as the hypotheses predicted. The last hypothesis predicted shorter reaction times and higher

success rates for both alternatives compared to the masculine generic. For both variables the results were only half in line with the hypothesis. The reaction times were shorter for feminizations compared to the masculine generic, but longer for neutralizations. For success rate it was the opposite; for neutralizations success rate was higher than for masculine generics, but for feminizations it was lower. Additionally, the results for reaction times and success rates did follow similar patterns in which type of noun yielded the lowest or highest scores. However, because of this, the results seem rather arbitrary, rather than being able to draw conclusions about them. Interestingly, for the Belgian participants the reactions times for masculine generics were shorter than both alternatives, rather than longer. This might indicate that in Belgium the masculine generic does not contribute to the problem of under-representation of women.

It could be possible that there were no significant differences between the Dutch and the Belgian sample because there may be no differences between Dutch and Belgian people after all. It was initially hypothesized that Dutch people would yield faster responses for neutralization trials and Belgian people for feminization trials, because of the difference in language. Belgian people seemed to still use the three different grammatical genders that the Dutch language used to have, whereas Dutch people only use two grammatical genders. However, it is possible that Belgian people have substituted the three-gender system for the two-gender system as well. The article that showed that the three-gender system is still found in most Belgian dialects is thirteen years old (De Vogelaer, 2010). It is conceivable that the situation has changed since then. It is also possible that the Belgian participants from our sample have substituted the three-gender system for the two-gender system. Both Belgian participants had been living in the Netherlands at the time of their participation. Possibly, they also had been for quite a while. It is therefore possible that they had been influenced by the Dutch people around them, causing them to adopt the two-gender system. This would eliminate the differences between the Belgian and Dutch sample that the hypotheses were based on. In future studies, a test should be included to determine whether the participants use two or three grammatical genders. Furthermore, for the Belgian sample it would be preferable to ask people who are living in Belgium to participate, or to include an item in the demographic questions about how long they have been living in the Netherlands. This way you could eliminate or take into account a possible influence of living in the Netherlands on the language they speak. On the other hand, there is a possibility that despite differences in

the gender system between the Netherlands and Belgium, there still are no differences in terms of mental representation elicited by the two alternatives to the masculine generic.

One possible explanation for why Dutch as spoken in Belgium seemed to tend more towards a grammatical gender language was that it had been influenced by French (Donaldson, 1983). This is a general influence on the language as a whole, but it is possible that actually speaking French has an individual difference as well. It would therefore have been useful to test or simply ask to what extent the participants from Belgium speak French. Future research could include this in their studies.

A specific problem that occurred for the hypotheses regarding success rate, was the violation of the assumptions. These suggested that the data did not follow a Poisson distribution. However, since the dependent variables contained count data, a Poisson model seemed like the best fit. Furthermore, the success rate data was not dispersed enough to draw any conclusions from. All participants had a very high success rate. This under dispersion of the data made it hard to find any differences between the two countries, or among the three different types of nouns.

The fifth hypothesis tested whether there were any differences between the masculine generic and neutralization or feminization at all. The results showed that there were no significant differences. It should therefore be considered that the masculine generic might not contribute to the linguistic under-representation of women in Dutch from a cognitive point of view. A study by Backer and Cuypere (2012) showed support for this notion. In this study, participants from Germany and Dutch speaking participants from Belgium were given a questionnaire in which they had to interpret masculine nouns by indicating what they thought was the natural gender of the referent of these nouns. In the Dutch sample the majority of the masculine nouns (63.2%) were interpreted as gender neutral. Furthermore, the number of gender neutral interpretations was higher in the Dutch sample than in the German sample. A number of empirical studies have shown that masculine generics cause the slowest or longest cognitive inclusion of women compared to feminizations or neutralization in German. However, the existing literature for Dutch mainly focuses on more theoretical perspectives rather than cognitive ones (Backer & Cuypere, 2012). The critiques that were the basis of this study were for Dutch mostly based on theoretical discussions and the cognitive effects that had been shown in other languages. For Dutch, it might therefore be possible that these findings do not apply.

This might once again have something to do with the nature of the Dutch language. Since Dutch tends more towards a natural gender language (at least in the Netherlands), this means that gender is not a critical part of it. In languages that do use a grammatical gender system with masculine and feminine gender, words are associated with one of the two genders. Therefore, words that refer to groups of people are associated with one of these two genders. This might also be part of the reason why there is a problem regarding visibility of women in the first place. Masculine generics are always at first associated with masculinity. Even neutralizations might be associated with one gender because of the assigned grammatical gender. The effect of masculine generics being associated with masculinity because of their grammatical gender could especially be stronger due to the mere presence of feminine grammatical gender in the language. Gygax and Gabriel (2008) conducted a study with native French speakers from Switzerland. One experiment consisted of two parts; in the first part the participants were presented with a grammatically masculine role name, like 'mechanics', and a kinship term, like 'brother' or 'sister'. The participants had to decide whether the person indicated by the kinship term could be part of the group indicated by the role name. In the second part, some of the role names were in the masculine form and some in the feminine form. The task remained the same. The results showed that in part one masculine role names were always less associated with women than with men, independent of the stereotypicality of the role name. This effect was even stronger in part two. This suggests that the mere presence of a feminine form in the experiment strengthens a male-biased interpretation of a masculine form. A possible reason for that could be that the presence of a feminine form is a cue that the masculine form should be interpreted specifically as well (Gygax & Gabriel, 2008). The general presence of feminine forms in a language might therefore indicate other options to refer to women than using a masculine generic, resulting in the latter to be more associated with masculinity.

In natural gender languages the possibility of alternatives to the masculine generic may be less obvious. Therefore, masculine generics could reasonably be associated with women. Additionally, because words are not assigned a grammatical gender, the initial association with a particular gender is eliminated. Interestingly, the sample used in the Backer and Cuypere (2012) paper was a sample from Belgium. If the results could be explained by the nature of the Dutch language, it might therefore show that in Belgium Dutch tends to a natural gender language as well. That could explain why there may be no differences between people from the Netherlands and Belgium as well.

Despite these alternative explanations for the results, it is interesting that this study did not find significant differences between the three types of nouns, whereas other studies did. In other studies, participants were presented with sentences containing either of the types of nouns, after which they had to fill in names (Nissen, 2013), or were presented with masculine generics and kinship names, after which they indicated whether those were compatible (Gygax & Gabriel, 2008). In these kinds of studies, inferences that are part of the mental representations associated with the stimuli were likely activated according to the memory-based approach of generating inferences. According to this approach, inferences are activated passively and may be irrelevant for text comprehension. This is further explained by the resonance model. This model suggests that concepts in active memory, like the stimuli that are read, send signals to long-term memory. Associated information stored in the long-term memory are triggered by these signals. The level of activation depends on how much the features of the stimuli correspond with the information stored in memory (Gygax et al., 2021). It is important to note that the words used in these studies are held in active memory, as the participants have to derive the meaning of the word to make a judgment of compatibility, or of which names could be associated with them. Additionally, according to the thinking-for-speaking hypothesis (Slobin, 2003), language users are forced to attend to world properties that are accentuated by the language. When these features, like grammatical gender, are activated when processing the language, people are biased in how they see the world. For grammatical gender specifically, this means that gender information is attended to, even when it is irrelevant (Gygax et al., 2021). The current study was unique in its design, as it used a modified version of the EAST. In this task, participants did not actively read the stimuli, as they are asked to focus on the color. However, much like the Stroop affect, the words are likely still read automatically, as reading is generally an automatic process (Groome, 2014). This therefore activates associated inferences. The processing of the word happens more implicitly, compared to the explicit processing in other studies. Because of this design, gender-associated suffixes can act as attentional cues towards gender, which can cause cognitive changes in the representation of reality, as contended by Whorf's (1956) linguistic relativity framework (Gygax et al., 2021). This difference in processing might explain why other studies did manage to find significant results, whereas this study did not. However, it does not explain it entirely. The modified version used in this study was based on the EAST used by De Houwer (2003). That study showed significant results, suggesting that the EAST is a valid measure of attitudes (De Houwer, 2003). The difference between that study and this one might be related to the differences in meaning of the stimuli. In a different study by de

Houwer (2001), an experiment was conducted in which the target stimuli did not differ in valence (positive or negative), like 'person' and 'animal'. The performance did not differ for these stimuli (De Houwer, 2001). Perhaps in the current study something similar is going on. Instead of using words that have different meanings as stimuli, words were used with essentially the same meaning, but in different forms. It is possible that because the words are only processed implicitly, the participant does not process them well enough to notice differences between the types of nouns. Therefore, there would not be significant differences in reaction times or success rates either.

There are some general things that could be improved about the design of the study. Something that Backer and Cuypere (2012) discussed as well is number (i.e., singular or plural) of the nouns. Singular masculine nouns were preferably interpreted as non-neutral and plural masculine nouns as neutral in both Dutch and German. In Dutch, singular nouns were interpreted as neutral in 33% of the cases, whereas plural nouns were interpreted as neutral in 93.4% of the cases (Backer & Cuypere, 2012). In this study, we only used plural nouns. It is therefore possible that we would have found significant differences between the masculine generic and its alternatives, if we had used singular nouns. Future research could therefore focus on singular nouns or compare singular with plural nouns.

Something we did not take into account either, was stereotypicality. Although we chose the stimuli in the modified EAST because they seemingly did not carry stereotype, we could still have tested for stereotypicality to make sure the participants did not perceive them as such either. Gabriel and Gyrax (2008) conducted a study on the influence of stereotypical information and masculine nouns on the representation of gender in Norwegian. Participants were presented with pairs of sentences. The first sentence of each pair introduced a group of people using a plural role name. The second sentence specified whether the group consisted of men or women. The participants had to indicate whether they thought the second sentence was a logical continuation of the first. The role names in the first sentence could be stereotypically female role names, stereotypically male role names, or neutral role names. They found that for stereotypically female role names there were more positive judgments when the second sentence specified women, then when it specified men; and for both stereotypically male and neutral role names there were more positive judgments when the second sentence specified men, then when it specified women. Thus, although there was a male bias based on the grammatical gender when the role name was not associated with any stereotypes, there was a female bias based on stereotypical information when the role name

was stereotypically female (Gabriel & Gyrax, 2008). This female bias caused the role name to have a more generic interpretation (Gabriel et al., 2018). However, you will not be able to draw conclusions about whether the masculine generic generally increases visibility of women in a certain language.

Furthermore, Gabriel et al. (2018) discussed the finding that even though neutralizations contribute to less biased representations, in the presence of gender cues like stereotypical information, other types of biases are facilitated. As a result, the gender neutrality of the word is counteracted. In this case, neutralizations might just reduce the visibility of gender biases, rather than actually correcting them (Gabriel et al., 2018), as the neutrality of the word covers up the stereotypes that are still associated with it.

Lastly, Vervecken et al. (2015) conducted a study on the influence of either feminizations or masculine generics used to describe stereotyped occupations on adolescents' perceptions of these occupations. French speaking participants from Switzerland were presented with occupations and asked to rate them on warmth, competence and perceptions of male and female success in the different occupations. The occupations were either female stereotyped, male stereotyped, or neutral. Half of the participants were presented with occupations in the feminization form and the other half with occupations in the masculine generic form. The results showed that, compared to the masculine generic, the perception that women and men are equally likely to succeed increased in the feminization form for male and female dominated occupations (Vervecken et al., 2015).

Not only does stereotypical information influence the mental representation of gender, it also differs between the different types of nouns how it influences mental representation. For the masculine generic there could be a bias based on stereotypical information for female stereotyped words, on grammatical gender for neutrally stereotyped words and on either for male stereotyped words (Gabriel & Gyrax, 2008). For male stereotyped words it is therefore not even sure where the bias comes from. For neutralizations any bias is always based on stereotypical information (Gabriel et al., 2018). For feminizations any bias based on stereotypical information seems to be attenuated, albeit that this bias is still present (Vervecken et al., 2015). Because of these differences you would not be able to draw any irrefutable conclusions about where differences in mental representation come from. And creating a gender inclusive language should be independent of other biases, like stereotypes. Additionally, a true gender inclusive language should be able to counteract biases like

stereotypes entirely. To test for that and to control for alternative explanations, a test of stereotypicality should therefore be added to a future research design.

In this study, we focused heavily on finding a way to create a more gender inclusive language, in terms of increasing visibility of women. What we did not take into account was creating a more truly gender inclusive language in terms of gender neutrality. Over the years, there has been increasingly more attention to gender diverse people. For example, more attention has been directed to school toilets, as they are identified as the least safe spaces at schools by gender diverse, but also sexuality diverse students. As a result, these students might limit how much they eat and drink throughout the day in order to avoid public toilets. This could lead to several health problems like dehydration, bladder infection, etc. Furthermore, so-called bathroom discrimination has been associated with depression and suicide attempts. Introducing gender neutral toilets at schools could prevent bullying and assault of gender and sexuality diverse students, and therefore these physical and mental health problems (Francis et al., 2022).

Similarly, there should be more attention to making language more inclusive for gender diverse people. The use of feminizations could increase visibility of women, but it causes a problem, as it only includes people who are on the gender binary. So, in your efforts of making language more gender inclusive you would still leave out a significant group of people. In 2017, the *Nederlandse Spoorwegen* (NS; 'Dutch Railways') decided to change the way they address people in their announcements because of this. Instead of 'dames en heren' ('ladies and gentlemen') they decided to address everyone with 'beste reizigers' ('dear passengers'; NOS Nieuws, 2017). This is one way to make language more inclusive. However, the word 'reiziger' is a masculine generic, as the female form 'reizigster' also exists (De Cocq & Redl, 2021). Like mentioned previously, it is possible that masculine generics are seen as gender neutral in Dutch, but it would be interesting to see if it is perceived like that for gender diverse people as well. Future research could focus on whether this or similar initiatives have the desired effect and make gender diverse people actually feel included.

In some languages gender neutral third-person pronouns have been created. For example, in English several pronouns have been suggested, like 'ze', 've', and 'xe', with 'ze' being the most well-known. Additionally, in Swedish the pronoun 'hen' was created. The problem, however, is that these new pronouns are either not known to the majority of the population or disliked (Lindqvist et al., 2019), which might mean that people refuse to use it. This way the problem of lack of inclusivity is still not corrected. Future research should

therefore focus on ways to make languages more inclusive for all gender identities, and on how to create popularity and acceptance for using these ways. Furthermore, it would be interesting to see if certain uses of language could elicit a mental presentation of people who identify as transgender, non-binary, genderfluid, et cetera, rather than just eliciting a mental representation of men or women, or even of neither specific gender. Visibility of gender diversity will increase awareness, which in turn might facilitate an increase in acceptance of gender diversity in general.

In conclusion, we found no answer to the question of whether there is a difference between the Netherlands and Belgium in which gender fair alternative to the masculine generic is the most effective. There were also no significant differences found between the three types of nouns in general. However, due to flaws in the study design and sample sizes future research should be able to draw a more reliable conclusion about that. Furthermore, it should be considered to shift the focus of future research from increasing visibility of women to increasing genuine gender neutrality in languages, to reflect society's recent increase in attention to inclusivity of gender diverse people.

References

Backer, M. D., & Cuypere, L. D. (2012). The interpretation of masculine personal nouns in German and Dutch: a comparative experimental study. *Language Sciences (Oxford), 34*(3), 253–268. https://doi.org/10.1016/j.langsci.2011.10.001

Brenner, C. J., & Inbar, Y. (2015). Disgust sensitivity predicts political ideology and policy attitudes in the Netherlands. *European Journal of Social Psychology, 45*(1), 27–38. https://doi.org/10.1002/ejsp.2072

Chen, J.-Y., & Su, J.-J. (2011). Differential sensitivity to the gender of a person by English and Chinese speakers. *Journal of Psycholinguistic Research, 40*(3), 195–203. https://doi.org/10.1007/s10936-010-9164-9

Cohen, H. (1997). Ministerin, première, bodin...Vrouwelijke beroepsnamen: Zuid versus noord? *Onze Taal, 66,* 10-12. https://www.dbnl.org/tekst/_taa014199701_01/_taa014199701_01_0006.php

De Cocq, E., & Redl, T. (2021). Neurologe of liever neuroloog?: Het effect van vrouwelijke functiebenamingen op de geloofwaardigheid van medisch specialisten. *Tijdschrift Voor Taalbeheersing (Assen), 43*(1), 35–63. https://doi.org/10.5117/TVT2021.1.002.DECO

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37*(6), 443–451. https://doi.org/10.1006/jesp.2000.1464

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*(2), 77–85. https://doi.org/10.1026//1618-3169.50.2.77

De Houwer, J., & De Bruycker, E. (2007). The implicit association test outperforms the extrinsic affective Simon task as an implicit measure of inter-individual differences in attitudes. *British Journal of Social Psychology, 46*(2), 401–421. https://doi.org/10.1348/014466606X130346

De Vogelaer, G. (2010). (Not) acquiring grammatical gender in two varieties of Dutch. In D. Geeraerts, G. Kristiansen, & Y. Peirsman (Eds.), *Advances in cognitive sociolinguistics* (pp. 167–190). De Gruyter Mouton. https://doi.org/10.1515/9783110226461.167

Donaldson, B. C. (1983). The linguistic situation in Belgium. In B. C. Donaldson (Ed.), *Dutch: A linguistic history of Holland and Belgium* (pp. 20-35). Uitgeverij Martinus Nijhof. https://www.dbnl.org/tekst/dona001dutc02_01/dona001dutc02_01_0006.php

Francis, J., Sachan, P., Waters, Z., Trapp, G., Pearce, N., Burns, S., Lin, A., & Cross, D. (2022). Gender-neutral toilets: A qualitative exploration of inclusive school environments for sexuality and gender diverse youth in Western Australia. *International Journal of Environmental Research and Public Health*, *19*(16), 10089. https://doi.org/10.3390/ijerph191610089

Gabriel, U., & Gygax, P. (2008). Can societal language amendments change gender representation? The case of Norway. *Scandinavian Journal of Psychology*, *49*(5), 451-457. https://doi.org/10.1111/j.1467-9450.2008.00650.x

Gabriel, U., Gygax, P. M., & Kuhn, E. A. (2018). Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations*, *21*(5), 844–858. https://doi.org/10.1177/1368430218771742

Gerritsen, M. (2002). Dutch: Towards a more gender-fair usage in Netherlands Dutch. In M. Hellinger & H. Bussmann (Eds.), *Gender across languages: The linguistic representation of women and men, vol. 2* (pp. 81-108), John Benjamins.

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory. *Journal of Personality and Social Psychology*, *70*(3), 491–512. https://doi.org/10.1037/0022-3514.70.3.491

Groome, D. (2014). *An introduction to cognitive psychology: processes and disorders* (3rd ed.). Routledge.

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. https://doi.org/10.1024/1421-0185.67.3.143

Gygax, P., Sato, S., Öttl, A., & Gabriel, U. (2021). The masculine form in grammatically gendered languages and its multiple interpretations: A challenge for our cognitive system. *Language Sciences (Oxford)*, *83*, 101328. https://doi.org/10.1016/j.langsci.2020.101328

Huijding, J., & De Jong, P. (2005). A modified extrinsic affective Simon task (EAST) to assess the affective value of pictorial stimuli: No influence of age and educational level. *Psychologica Belgica*, *45*(4), 241–255. https://doi.org/10.5334/pb-45-4-241

Lindqvist, A., Renström, E. A., & Gustafsson Sendén, M. (2019). Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies. *Sex Roles*, *81*(1-2), 109–117. https://doi.org/10.1007/s11199-018-0974-9

Mortelmans, T. (2008). Zij is een powerfeministe. Nog eens functie-en rolbenamingen in het

Nederlands vanuit contrastief perspectief. *Tijdschrift voor Genderstudies, 11*(1). https://ugp.rug.nl/genderstudies/article/view/1915/1908

Nissen, U. K. (1997). Do sex-neutral and sex-specific nouns exist: The way to non-sexist Spanish? In G. Linck, & U. Pasero, (Eds.), *Kommunikation von Geschlecht. Communication of gender* (pp. 222-241). Pfaffenweiler: Centaurus.

Nissen, U. K. (2013). Is Spanish becoming more gender fair? A historical perspective on the interpretation of gender-specific and gender-neutral expressions. *Linguistik Online, 58*(1), 99-117. ISSN: 1615-3014

NOS Nieuws. (2017, July 28). *Bij de NS hoor je straks geen 'dames en heren' meer*. NOS. https://nos.nl/artikel/2185349-bij-de-ns-hoor-je-straks-geen-dames-en-heren-meer

Romein-Verschoor, A. (1975). Over taal en seks, seksisme en emancipatie. *De Gids, 138*, 3-36. https://www.dbnl.org/tekst/_gid001197501_01/_gid001197501_01_0001.php

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 157–192). MIT Press.

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*, 1437–1437. https://doi.org/10.3389/fpsyg.2015.01437

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Bejamin Lee Whorf* (Edited by J. B. Carroll). MIT Press.

Appendix A

Modified EAST stimuli

White words

Lucas	Mila
Finn	Sophia
Liam	Lily
Thomas	Zoë
Oscar	Emily
Matthijs	Liv

Colored words

Leraren	Leraren en leraressen	Leerkrachten
Medewerkers	Medewerkers en medewerksters	Het personeel
Redacteurs	Redacteurs en redactrices	De redactie
Verplegers	Verplegers en verpleegsters	Verpleegkundigen

Appendix B

Political preference inventory statements

Conservative statements

Het zou onmogelijk moeten zijn om nieuwe, niet-Christelijke gebedshuizen te bouwen in België/Nederland.

Het zou mogelijk moeten zijn om mensen de doodstraf te geven.

Immigranten die een misdaad hebben gepleegd zouden terug naar hun land van oorsprong moeten worden gestuurd.

Vrouwen die voor de overheid werken zouden geen hoofddoek mogen dragen.

Mensen die meer verdienen zouden niet meer belasting moeten betalen.

Progressive statements

Kraken in leegstaande gebouwen zou niet illegaal moeten zijn.

Er moet een verbod komen op intensieve veehouderij.

Kraken is een goede oplossing voor het tekort aan betaalbare woonruimte.

De verplichte bedenktijd voor abortus moet worden afgeschaft.

Het is goed dat de overheid groene stroom subsidieert.