Research Article

Jonathan D. Kim, Anton Öttl, Pascal Gygax, Dawn M. Behne, Jukka Hyönä, Ute Gabriel*

"The Authors" Make Me Think Equally of Women and Men: Exploring Mixed-Gender Representations in a Visual Categorisation Task

https://doi.org/10.1515/psych-2022-0136
received September 7, 2023; accepted November 7, 2023

Abstract: A common goal for gender-fair language policies is to promote terms that elicit balanced activation of gender categories. Expanding previous research on the activation of feminine versus masculine categories through person nouns, we used a word-picture response priming design with gendered human faces as target stimuli, to explore whether a simultaneous activation of more than one gender category can be captured empirically. Focusing on Norwegian (Bokmål), we tested whether reading stereotypical (i.e. role nouns, e.g. "*care givers*") and categorical gendered person nouns (i.e. name pairs, e.g. "*Elin and Sandra*") facilitates the categorisation of face pairs that match the gender of the designated people. In Experiment 1 (N = 32), gender-specific (feminine or masculine) word primes were tested, before gender-balanced word primes (non-stereotyped role nouns; pairs of a female and a male name) were added in Experiment 2 (N = 39). In both experiments, the visual targets were pairs of faces (two female faces, two male faces, or one male and one female face). Consistent with previous results for English, we found gender-specific priming effects, supporting the notion that gender categories activated by linguistic stimuli may also exert influence outside of language processing. Most importantly, mixed-gender faces were successfully primed by non-stereotypical role nouns providing initial support for the idea of a balanced activation of gender categories.

Keywords: gender stereotypes, face pair task, linguistic-visual priming, social cognition, psycholinguistics

Gender-Fair Language (GFL) can broadly be defined as linguistic practices and policies that intend to remove gender discrimination (Sczesny et al., 2016). One of the common goals for GFL policies is the promotion of terms that elicit balanced activation of gender categories. This includes promoting the use of language where designations for social roles – especially job titles, but also hobbies, offices, or nationalities – do not unnecessarily make reference to specific gender(s) or group(s). Instead, in both official communications and everyday discussions, role nouns that are assumed to be equally associated with all genders should be used, such as *businesspeople* instead of *businessmen* and *cleaners* instead of *cleaning ladies*. But do the proposed alternatives really activate all gender categories equally?

*** Corresponding author: Ute Gabriel,** Department of Psychology, Norwegian University of Science and Technology, 7491 Trondheim, Norway, e-mail: ute.gabriel@ntnu.no
Jonathan D. Kim, Anton Öttl, Dawn M. Behne: Department of Psychology, Norwegian University of Science and Technology, 7491 Trondheim, Norway
Pascal Gygax: Department of Psychology, University of Fribourg, Rue P.A. de Faucigny 2, 1700 Fribourg, Switzerland
Jukka Hyönä: Division of Psychology, University of Turku, 20014 Turku, Finland
ORCID: Jonathan D. Kim 0000-0002-7926-6834; Anton Öttl 0000-0002-5711-4362; Pascal Gygax 0000-0003-4151-8255; Dawn M. Behne 0000-0002-2263-4480; Jukka Hyönä 0000-0002-5950-3361; Ute Gabriel 0000-0001-6360-4969

The search for GFL is founded on research indicating gender bias in language comprehension: even when statements are meant to refer to all genders, one gender category is often systematically activated preferentially during the comprehension process, with some research suggesting that the activation of gender categories can already take place at the word level (e.g. Cacciari & Padovani, 2007; Gygax & Gabriel, 2008; Kim et al., 2019; Oakhill et al., 2005). However, the research designs employed to investigate word-level representation of potentially less biased linguistic alternatives (e.g. Fatfouta & Sczesny, 2023; Kim et al., 2023; Lemm et al., 2005) typically do not allow for a direct test of whether there is a balanced activation of gender categories in the sense of GFL; instead, they infer this, for example, from an on average similarly strong activation of the female and male gender categories.

In an attempt to overcome this limitation of previous studies, the present study explores an alternative design to see if it is at all possible to empirically capture the simultaneous activation of gender categories. While beyond the scope of this article, such kinds of experimental tasks would provide direct benefits, such as for the evaluation of language policies that have been implemented to promote the use of a language that is not biased against gender (Gabriel et al., 2018).

For this alternative design, we opted for a cross-modal (linguistic-visual) paradigm. A uni-modal, i.e. word-to-word paradigm (such as those used by Banaji & Hardin, 1996; Oakhill et al., 2005) would have meant a reliance on gender-related words not only for activating the gender categories but also as target stimuli. This is challenging because the number of uniquely gendered and gender-balanced word clusters (e.g. wives-husbands-heterosexual couples) is very limited. Further, it has been argued (e.g. Sato et al., 2016; Wilson & Ng, 1988) that while the exclusive use of verbal stimuli informs us about linguistic processing, results obtained this way cannot necessarily be generalised to other, non-linguistic, representations of gender, since information activation could serve exclusively to establish textual coherence. Thus, as an alternative, we used a word-picture design that included a three-alternative classification task consisting of pairs of gendered faces that were either gender-matched or gender-mixed.

Since previous research has focused almost exclusively on the activation of female and male gender categories, we followed this gender binary tradition to ensure comparability of results in our new experimental task. Although in this study the gender-matched pairs therefore consisted of two female and two male faces and the mixed pairs consisted of a female and a male face, this does not preclude future research with a broader range of gender categories.

Focusing on Norwegian (Bokmål), we tested whether the reading of person nouns facilitates the categorisation of face pairs that match the gender of the designated persons (Experiments 1 and 2), with particular attention to whether gender-balanced designations facilitate responses to mixed-gender face pairs (Experiment 2). Such a facilitation effect for the categorisation of corresponding face pairs would indicate that gender information was sufficiently activated during the reading of the person nouns to influence the categorisation task. In what follows, we further explain the theoretical foundations that guide our predictions and design decisions.

1 Conceptual Background

Text comprehension is generally understood as a process of constructing a coherent mental representation of the text that goes beyond its mere linguistic representation. In this process, the surface meaning of words is merged with prior knowledge from long-term memory to form situation models, as described for example, in Kintsch's Construction-Integration Model (Kintsch, 1988; Kintsch & Van Dijk, 1978). Concerning the surface meaning of words, person nouns can linguistically refer to the gender of individuals in different ways (Motschenbacher, 2010; Stahlberg et al., 2007): *lexically*, i.e. in the semantics of nouns such as *mother*, *father*, police*woman* or police*man*, *morphologically* such as in wait*ress* (female waiter) and widow*er* (male widow) and *grammatically,* such as in French la_{FEM} *violiniste autrichienne*$_{\text{FEM}}$ and le_{MASC} *violiniste autrichien*$_{\text{MASC}}$ (*the Austrian [female] violinist* and *the Austrian [male] violinist*). The extent to which human referent gender is linguistically encoded varies greatly between languages (Dryer & Haspelmath, 2022). Further, role nouns are

often associated with a gender category as part of speakers' extra-linguistic knowledge, thus biassing word comprehension. This gender bias is especially true for many job titles and has been theorised by researchers to indicate conceptual (Irmen, 2007; Sera et al., 1994), social (Motschenbacher, 2010), typical (Egan & Perry, 2001), or stereotypical gender (Carreiras et al., 1996; Gabriel et al., 2008).

Much of the previous research has examined gender representations of person nouns in the context of discourse processing, where person nouns are inserted into short text passages (recent examples are Körner et al., 2022; Tibblin et al., 2023; Xu & Sturt, 2023). Studies conducted at the word level can be found mainly in the context of the discussion on the "automaticity" of the activation of stereotypical gender information (e.g. Banaji & Hardin, 1996; Cacciari & Padovani, 2007; Fatfouta & Sczesny, 2023; Oakhill et al., 2005) and to investigate the interplay between linguistic (grammatical, morphological) and extralinguistic (stereotypical) information (e.g. Gygax et al., 2012; Gygax & Gabriel, 2008).

In the experimental tasks typically used in this research, participants work through a sequence of trials in which two stimuli are presented either simultaneously or sequentially and to which participants make a dichotomous decision (e.g. yes vs no, same vs different). One type of task requires participants to judge whether or not the two stimuli are associated/linked (explicit association, verification, or matching tasks, e.g. Oakhill et al., 2005; Sato et al., 2016). Participants explicitly compare the stimuli and are thus encouraged to invoke (among others) gender-specific information when evaluating semantic associations.

In another type of task, only one of these stimuli – the target stimulus – must be processed in terms of a basic feature, while the other – the prime stimulus – is task-irrelevant (priming tasks, e.g. Banaji & Hardin, 1996). Such priming tasks can be further differentiated according to whether the target processing task does or does not reflect the assumed relationship between prime and target. Wentura and Degner (2010) accordingly distinguish *response priming* and *semantic priming*. Applied to the activation of gender categories by role nouns, the difference lies in whether the (gender) category that is (presumably) activated by the primary stimulus is a relevant feature for the task that is performed on the target stimulus (White et al., 2018). For example, if the primes are lexically feminine or masculine nouns, and the task is to decide whether a target pronoun is feminine or masculine (Banaji & Hardin, 1996, Exp. 1), gender category activation is immediately relevant for the task (response priming). However, if the task is to decide whether a target word is a pronoun or not (Banaji & Hardin, 1996, Exp. 2), gender category activation is not immediately relevant for the task (semantic priming). Response priming effects tend to be more robust than semantic priming effects (e.g. De Houwer, 2003; Wentura & Degner, 2010; for gender categories e.g., Kidder et al., 2018; Müller & Rothermund, 2014; White et al., 2018), which is explained by the priming effects in response priming reflecting the combined influence of memory and response activation processes (Wentura & Rothermund, 2014). Furthermore, while most of the previous research on the gender representations of role nouns exclusively relied on lexical stimuli, some studies have used a Linguistic Visual paradigm (e.g. Irmen & Köhncke, 1996; Lemm et al., 2005; Sato et al., 2016; Zacharski & Ferstl, 2023), in which participants are typically shown written text (lexical stimulus) prior to observing a visual scene (pictorial stimulus).

As we were interested in whether linguistic stimuli spontaneously activate gender categories, we chose a priming design in which participants were not explicitly asked to compare stimuli to complete the task (as in explicit association, verification, or matching tasks). More specifically, we chose a response priming design, in which response competition is expected to produce faster and more accurate responses when the target stimulus and prime activate the same response (response facilitation) and that produces slower and less accurate responses when the target stimulus and prime activate different responses (response interference; Wentura & Rothermund, 2014, p. 53).

Even though cross-modal priming effects are weaker than effects observed with same-modality pairs (e.g. Carr et al., 1982), we decided to use visual target stimuli, namely gendered human faces, to explore whether nouns can simultaneously activate more than one gender category. Using a visual categorisation task allows us to investigate whether gender categories are being activated even when they neither serve to establish textual coherence nor generally facilitate linguistic processing hence informing us about the impact that linguistic choices, such as GFL, have beyond the purely linguistic realm.

We are aware of only two studies that investigated the activation of gender information in a cross-modal response priming design. Lemm et al. (2005, Exp. 1) used gender stereotypical role nouns (e.g. "ballerina',"

"banker," or "author") and lexically gendered words (e.g. "chairwoman," "chairman," or "chairperson") to prime female, male, and neutral representations. These primes were followed by target drawings representing either a female or a male person that had to be categorised according to gender. Results showed that participants were quicker to classify targets when their gender was consistent with the preceding prime, though the contrast reached significance for female targets only. Generally, the priming paradigm employed by Lemm et al. (2005) thus demonstrates the adequacy of text-to-image priming to assess the activation of gender representations across modalities.

Discussing the relationship between prime and target stimuli, Kawakami and Dovidio (2001) argue that photographs may provide more ecologically valid stimuli than more abstract category labels, as the photographs represent actual members of the primed category. In their study (Exp. 1), positive and negative gender stereotypical trait words were used as primes (e.g. "fashionable" and "gossipy" for female traits and "brave" and "dominant" for male traits) to a categorisation task in which participants had to decide whether a yearbook photo of a college student represented a woman or a man. Results showed that participants were both quicker and better at categorising the pictures when the gender of the individual in the picture was congruent with the preceding prime, relative to when it was inconsistent.

Building off this previous research, we have designed an experimental set-up using person nouns as the prime stimuli and photographs of faces as the target stimuli. However, to explore the possibility of simultaneous activation of more than one gender category, we use *pairs* of faces as target stimuli. To match targets and primes in number, the primes also describe people in the plural.

The main prime type we use follows the perspective of GFL as equally activating feminine and masculine beliefs. Specifically, we examine stereotypically gendered pluralised role nouns as a microcosm of this conceptualisation, on the basis that normed non-gender stereotyped role nouns should equally activate both feminine and masculine gender information.

Previous research (Banaji & Hardin, 1996; Müller & Rothermund, 2014) suggests that *stereotypically* gendered person nouns produce smaller priming effects than *linguistically* (i.e. lexically, morphologically, or grammatically) gendered person nouns, and although the relevant interaction was not significant in Lemm et al. (2005, Exp. 1), there was evidence that the strength of the priming effects was related to how strongly the primes were perceived to be associated with gender. This is consistent with the notions that linguistically gendered nouns provide categorical gender information, while stereotypically gendered nouns provide probabilistic gender information (Kreiner et al., 2008, for a discussion on differences in gender representations), and that it is easier to overcome cognitive discrepancies/mismatches in gender information between probabilistic primes and categorical targets than between the categorical primes and categorical targets.

Since the sensitivity of our experimental design to detect priming effects has not yet been tested, we decided to also include a word type, which is expected to have a strong gender priming effect, namely gender-typical names. First names have traditionally been selected to indicate a person's gender. Their use allows us – by combining them into pairs – to maximise the match/mismatch of gender information between prime (name pairs) and target (face pairs), hence serving as a benchmark for the functioning of the design.

An important consideration for priming studies is the stimulus onset asynchrony (SOA; i.e. the time interval between prime onset and target onset). In a recent meta-analysis, Brysbaert (2019) found that across studies conducted in English (N = 27), the average reading speed for non-fiction was 238 words per minute, or 3.9 words per second, and that across studies conducted in Swedish (N = 5), the average reading speed (combined fiction and non-fiction) was 218 words per minute, or 3.6 words per second. As the primes used in this study were a mix of one- and three-word items, this suggests that participants require a longer SOA. Hutchison et al. (2008) conducted a study on associative strength which refers to the level of association between a specific prime and target in both directions: forward strength and backward strength. They demonstrated that forward strength was present in both short SOAs (250 ms; 200 ms prime presentation, 50 ms inter-stimulus period) and long SOAs (1,250 ms; 1,000 ms prime presentation, 250 ms inter-stimulus period); however, backward strength was only present in the long SOA. As this study involves visual targets of linguistic primes, which complicates the estimation of their associative strength, a prolonged SOA that allows for both forward and backward associative strength to inform responses is beneficial. Nevertheless,

research has indicated that 2,000 ms SOAs are excessively long, leading to conscious responses that might possibly reverse stereotype effects (e.g. Blair & Banaji, 1996). Following Hutchison et al. (2008), we implemented an SOA of 1,250 ms in our study which encompasses a 1,000 ms prime presentation and a 250 ms inter-stimulus period. This temporal arrangement met all our requirements by providing enough time for participants to read the primes, and for the backwards associative strength to facilitate responses, without inducing conscious responses.

The present study was conducted in Norwegian (Bokmål). Even though Norwegian possesses a grammatical three-gender system (masculine, feminine, and neuter), the language is experiencing the erosion of the correspondence between grammatical and referential gender for personal nouns (Bull & Swan, 2002) where lexically masculine and feminine nouns merge into a common gender (the former masculine). To illustrate, *kvinne*, the Norwegian word for *woman*, can be used with both the grammatically feminine determiner (ei_{FEM} *kvinne*) and grammatically masculine determiner (en_{MASC} *kvinne*). It is possible to morphologically mark femininity via suffixes (e.g. *skuespillerinne* [actress] instead of *skuespiller* [actor] to indicate that the actor is female) but these suffixes are no longer productive although a few remnants can still be found (e.g. *en venn – en/ei venninne [a friend], en elsker- en/ei elskerinne [a lover], en svoger – en/ei svigerinne [a sibling-in-law]).* Personal nouns assigned to the grammatical neuter class such as *barn* (child) or *menneske* (human being) remain unchanged, and there is no grammatical gender to designate people who assign themselves to neither the female nor the male gender. As such Norwegian (bokmål) is similar to English in terms of referential gender and is therefore a good choice to overcome the heavy reliance on English in experimental research (e.g. Levisen, 2019) without having to introduce grammatical aspects regarding referential gender.

To summarise, in this study, we utilise a face pair categorisation task (Face Pair Task) in a response priming setup to examine text-to-image gender priming in Norwegian. Specifically, we aim to (a) investigate the existence and reliability of gender priming effects on non-linguistic visual gendered targets and (b) explore whether mixed gender information can both prime and be primed.

2 Overview and Hypotheses

This study comprises two experiments. We first wanted to see if we could replicate previous findings with our triple-choice categorisation task (Face Pair Task). For this purpose, only feminine and masculine primes were used in Experiment 1. Two word-types, role nouns (e.g. “the nurses” [feminine], “the programmers” [masculine]), and name pairs (three words: e.g. “Ann and Catherine” [feminine], “Håvard and Hans” [masculine]) were used as primes. Since the form of these word types differs greatly, they were presented by block. Due to its exploratory nature and in line with Lemm et al. (2005), in Experiment 1, the block order (role nouns before name pairs) was fixed.

Next, we wanted to explore whether *mixed-gender* targets (i.e. mixed-gender faces) can be successfully primed. To this end, in Experiment 2, gender-balanced primes were included for both role noun primes (role nouns considered suitable roles for both men and women to hold, e.g. “the oceanographers” [non-stereotyped]) and name pairs (pairs of names composed of one female and one male name, e.g. “Ann and Hans” [gender-mixed]). To provide a within-block baseline, control primes (nonsense strings designed to mimic the role noun and name pair primes) were also included. Further, to allow for a direct comparison between role noun and name pair primes, in Experiment 2, the block order was counterbalanced between participants.

Both experiments were composed of five experimental blocks: (1) a button task, (2) the Face Pair Task without primes, (3) and (4) the critical Face Pair Task with primes, and (5) the button task again. Within each block, the trial order was randomised by participants. Responses and response times were collected. The button task was a simple task that allowed for familiarisation with the inputs needed to give responses, while the Face Pair Task without primes allowed for familiarisation with the face pair task specifically. As such, the button task can be seen to provide a measure of baseline motoric costs, while the Face Pair Task without primes provides a measure of both motoric costs and cognitive costs associated with facial perception. Data for

both experiments, and code used to explore the data, are available at [https://dataverse.no/privateurl.xhtml?token=1780b24d-1bca-45de-a699-b8c4a2302d05].

In line with previous findings (e.g. Lemm et al., 2005, Exp. 1), in this study, we hypothesise (H1) that a response facilitation effect will be found for the gender-specific face pairs following gender congruent compared to incongruent primes and (H2) that this effect will be stronger for the primes that provide categorical gender information (i.e. gender-matched name pairs) than those that provide probabilistic gender information (i.e. gendered role nouns). Regarding our research question of whether gender-mixed representations can be activated, our assumptions differ for word types. For the role noun primes, the non-stereotyped roles should broadly activate gender such that feminine and masculine information is equally activated, leading to a processing advantage compared to the feminine stereotyped and masculine stereotyped roles when responding to the mixed gender faces. We therefore hypothesise (H3) that, for the role nouns, a response facilitation effect will be found for the mixed-gender faces following gender-balanced compared to gender-specific primes. This is not necessarily the case for the name pair primes, due to match-mismatch bias (e.g. Proctor, 1981), which holds that the decision that two stimuli are similar is, in general, made more quickly and accurately than the decision that they are different. Related to our task, for the gender-specific name pairs, the gendered attributes activated by the first name are reinforced by the second name, while for the mixed-gender name pairs, these attributes are contradicted by the second name. As such, responses to the mixed gender name pairs should have higher error rates and slower response times to an as-yet-unknown degree; this is why we neither theorise for nor against a processing advantage for mixed gender compared to gender-specific face pairs.

3 Experiment 1

The data presented in Experiment 1 were recorded as part of a study exploring the adequacy of different stimulus-response setups (Öttl et al., 2023). For the present study, data from the *bimanual set-up* from the above study were used. The results of the experimental blocks without priming (i.e. Block 1, Block 2, and Block 5) are presented in Öttl et al. (2023, Bimanual Setup). In contrast, we present the data from the Face Pair Task with priming (Blocks 3 and 4), with the results of the Face Pair Task without primes (Block 2) used as a potential control variable in the data analyses,

3.1 Methods

3.1.1 Participants

Thirty-two native Norwegian-speaking participants (15 male) with a mean age of 23.8 (SD = 3.5) were recruited at the Norwegian University of Science and Technology (NTNU). Since both hands were used to register a response, both left- and right-handed participants were included (5 and 27 participants, respectively).

3.1.2 Apparatus

This experiment was run using a laptop computer in a laboratory setting. E-Prime 2.0.10.356, in association with an eye-tracking system, was used for stimulus presentation and logging responses. Responses were collected using a Chronos response box. For further details see Öttl et al. (2023).

The experimenter was seated facing the participant, with eye contact obstructed by the computer screens placed between them. This arrangement was used to offer the participant as much privacy as possible, with the experimenter still being able to indirectly monitor participants' progress through access to both their gaze

coordinates and a duplication of the stimulus screen. Participants were instructed to rest their fingers (middle and index fingers of both hands) on the response box throughout the experiment's duration.

3.1.3 Materials

For both the button task and face pair task, responses were given through four buttons, with participants pressing two designated buttons to respond to each response condition (two with either the left or right hand for the matched-response categories or one with each hand for the mixed-response condition, Figure 1). For the button task, stimuli were basic representations of the response options, with the relevant buttons to press for each trial presented on the screen. For a discussion of the adequacy of the response set up, see Öttl et al. (2023).

To control for any order effect of stimulus position, the spatial arrangement (left–right) of the stimuli and responses was counterbalanced between participants but kept constant across tasks. Therefore, stimuli were created in two sets, as described below. Participants were pseudo-randomly assigned to the spatial arrangement version to ensure roughly equal numbers of women and men in each version.

3.1.3.1 Face Pair Task

Stimuli were created from pictures of 45 female and 45 male faces taken from the Chicago Face Database (Ma et al., 2015). These were combined into three sets of face pairs: a) 15 image pairs consisting of two female faces, b) 15 pairs consisting of two male faces, and c) 15 pairs consisting of one female and one male face. The faces were pretested to ensure that female and male faces were categorised according to their respective gender category equally rapidly and that comparable pictures were assigned to the matched and mixed gender categories. Further information on this process (image selection-from-database criteria, image post-processing, and image pairing procedure) is described in the project protocol [https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/WWPPCA#]. The dimensions for each face pair stimulus were 300 × 508 pixels (8.5 cm × 14.3 cm). Examples of the faces used to produce the stimuli are presented in Figure 2. Two sets of face pairs were created to match the response setups: spatial arrangement 1 (female left, male right) and spatial arrangement 2 (male left, female right).

3.1.3.2 Primes

Primes (Appendix A, Table A1) were two sets of Norwegian person nouns: definite pluralised role nouns that were stereotypically gendered feminine (e.g. *ballettdanserne* [the ballet dancers]) or masculine (e.g. *gruvearbeiderne* [the miners]) and pairs of names that were typically female (e.g. *Anna og* [and] *Marianne*) or typically male (e.g. *Tommy og* [and] *Martin*).

Stereotypicality of the role nouns was based on explicit ratings from Misersky et al. (2014), with role nouns selected to be equally stereotypical across genders (estimated proportion of female members for feminine role nouns were $M = 0.77$, SD = 0.05 and for masculine role nouns $M = 0.21$, SD = 0.04). The definite form was chosen

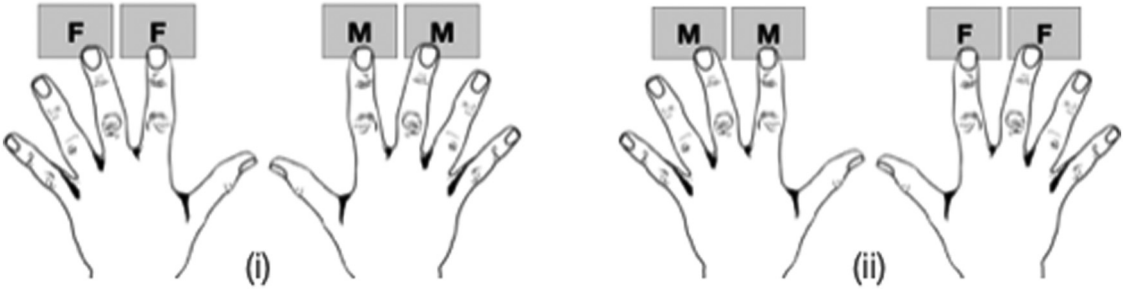

Figure 1: Bimanual response setup for two spatial arrangements: (i) response set-up 1 (FF left); (ii) response set-up 2 (MM left).

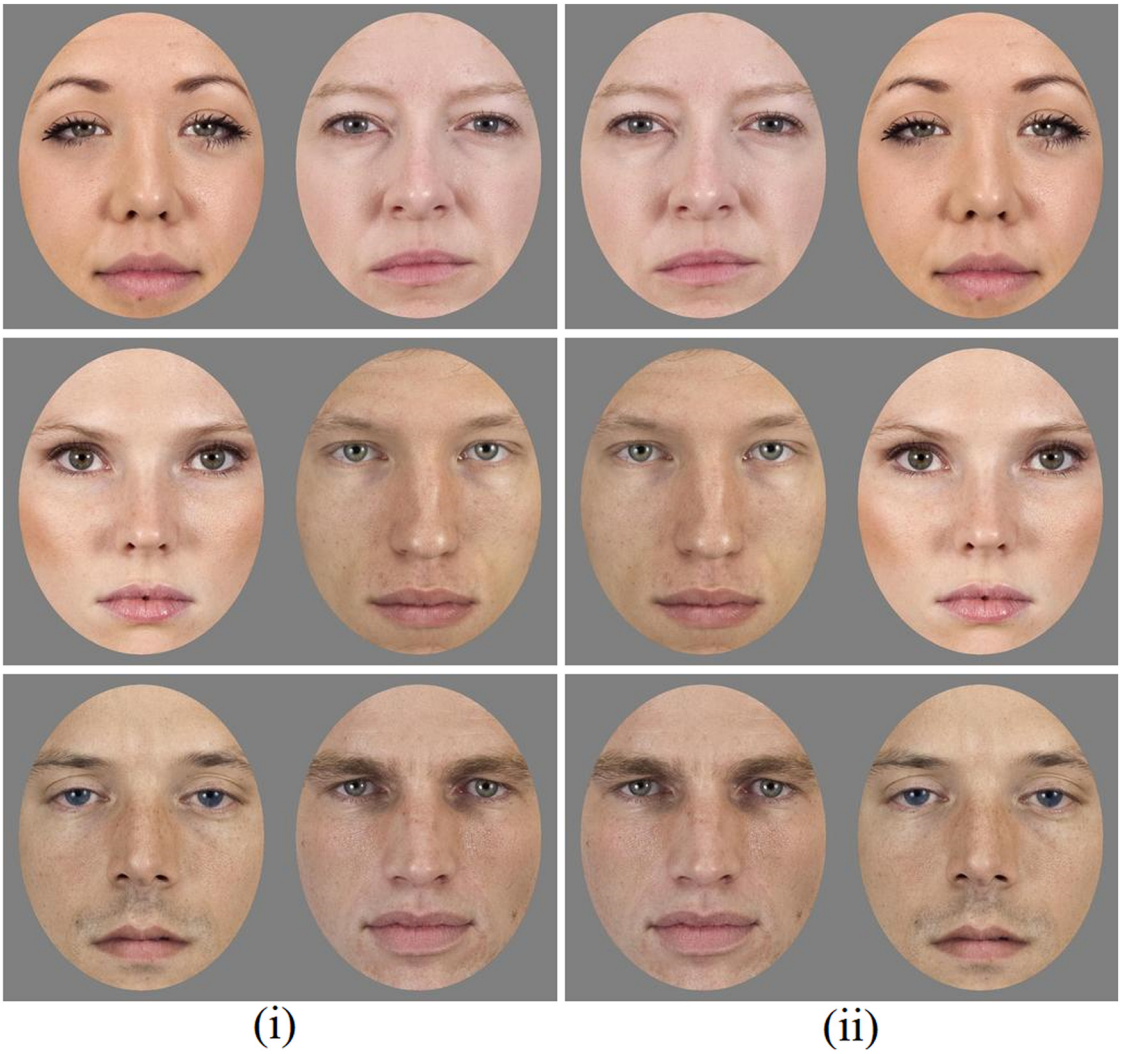

Figure 2: Examples of the Visual Stimuli Used in the Face Pair Task Note: (i) Examples from spatial arrangement 1 (female left); (ii) Examples from spatial arrangement 2 (male left). Faces obtained from the Chicago Face Database (Ma et al., 2015).

over the indefinite form as it seems more consistent with the specificity of the name pairs. Nevertheless, since definiteness is marked by suffixation in Norwegian, this feature may be less salient with words presented in isolation than would be the case in, e.g. English (e.g. *ballettdansere* [ballet dancers] vs *ballettdanserne* [the ballet dancers]).

Names were selected from the list of most common names of persons born in Norway between 1976 and 1996 (Navnestatistikken, 2022). This range of birth years was chosen to exclude names that, although very common, are predominantly associated with very young or very old people, which could lead to a perceived mismatch between face and name. Only names that are used exclusively as unisex names in the cultural context of our study were selected.

3.1.4 Procedure and Research Design

This experiment has a 3 (Face Pair Gender: female vs mixed vs male) by 2 (Prime Gender: feminine vs masculine) by 2 (Prime Type: role nouns vs name pairs) by 2 (Spatial Arrangement: female left vs male left) mixed design. Face Pair Gender, Prime Gender, and Prime Type are all within-participant factors, while Spatial Arrangement is a between-participant factor.

Participants were informed that the purpose of the experiment was to investigate attention switching between different visual tasks, especially between reading and face recognition, and that they would be undertaking five visual categorisation tasks, followed by a memory task. Participants gave written informed consent prior to participation. At experimental onset, they were presented with an information screen that restated that the purpose of the experiment was to learn more about what happens when attention shifts between different tasks, and informed them that, to this end, they would be undertaking two tasks: one consisting of categorising a picture of two faces according to the gender of the faces, and one consisting of reading words and remembering them.

Prior to the first Face Pair Task block with primes (Block 3), participants were told that they would be given the reading task in addition to the categorisation task with which they were already familiar. Specifically, they were told that they would see a word in the centre of the screen before each face pair and that they would have to remember these words as part of a memory task conducted at the end of the experiment, which assessed how many of the words the participant remembered. Prior to the second block with primes (Block 4), participants were told that they would now perform the same tasks, but with pairs of names rather than individual words for the reading task and were informed that the memory task would include both the words and the names that the participant had seen.

Following a 45-trial button block and a 45-trial face pair block, participants completed a 45-trial critical face pair block with role noun primes and a 45-trial critical face pair block with name primes, followed by a 30-trial button block (repetition). All face pair blocks were preceded by six practice trials.

Each trial in the critical blocks began with a gaze-contingent fixation cross being displayed in the centre of the screen. When the participant had fixated on the cross for 500 ms, a prime word appeared for 1,000 ms centrally on the screen, followed by a 250 ms blank screen after which a face pair was presented until the participant gave a response (with no upper time limit). Stimuli remained on the screen for 500 ms after the participant registered their response. Response and response time were collected for each trial.

Finally, participants were asked to complete a post-test during which they were presented with 40 words (20 pluralised role nouns [10 stereotypically feminine, 10 stereotypically masculine], 20 gendered names [10 typically female, 10 typically male]) in randomised order, half of which (five from each group) they had encountered as primes in the experiment. For each item, participants were asked to indicate whether they thought they had read the word during the experiment. This was included to assess whether participants were paying attention during the experiment, as, if they were, they should respond correctly above chance. After completing the post-test, participants were informed of their performance. Individual difference measures, not included in this article, were also collected. The entire experimental session lasted approximately 25–30 min. After the experiment, participants were debriefed and received a cafeteria voucher as compensation for their participation.

3.1.5 Data Preparation

Prior to analyses, both by-participant and item-by-participant data screening were used. The item-by-participant exclusion was based on both response (trials were excluded when more than two button presses were registered) and response time (trials were deselected if they were shorter than 200 ms, Baayen, 2008, or more than three standard deviations above a participant's mean in each block) and led to the removal of 3.21% of the data. By-participant exclusion was based on individual error (participants who had more than 10% of their responses to the face pair blocks removed during item-by-participant screening were deselected) and on prime recognition (deselection criteria were set at 50% or less of the items in the prime recognition task responded to correctly). No participants were deselected due to these measures.

Error rate (percentage of experimental items responded to incorrectly), and response time (speed of responses to experimental items *to which participants responded correctly*), were statistically analysed using generalised linear mixed effects modelling (GLM) in R (version 4.0.3; R Core Team, 2018). Data analysis was conducted through the *glmer* function of Version 1.1-23 of the *lme4* package (Bates et al., 2015). To prevent misattribution of order effects as differences between prime types, error rate, and response time analyses

were conducted separately for role noun and name pair primes. For all analyses, initial models were designed that captured the maximal random effects structure as justified by the design. These were composed of the experimental factors Prime Gender (feminine vs masculine) and Face Pair Gender (female vs mixed vs male), and their interaction, as well as fixed factor covariates of Participant Gender, Handedness, Age, Trial Number, and Spatial Arrangement, random intercepts of Participant, Image Shown, and Prime Shown, and random slopes of each experimental fixed factor by each random factor.

Further, to account for effects related to the motoric and cognitive costs associated with the different responses in the Face Pair Task (Öttl et al., 2023), responses to the Face Pair Task without primes (Block 2) were also included in the analyses. For each participant, the mean error rates, and response times on the Face Pair Task without primes (Block 2) were calculated per Face Pair Gender and then included in the maximal models and during model fitting as a potential random by-participant slope. To aid with comprehension, the mean value for the Face Pair Task without primes block (Block 2; overall, and per face pair condition) is reported in Table 1. Model refinement to find the model of best fit followed the approach outlined in the project protocol (https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/WWPPCA#). Comparisons between the maximal models and models of best fit for each analysis can be found in Appendix B, Tables B1 (error rate), and B2 (response time).

In line with current recommendations (Nakagawa & Schielzeth, 2013), we report conditional R^2 as an approximation of the global effect size for the best-fitting model and partial eta squared (η_p^2) as an approximation of the local effect sizes for significant main effects and interactions. Conditional R^2 was calculated using the *r.squaredGLMM()* function of the *MuMIn* package (Version 1.43.17; Bartoń, 2020), while η_p^2 was calculated using the *anova_stats* function of the *sjstats* package (Version 0.18.2; Lüdecke, 2022). Tukey- and bias-adjusted post-hoc contrasts were examined through the *emmeans()* function of the *emmeans* package (Version 1.5.0; Lenth, 2020). Bias adjustment occurred based on the estimated standard deviation of the best fitting models' residuals, calculated via the *sigma()* function of the base R package *stats.*

The analysis of the response time data is consistent with the analysis plan pre-registered with the Center for Open Science (https://osf.io/njgs4/. This preregistration is based on a larger project, and the relevant plans are described in Part II ("Gender Priming"). The analyses presented here deviate from the pre-registered plan in two respects. First, in addition to an examination of response times, we present results for error rates. Second, while not specified in the analysis plan, participant gender, participant age, and trial number were tested for inclusion as random factors and slopes.

3.2 Results

3.2.1 Error Rate

3.2.1.1 Role Noun Primes

The model of best fit (conditional R^2 = 0.17, AIC = 324 [16 lower than maximal model]) indicated no significant effects of Prime Gender, Wald $X^2 < 1$, or Face Pair Gender, Wald X^2 = 3.56, p = 0.17, η_p^2 = 0.01. Descriptively, participants produced very few errors overall (M_{ERROR} = 2.4%, SE = 0.4%).

3.2.1.2 Name Pair Primes

The model of best fit (conditional R^2 = 0.09, AIC = 304 [11 lower than maximal model]) indicated no significant effects of Prime Gender, Wald $X^2 < 1$, or Face Pair Gender, Wald X^2 = 1.34, p = 0.51, η_p^2 = 0.01. Participants tended to respond less accurately as the block progressed (Trial Number, Wald X^2 = 3.09, p = 0.08, η_p^2 = 0.01) and produced very few errors overall (M_{ERROR} = 2.2%, SE = 0.4%).

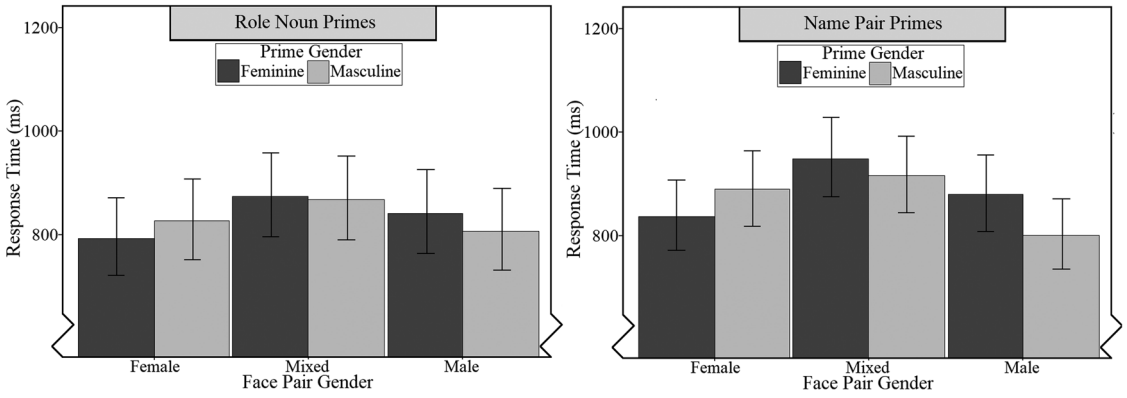

Figure 3: Estimated Means (and 95% Confidence Intervals) for Response Time (ms) as a Function of Face Pair Gender and Prime Gender for Role Noun Primes and Name Pair Primes (Experiment 1).

3.2.2 Response Time

3.2.2.1 Role Noun Primes

The model of best fit (conditional R^2 = 0.26, AIC = 17,669 [605 lower than maximal model]) indicated a small two-way interaction between Prime Gender and Face Pair Gender, Wald X^2 = 9.69, p = 0.008, η_p^2 = 0.01 (Table 2, Figure 3 [left panel]). Participants tended to respond faster to male and female faces following gender congruent compared to incongruent primes for both male (M_{DIFF} = 35 ms, z = 1.80, p = 0.04 [one-tailed]) and female faces (M_{DIFF} = 34 ms, z = 1.75, p = 0.04 [one-tailed]). Participants responded most slowly, but equally quickly, to mixed-gender faces regardless of the gender of the preceding prime, (M_{DIFF} = 6 ms, z = 0.31, p = 0.76).

3.2.2.2 Name Pair Primes

The model of best fit (conditional R^2 = 0.28, AIC = 17,716 [612 lower than maximal model]) indicated a small significant main effect of Face Pair Gender, Wald X^2 = 10.55, p = 0.005, η_p^2 = 0.02, which was qualified by a two-way interaction between Prime Gender and Face Pair Gender, Wald X^2 = 29.49, p < 0.001, η_p^2 = 0.02 (Table 2, Figure 3 [right panel]). Participants responded significantly faster to male and female faces following gender congruent compared to incongruent primes. This was more pronounced for male faces (M_{DIFF} = 78 ms, z = 3.81, p < 0.001 [one-tailed]) than for female faces (M_{DIFF} = 52 ms, z = 2.52, p = 0.006 [one-tailed]). Participants responded most slowly to mixed-gender faces regardless of the gender of the preceding prime, with participants responding non-significantly faster to male names (M_{DIFF} = 33 ms, z = 1.46, p = 0.15).

3.2.3 Summary

The results indicated very small error rates overall for both role noun and name pair primes, suggesting a potential floor effect. In line with previous research, for gender-specific face pairs, participants responded

Table 1: Mean Error Rates & Response Times per Face Pair Gender for Block 2: Face Pair Task without primes (Experiment 1)

Face Pair Gender	Mean Error Rate (%)	Mean Response Time (ms)
Female	2.1	863
Mixed Gender	3.8	982
Male	1.9	862
Overall Mean	2.6	903

Table 2: Estimated Means [and 95% Confidence Intervals] for Response Time (ms) as a Function of Face Pair Gender and Prime Gender for Role Noun Primes and Name Pair Primes (Experiment 1)

Face Pair Gender	Prime Gender		
	Feminine	Masculine	M_{DIFF}
Role Noun Primes			
Female	792 [721, 870]	826 [752, 906]	−34 [−185, 118]*
Mixed Gender	873 [796, 959]	867 [790, 952]	6 [−156, 169]
Male	841 [763, 926]	806 [732, 888]	35 [−125, 194]*
Name Pair Primes			
Female	839 [773, 911]	891 [820, 968]	−52 [−225, 91]**
Mixed Gender	951 [878, 1,031]	918 [847, 996]	33 [−118, 184]
Male	881 [810, 958]	803 [738, 873]	78 [−63, 220]***

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

faster when the gender of the primes (feminine, masculine) and faces (female, male) were congruent compared to when they were incongruent. Descriptively, the effect was stronger for the name pair primes, compared to the role noun primes, and was for name pair primes stronger when participants responded to male faces. For mixed gender face pairs, participants responded equally quickly to role noun primes regardless of prime gender, but responded faster to masculine compared to feminine name pair primes. Finally, participants responded more slowly to the mixed gender face pairs compared to the matched gender face pairs following both role noun and name pair primes.

4 Experiment 2

The goals for the second experiment were to test whether the results found for the gender-specific primes in Experiment 1 are reliable, to further investigate prime type differences (role nouns vs name pairs), and to explore whether mixed gender targets (i.e. mixed gender faces) can be successfully primed by gender-balanced primes. Therefore, the experiment's design was changed as follows: Block order (Role noun block first vs Name pair block first) was added as a factor and counterbalanced by participant; prime gender was expanded by gender-balanced primes (non-stereotypical role nouns and gender-mixed name pairs), and to provide a within-block baseline, control primes (nonsense strings) were added.

With the addition of gender-balanced and control primes, the number of trials in all three Face Pair Task blocks was increased from 45 to 60. The results of the pre-test for Experiment 1 did not support enough additional faces as being perceived as feminine or masculine to match the increased number of primes. We decided to reduce the number of faces used, and consequently, participants observed each face pair twice as often (two times per block in Experiment 2, compared to one time per block in Experiment 1). As this additional complexity may alleviate the Error Rate floor effect found in Experiment 1, both Error Rate and Response Time were examined in this experiment.

4.1 Methods

4.1.1 Participants

Thirty-nine native Norwegian-speaking participants (12 male) with a mean age of 23.1 (SD = 2.4) were recruited at the NTNU. Since both hands were used to register a response, left-handed ($N = 4$), right-handed ($N = 34$), and ambidextrous ($N = 1$) participants were included.

4.1.2 Apparatus

Apparatus was identical to that used in Experiment 1. The experimenter and participant were seated as in Experiment 1.

4.1.3 Materials

4.1.3.1 Face Pairs

Following the design rules used in Experiment 1, 30 images per gender were selected from the pre-test results and combined into three sets of face pairs for Experiment 2: a) 10 image pairs consisting of two female faces, b) 10 pairs consisting of two male faces, and c) 10 pairs consisting of one female and one male face. As in Experiment 1, two sets were prepared to account for spatial arrangement.

4.1.3.2 Primes

The same text-based priming stimuli sets as in Experiment 1 were used (Appendix A, Table A1) and complemented by gender-balanced primes (Appendix A, Table A2): for role noun primes, these were definite pluralised role nouns that were non-stereotyped, and for name pair primes, these were pairs of gender-mixed names (one typically female and one typically male). Based on the explicit assessments of Misersky et al. (2014), roles that were equally perceived as female and male were selected as non-stereotypical roles (estimated proportion of female members was $M = 0.50$, SD = 0.03).

4.1.3.3 Control primes

Nonsense strings designed to mimic both role noun primes (e.g. *ømorsnkrnveysnem*) and name pair primes (e.g. *Tinses os laga*) were used as control primes to establish a within-block baseline measure (Appendix A, Table A2). The nonsense strings were created by taking roughly equal numbers of feminine, masculine, and mixed-gender primes and, for each, randomising the letter order. For the role noun primes, the randomised string was taken as one word. For the name pair primes, the random string was split into three "words," with the middle "word" being two letters long. Following this, a native Norwegian speaker checked the nonsense strings to ensure that the randomisation had not resulted in the strings containing any Norwegian words.

4.1.4 Procedure and Research Design

This experiment had a 3 (Face Pair Gender: female vs mixed vs male) by 3 + 1 (Prime Gender: feminine vs gender-balanced vs masculine + control) by 2 (Prime Type: role nouns vs name pairs) by 2 (Block Order: role nouns first vs name pairs first) by 2 (Spatial Arrangement: female left vs male left) mixed design. Face Pair Gender, Prime Gender, and Prime Type are within-participant factors, while Block Order and Spatial Arrangement are between-participant factors.

Participants gave informed consent before undertaking the experiment. Throughout the experiment, the instructions to participants were largely the same as in Experiment 1, apart from the instructions immediately prior to the blocks with priming, which were modified to take account of the randomisation of the block order.

Notably, for the mixed primes in the name pair prime block, the order of gender presentation was consistent with the spatial arrangement of the response box. That is, if participants responded to female-specific face pairs with their left hand, the female name would be presented on the left, while if they responded to male-specific face pairs with their left hand, the male name would be presented on the left. Due to the increase in number of items responded to, the entire experimental session lasted approximately 35–40 min. After the experiment, to test for a potential influence on participants' awareness of the purpose of

the experiment, participants were asked what they thought the aim of the experiment was and were then debriefed and given a voucher for a local shopping centre as compensation for their participation.

4.1.5 Data Preparation

Participants' responses to the question about the aim of the experiment were coded as "understood" (n = 0), "somewhat understood" (n = 5), "partially understood" (n = 7), and "not understood" (n = 26) by two of the experimenters (inter-rater agreement Cohen's K = 0.81, z = 6.96, p < 0.001, with only four mismatches that were then resolved in conversation). Because of the large differences in size between the groups, the data were combined into a two-level factor: those who did not identify (n = 26) and those who at least partially identified (n = 12) the actual goal of the experiment.

Prior to analyses, the same item-by-participant and by-participant data screening as was used in Experiment 1 was applied to the data set. For the item-by-participant exclusion, this accounted for 3.46% of the data. For the by-participant exclusion, one participant was removed as their performance in the prime recognition task was below 50%. As such, the final dataset was composed of 38 participants (11 male; 3 left-handed, 34 right-handed, 1 ambidextrous) with a mean age of 23.1 (SD = 2.4). In keeping with Experiment 1, the analysis focused on Error Rate (percentage of experimental items responded to incorrectly) and Response Time (speed of responses to face pairs to which participants responded correctly).

The statistical analyses conducted were the same as in Experiment 1 apart from three important changes; first, Prime Type (role noun vs name pair) and Block Order were included as experimental fixed factors in model fitting. Second, nonsense text was included as a "prime gender" level. And lastly, Identification of Experimental Aim was included as a Fixed Factor Covariate. The mean values for the Face Pair Task without primes block (Block 2; overall, and per face pair condition) are reported in Table 3. Comparisons between the maximal models and models of best fit for each analysis can be found in Appendix B, Tables B3 (error rate), and B4 (response time).

4.2 Results

4.2.1 Error Rate

The model of best fit (conditional R^2 = 0.34, AIC = 898 [50 lower than maximal model]) indicated no significant effects of Face Pair Gender, Wald X^2 = 3.94, p = 0.14, η_p^2 = 0.01, Prime Gender, Wald X^2 = 6.00, p = 0.11, η_p^2 = 0.01, or Prime Type, Wald X^2 = 1.44, p = 0.23, η_p^2 < 0.01. Participants tended to respond less accurately as the block progressed (Trial Number, Wald X^2 = 3.05, p = 0.09, η_p^2 < 0.01) and produced very few errors overall (M_{ERROR} = 2.9%, SE = 0.3%).

4.2.2 Response Time

The model of best fit (conditional R^2 = 0.24, AIC = 57,684 [390 higher than maximal model, which reached singularity so was ignored]) indicated small significant effects of Face Pair Gender, Wald X^2 = 54.90, p < 0.001, η_p^2 = 0.02, Prime

Table 3: Mean Error Rates and Response Times per Face Pair Gender for Block 2: Face Pair Task Without Primes (Experiment 2)

Face Pair Gender	Mean Error Rate (%)	Mean Response Time (ms)
Female	1.9	927
Mixed Gender	6.9	1,070
Male	1.6	888
Overall Mean	3.5	962

Gender, Wald $X^2 = 53.32$, $p < 0.001$, $\eta_p^2 = 0.01$, and Prime Type, Wald $X^2 = 33.70$, $p < 0.001$,and $\eta_p^2 = 0.01$. These were qualified by a small significant interaction between Face Pair Gender and Prime Gender, Wald $X^2 = 43.04$, $p < 0.001$, $\eta_p^2 = 0.01$, and a very small significant interaction between Prime Gender and Prime Type, Wald $X^2 = 16.16$, $p = 0.001$, $\eta_p^2 < 0.01$, which were qualified in turn by a very small significant three-way interaction between Face Pair Gender, Prime Gender, and Prime Type, Wald $X^2 = 13.79$, $p = 0.032$, $\eta_p^2 < 0.01$ (Table 4, Figure 4).

For role noun primes, participants responded equally quickly to female faces following feminine, masculine, and gender-balanced primes – in all cases, this was significantly faster than the nonsense text baseline (feminine $M_{DIFF} = 81$ ms, $z = 3.44$, $p = 0.003$; masculine $M_{DIFF} = 80$ ms, $z = 3.34$, $p = 0.005$; gender balanced $M_{DIFF} = 75$ ms, $z = 3.17$, $p = 0.008$). Participants responded significantly faster to male faces following gender congruent compared to both incongruent primes (feminine $M_{DIFF} = 79$ ms, $z = 3.51$, and $p = 0.033$; gender balanced $M_{DIFF} = 71$ ms, $z = 3.19$, and $p = 0.008$) and to the nonsense text baseline ($M_{DIFF} = 61$ ms, $z = 2.72$, $p = 0.033$). Finally, for the mixed gender faces, participants responded significantly faster following gender-balanced primes compared to the nonsense text baseline ($M_{DIFF} = 62$ ms, $z = 2.65$, $p = 0.041$) and non-significantly faster following feminine ($M_{DIFF} = 26$ ms, $z = 1.03$, $p = 0.731$) and masculine ($M_{DIFF} = 49$ ms, $z = 1.93$, $p = 0.217$) primes.

For name pair primes, participants responded significantly faster to female faces following gender congruent primes compared to incongruent primes (masculine $M_{DIFF} = 78$ ms, $z = 3.37$, $p = 0.004$; gender balanced $M_{DIFF} = 68$ ms, $z = 6.03$, $p < 0.001$) and to the nonsense text baseline ($M_{DIFF} = 154$ ms, $z = 6.46$, $p < 0.001$). Participants responded significantly faster to male faces following gender congruent compared to gender-balanced primes ($M_{DIFF} = 91$ ms, $z = 3.88$, $p = 0.001$) and tended to respond more quickly to male faces following gender congruent compared to feminine primes ($M_{DIFF} = 44$ ms, z = 1.79, $p = 0.28$) and to the nonsense text baseline ($M_{DIFF} = 48$ ms, $z = 2.08$, $p = 0.159$). For the mixed gender faces, participants responded non-significantly faster following gender-balanced primes compared to the nonsense text baseline ($M_{DIFF} = 61$ ms, $z = 2.19$, and $p = 0.127$), similarly fast compared to the feminine primes ($M_{DIFF} = 1$ ms, $z = 0.03$, and $p > 0.999$), and slower compared to the masculine primes ($M_{DIFF} = 55$ ms, $z = 2.12$, and $p = 0.148$). Interestingly, for the mixed-gender faces, participants responded significantly faster following masculine compared to the nonsense text baseline ($M_{DIFF} = 116$ ms, $z = 4.25$, and $p < 0.001$).

4.2.3 Summary

Reproducing the findings from Experiment 1, the results indicated very small error rates overall, suggesting a potential floor effect, even with the additional complexity from this experiment. The response time results were complex. For the role noun primes, participants responded significantly faster to male faces paired with masculine primes than to any of the other prime genders, and the nonsense text baseline, responded significantly faster to mixed gender faces paired with gender-balanced primes compared to the nonsense text

Table 4: Estimated Means [and 95% Confidence Intervals] for Response Time (ms) as a Function of Face Pair Gender, Prime Type, and Prime Gender (Experiment 2)

	Prime Gender			
Face Pair Gender	**Feminine**	**Gender-Balanced**	**Masculine**	**Nonsense**
Role Noun Primes				
Female	890[785, 1,010]$_a$	896[790, 1,017]$_a$	891[785, 1,012]$_a$	971[856, 1,102]$_b$
Mixed	989[875, 1,118]$_{ab}$	963[852, 1,089]$_a$	1,012[895, 1,144]$_{ab}$	1,031[912, 1,165]$_b$
Male	936[824, 1,065]$_a$	928[816, 1,055]$_a$	857[754, 974]$_b$	918[807, 1,043]$_a$
Name Pair Primes				
Female	878[774, 996]$_a$	1,022[901, 1,160]$_b$	956[843, 1,086]$_c$	1,032[910, 1,172]$_b$
Mixed	1,038[918, 1,174]$_{ab}$	1,039[919, 1,174]$_{ab}$	984[871, 1,112]$_a$	1,100[972, 1,244]$_b$
Male	930[818, 1,057]$_{ab}$	980[862, 1,114]$_a$	889[782, 1,010]$_b$	937[824, 1,065]$_{ab}$

Note. Per row, means sharing a common subscript are not significantly different at $\alpha = 0.05$ (two-tailed) according to contrast differences.

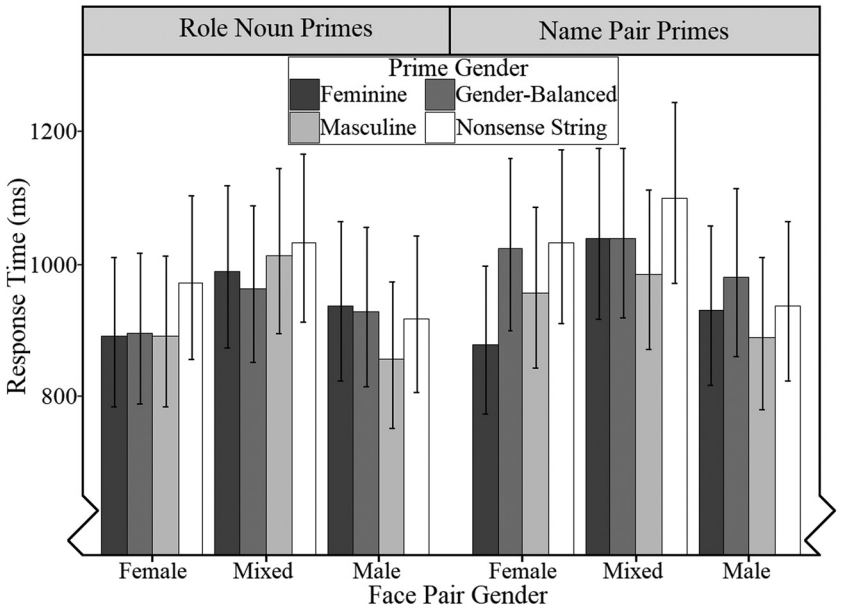

Figure 4: Estimated Means [and 95% Confidence Intervals] for Response Time (ms) as a Function of Face Pair Gender, Prime Type, and Prime Gender (Experiment 2).

baseline, and responded equally quickly to the female faces regardless of the prime gender – in all cases, significantly faster than the nonsense text baseline. For the name pair primes, participants responded faster to the female and male faces when paired with congruently gendered primes than when paired with incongruent primes and the nonsense text baseline. This effect was much stronger for the primes paired with the female compared to male faces. For the mixed gender faces, participants tended to respond more quickly to masculine primes than to the other prime gender categories – and responded significantly faster following masculine primes compared to the nonsense text baseline. In keeping with Experiment 1, the effects were descriptively larger for the name pair primes compared to the role noun primes.

5 General Discussion

This study was aimed at exploring the possibility of simultaneous activation of more than one gender category following the perspective of GFL, which promotes language that seeks to avoid unnecessary references to a specific gender. We employed a text-to-image response priming design utilising a three-alternatives classification task that complements the usual specifically female and specifically male categories with a gender-mixed (female and male) category. Due to the very exploratory nature of this study, the first experiment focused on gender-specific priming effects, and the second incorporated an examination of gender priming for primes that do not correspond to an either-or gender paradigm.

In line with previous research applying a similar design (e.g. Kawakami & Dovidio, 2001; Lemm et al., 2005) and consistent with H1, faster response times were found for the female- and male-specific face pairs following gender congruent compared to incongruent primes (Exp. 1 [all] and Exp. 2 [all except for female face pairs following stereotypical primes]). Also, consistent with H2, the results for both experiments indicated stronger facilitation effects for the primes that provide categorical gender information (i.e. gender-matched name pairs) than those that provide probabilistic gender information (i.e. gendered role nouns). Overall, these results suggest that non-linguistic targets are successfully primed by categorical female and categorical male name pairs and show evidence of successfully being primed by probabilistic female and probabilistic male role nouns (especially, in Exp. 2, compared to the nonsense text baseline).

With reference to the gender-mixed category, in Experiment 1, responses to gender-mixed face pairs were slower than those to gender-matched face pairs – uniformly for role noun primes, but slower for feminine compared to masculine name pair primes. When the gender-balanced primes were incorporated in Experiment 2, consistent with H3, responses to gender-mixed face pairs were faster when preceded by non-stereotypical role noun primes compared to gender-stereotypical role noun primes and the nonsense text strings. When preceded by gender-mixed name pairs, responses to gender-mixed face pairs were faster compared to the nonsense text strings, but similarly fast compared to feminine name pairs and even slower compared to masculine name pairs. This suggests that mixed gender targets are successfully primed by probabilistically gender-balanced information and show some benefit from the presentation of categorical gender-balanced information, but that categorical masculine information is perceived as more relevant.

The lack of a response time priming effect in Experiment 2 for female face pairs following a role noun prime was unexpected. The response times for female face pairs following a stereotypical prime were relatively quick, with feminine, masculine, and gender-balanced primes all resulting in significantly faster responses compared to the nonsense text strings. This contrasts with the results of Experiment 1.

Since far more women than men participated in Exp. 2 while participant gender was balanced in Exp. 1, the difference in findings may indicate an in-group bias, i.e. that participants implicitly consider members of their own gender as more capable of holding any role. However, if this were the case, one could expect participant gender to significantly improve the model of best fit during model fitting, which was not found.

The difference in the findings could also be related to the different number of prime types per experiment. The use of only female and male primes in Experiment 1 may have led to a contrast effect, whereas the addition of gender-balanced primes and nonsense text strings in Experiment 2 may have led to more complex strategies involving other factors such as humanness. In addition, one-third of the prime-target combinations were congruent in Experiment 1, but only one-quarter in Experiment 2, so the role of congruency proportion would need to be explored in future studies. With this in mind, one can speculate whether the finding that women – when gender-balanced primes and nonsense text strings are included – are perceived as capable of holding roles regardless of these roles' gender stereotypicality can be explained as a combination of historical gender roles and modern approaches to improving gender equality.

A common approach to improving gender equality in role selection (especially occupation selection) is to convey messages (Moser & Branscombe, 2023) aimed at affirming that the role can be held by anyone regardless of gender; e.g. girls can be doctors, and boys can be nurses. Historically, there have been far more "men-only" compared to "women-only" roles, meaning that the majority of gender equality messaging by necessity has reinforced the capability of women to hold any role (reflected in the equal response times between all three occupation gender stereotype categories), seemingly without reaching the same tipping point for the capability of men to hold any role. If this is true, then it can also be seen as an indicator that GFL policies (in this case, the renunciation of morphological female markers in Norwegian) can positively impact how social perceptions change over time and suggests that it would be beneficial to investigate how the perception of men's abilities to hold counter-stereotyped roles may also be shifted.

Returning to the mixed gender targets for the name pairs in Experiment 2, the finding that participants responded most quickly following masculine name pairs, and equally quickly following feminine and gender-mixed name pairs, with responses to all three being faster than responses to the nonsense strings, is broadly in keeping with the expected match-mismatch bias. In other words, it is possible that the names provided a general processing advantage compared to the nonsense strings, with a second processing advantage for masculine compared to feminine names. It is possible that a third processing advantage of the conceptual match between the gender-mixed names and mixed-gender faces was masked by the match-mismatch effect as participants had to spend more cognitive energy, and therefore time, in correctly interpreting the prime before responding. As such, it would follow that if single words (e.g. gendered personal pronouns) were used as categorical primes instead, thus avoiding the match-mismatch bias, the priming effect of the categorical gender-balanced primes might be much stronger. However, since the number of these forms per category is very limited, each prime would need to be repeated multiple times. This would potentially introduce a repetition bias, would strengthen the bias that each individual prime would have as an outlier, and would severely complicate comparisons of the categorical and stereotypical primes.

In relation to the question of why masculine name pair primes would lead to an improvement in response times for mixed-gender face pairs, some research (e.g. Koivula, 2001) has found that masculine items create far richer cognitive representations than feminine items. As the gender-balanced name pair primes contained one male and one female name, it could be that this provides a comparative processing time advantage when observing ambiguous visual targets.

The priming effects found in both experiments were relatively large compared to similar priming studies (e.g. Kawakami & Dovidio, 2001; Lemm et al., 2005). It is possible that this is related to the length of the SOA, as longer SOAs (such as those chosen by us) generally appear to produce larger effects (e.g. Avneon & Lamy, 2018; Greenwald et al., 1996). It would be interesting to investigate whether a shorter SOA, more in line with conventional response priming studies but, following Hutchison et al. (2008), lacking backwards associative strength, would yield the same or different results. If comparable results were found, they would suggest a strong forward prime-target associative strength – powerful enough to surpass any backward associative strength effects – and would allow for an assessment of semantic feature (e.g. gender) strength for specific primes (e.g. role nouns, first names) via manipulation of SOA. Another possible explanation for these comparatively large priming effects is the experimental design; the experimental setup modelled the face pair task as closely as possible, and stimulus selection and development was also based on tasks using response time measures, which may provide more accurate measures than commonly used approaches such as explicit ratings. A further possible explanation is that this experiment increased in difficulty across experimental tasks, which would allow participants to become familiarised with the task. However, the effect of Trial Number was only strong enough to be added to the models of best fit for the Error Rate analyses, suggesting that any effect of Trial Number on Response Time would be minimal. Another possible explanation is that the inclusion of the data from the Face Pair Task without primes (Block 2) in the analyses may have removed a large amount of noise related to idiosyncratic biases (e.g. between-participant differences in which face gender is easier to recognise) or personal factors (e.g. handedness, factors related to motoric ease of response registration).

As the findings of this study indicate that non-stereotyped role nouns successfully prime mixed-gender targets, but that gender-mixed first name pairs appear less effective than masculine information in priming mixed-gender targets, these results have practical and theoretical implications for gender-fair language policies and open the door for further examinations of how different forms of a) gender-balanced information, and b) truly gender-neutral information affects our perceptions. Research based on linguistic-visual paradigms can thus complement current research on the gender associations of new linguistic forms, which relies heavily on estimation tasks (e.g. Pozniak et al., 2023; Schunack & Binanzer, 2022; Tibblin et al., 2023; Xiao et al., 2023) and reading tasks (e.g. Stetie & Zunino, 2022; Tibblin et al., 2023). Further, as exemplified by Sato et al. (2016), linguistic-visual paradigms are very usable in cross-linguistic experimentation, as the same target stimuli can be used across many different language contexts, allowing for more direct comparisons to be drawn. Similarly, there is the potential for these advantages to be extended to comparisons between spoken and written modes of language comprehension, and to access priming effects across different levels of literacy (e.g. due to differences in age or educational opportunities). As such, these findings indicate that the presented three categorisation-alternatives face pair task is well positioned to be used in exploring the impacts of GFL cross-culturally and cross-lingually.

In conclusion, our results replicate gender-specific priming effects for both gender categorical primes (name pairs) and gender probabilistic primes (gender stereotyped role nouns), with the stronger priming effect found for the categorical primes, as expected. The results also indicated that non-stereotyped role nouns successfully primed responses to mixed-gender faces, although the same was not found for the gender-mixed name pairs. As the non-stereotyped role nouns can be interpreted as a microcosm of GFL focusing on equally activated gender beliefs, these results offer some preliminary support for the idea that GFL can positively impact our perceptions of and interactions with others outside of the linguistic domain and support the use of the face pair task as a method of further exploration. This study provides important insights into how linguistic information can subtly influence our perceptions of the people around us. While these findings cannot be used directly to improve gender-sensitive language, they provide a starting point on which future interventions can build.

Acknowledgements: We would like to express our gratitude to Martine Aune, Ane-Kristine Byrvik, and My Nguyen for collecting the data.

Funding information: This research was funded by the Norwegian Research Council, with the funding awarded to UG [PI], PG, DB, and JH. The grant number for this research is FRIPRO 240881. The Norwegian Research Council website is https://www.forskningsradet.no/. The Norwegian Research Council had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

Author contributions: The following list of contributions follows the CRediT Taxonomy of author's individual contributions to the work. *Jonathan D. Kim:* data curation, formal analysis, investigation, methodology, software, validation, visualisation, roles/writing – original draft, and writing – review & editing. *Anton Öttl:* Conceptualisation, data curation, formal analysis, investigation, methodology, software, validation, visualisation, roles/writing – original draft, and writing – review & editing. *Pascal Gygax:* conceptualisation, funding acquisition, methodology, and writing – review & editing. *Dawn M. Behne:* conceptualisation, funding acquisition, methodology, and writing – review & editing. *Jukka Hyönä:* conceptualisation, funding acquisition, methodology, and writing – review & editing. *Ute Gabriel:* conceptualisation, funding acquisition, methodology, project administration, supervision, validation, roles/writing – original draft, and writing – review & editing.

Conflict of interest: The authors have no competing interests to declare that are relevant to the content of this article.

Compliance with ethical standards: In this research, no personally identifiable information and no sensitive personal data were gathered. All participants gave their written informed consent for inclusion before they participated in the study. This study was conducted in accordance with the Declaration of Helsinki and the guidelines of the Norwegian Committees for Research Ethics (https://www.forskningsetikk.no/en/). According to Norwegian law, this research did not have to go through a process of ethical review in Norway.

Data availability statement: The dataset and analysis code generated during and/or analysed during the current study can be found at https://doi.org/10.18710/WGSPJG.

References

Avneon, M., & Lamy, D. (2018). Reexamining unconscious response priming: A liminal-prime paradigm. *Consciousness and Cognition*, *59*, 87–103. doi: 10.1016/j.concog.2017.12.006.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x.

Bartoń, K. (2020). *MuMIn: Multi-Model Inference (1.43.17). CRAN*. [Computer software]. https://cran.r-project.org/src/contrib/Archive/MuMIn/.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*(6), 1142–1163. doi: 10.1037/0022-3514.70.6.1142.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 1–30. doi: 10.1016/j.jml.2019.104047.

Bull, T., & Swan, T. (2002). The representation of gender in Norwegian. In *Gender across languages: The linguistic representation of women and men* (pp. 219–249). John Benjamins Publishing Company.

Cacciari, C., & Padovani, R. (2007). Further evidence of gender stereotype priming in language: Semantic facilitation and inhibition in italian role nouns. *Applied Psycholinguistics*, *28*(2), 277–293. doi: 10.1017/S0142716407070142.

Carr, T. H., McCauley, C., Sperber, R. D., & Parmelee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(6), 757–777. doi: 10.1037/0096-1523.8.6.757.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A*, *49*(3), 639–663. doi: 10.1080/713755647.

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition*, *31*(3), 353–359. doi: 10.3758/BF03194393.

Dryer, M., & Haspelmath, M. (2022). *The World Atlas of Language Structures Online* (v2020.3) [dataset]. Zenodo. doi: 10.5281/ZENODO.7385533.

Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, *37*(4), 451–463. doi: 10.1037/0012-1649.37.4.451.

Fatfouta, R., & Sczesny, S. (2023). Unconscious Bias in Job Titles: Implicit associations between four different linguistic forms with Women and Men. *Sex Roles*. doi: 10.1007/s11199-023-01411-8.

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

Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., & Oakhill, J. (2008). Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods*, *40*(1), 206–212. doi: 10.3758/BRM.40.1.206.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science*, *273*(5282), 1699–1702. doi: 10.1126/science.273.5282.1699.

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/Schweizerische Zeitschrift Für Psychologie/Revue Suisse de Psychologie*, *67*(3), 143–151. doi: 10.1024/1421-0185.67.3.143.

Gygax, P., Gabriel, U., Lévy, A., Pool, E., Grivel, M., & Pedrazzini, E. (2012). The masculine form and its competing interpretations in French: When linking grammatically masculine role names to female referents is difficult. *Journal of Cognitive Psychology*, *24*(4), 395–408. doi: 10.1080/20445911.2011.642858.

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *Quarterly Journal of Experimental Psychology (2006)*, *61*(7), 1036–1066. doi: 10.1080/17470210701438111.

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, *36*(6), 431–456. doi: 10.1007/s10936-007-9053-z.

Irmen, L., & Köhncke, A. (1996). Zur Psychologie des 'generischen'Maskulinums. [On the psychology of the 'generic' masculine.]. *Sprache & Kognition*, *15*, 152–166.

Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin*, *27*(2), 212–225. doi: 10.1177/0146167201272007.

Kidder, C. K., White, K. R., Hinojos, M. R., Sandoval, M., & Crites, S. L. (2018). Sequential stereotype priming: A meta-analysis. *Personality and Social Psychology Review*, *22*(3), 199–227. doi: 10.1177/1088868317723532.

Kim, J., Angst, S., Gygax, P., Gabriel, U., & 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), 1–26. doi: 10.1017/S095926952200014X.

Kim, J., 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*(9), e0221802. doi: 10.1371/journal.pone.0221802.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. doi: 10.1037/0033-295X.95.2.163

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. doi: 10.1037/0033-295X.85.5.363.

Koivula, N. (2001). Perceived characteristics of sports categorized as gender-neutral, feminine and masculine. *Journal of Sport Behavior*, *24*, 377–393.

Körner, A., Abraham, B., Rummer, R., & Strack, F. (2022). Gender Representations Elicited by the Gender Star Form. *Journal of Language and Social Psychology*, *41*(5), 553–571.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*(2), 239–261. doi: 10.1016/j.jml.2007.09.003.

Lemm, K. M., Dabady, M., & Banaji, M. R. (2005). Gender picture priming: It works with denotative and connotative primes. *Social Cognition*, *23*(3), 218–241. doi: 10.1521/soco.2005.23.3.218.

Lenth, R. (2020). *Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.0*. [Computer software]. https://CRAN.R-project.org/package=emmeans.

Levisen, C. (2019). Biases we live by: Anglocentrism in linguistics and cognitive sciences. *Language Sciences*, *76*, 101173. doi: 10.1016/j.langsci.2018.05.010.

Lüdecke, D. (2022). *sjstats: Statistical Functions for Regression Models (Version 0.18.2)* [Computer software]. Zenodo. doi: 10.5281/zenodo.1489175.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. doi: 10.3758/s13428-014-0532-5.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T., Englund, K., Hanulikova, A., Öttl, A., Valdrova, J., Stockhausen, L. V., & Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, *46*(3), 841–871. doi: 10.3758/s13428-013-0409-z.

Moser, C. E., & Branscombe, N. R. (2023). Communicating inclusion: How men and women perceive interpersonal versus organizational gender equality messages. *Psychology of Women Quarterly*, *47*(2), 250–265. doi: 10.1177/03616843221140300.

Motschenbacher, H. (2010). *Language, gender and sexual identity: Poststructuralist perspectives*. John Benjamins Publishing.

Müller, F., & Rothermund, K. (2014). What does it take to activate stereotypes? Simple primes don't seem enough: A replication of stereotype activation. *Social Psychology*, *45*(3), 187–193. doi: 10.1027/1864-9335/a000183.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. doi: 10.1111/j.2041-210x.2012.00261.x.

Navnestatistikken. (2022). *SSB*. Retrieved 6 September 2023, from https://www.ssb.no/befolkning/navn/statistikk/navn.

Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition*, *33*(6), 972–983. doi: 10.3758/BF03193206.

Öttl, A., Kim, J. D., Behne, D. M., Gygax, P., Hyönä, J., & Gabriel, U. (2023). Exploring the comparative adequacy of a unimanual and a bimanual stimulus-response setup for use with three-alternative choice response time tasks. *PLOS ONE*, *18*(3), e0281377. doi: 10.1371/journal.pone.0281377.

Pozniak, C., Corbeau, E., & Burnett, H. (In Press). *Contextual dilution in French gender inclusive writing: An experimental investigation. Journal of French Language Studies*. doi: 10.31234/osf.io/hjfc4.

Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychological Review*, *88*(4), 291–326. doi: 10.1037/0033-295X.88.4.291.

R Core Team. (2018). *R: A Language and environment for statistical computing* (4.0.3) [Computer software]. https://www.R-project.org/.

Sato, S., Gygax, P. M., & Gabriel, U. (2016). Gauging the impact of gender grammaticization in different languages: Application of a linguistic-visual paradigm. *Language Sciences*, *7*, 140. doi: 10.3389/fpsyg.2016.00140.

Schunack, S., & Binanzer, A. (2022). Revisiting gender-fair language and stereotypes – A comparison of word pairs, capital I forms and the asterisk. *Zeitschrift Fur Sprachwissenschaft*, *41*(2), 309–337. Scopus. doi: 10.1515/zfs-2022-2008.

Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, *7*, 1–11. https://www.frontiersin.org/articles/ 10.3389/fpsyg.2016.00025.

Sera, M. D., Berge, C. A. H., & Pintado, J. D. C. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*(3), 261–292. doi: 10.1016/0885-2014(94)90007-8.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In *Social Communication* (K. Fiedler, pp. 163–187). Psychology Press.

Stetie, N. A., & Zunino, G. M. (2022). Non-binary language in Spanish? Comprehension of non-binary morphological forms: A psycholinguistic study. *Glossa: A Journal of General Linguistics*, *7*(1), Article 1. doi: 10.16995/glossa.6144.

Tibblin, J., Granfelt, J., van de Weijer, J., & Gygax, P. (2023). The male bias can be attenuated in reading: On the resolution of anaphoric expressions following gender-fair forms. *Glossa Psycholinguistics*, *2*(1), 1–33. doi: 10.5070/G60111267.

Tibblin, J., van de Weijer, J., Granfeldt, J., & Gygax, P. (2023). There are more women in joggeur·euses than in joggeurs: On the effects of gender-fair forms on perceived gender ratios in French role nouns. *Journal of French Language Studies*, *33*(1), 28–51. doi: 10.1017/S0959269522000217.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). The Guilford Press.

Wentura, D., & Rothermund, K. (2014). Priming is not priming is not priming. *Social Cognition*, *32*, 47–67. doi: 10.1521/soco.2014.32.supp.47.

White, K. R. G., Danek, R. H., Herring, D. R., Taylor, J. H., & Crites, S. L. (2018). Taking priming to task. *Social Psychology*, *49*(1), 29–46. doi: 10.1027/1864-9335/a000326.

Wilson, E., & Ng, S. H. (1988). Sex bias in visual images evoked by generics: A new zealand study. *Sex Roles*, *18*(3), 159–168. doi: 10.1007/BF00287786.

Xiao, H., Strickland, B., & Peperkamp, S. (2023). How fair is gender-fair language? Insights from gender ratio estimations in French. *Journal of Language and Social Psychology*, *42*(1), 82–106. doi: 10.1177/0261927X221084643.

Xu, Z., & Sturt, P. (2023). The influence of stereotypical information on gender inference in Chinese. *SN Social Sciences*, *3*(2), 30. doi: 10.1007/s43545-023-00618-6.

Zacharski, L., & Ferstl, E. C. (2023). Gendered representations of person referents activated by the nonbinary gender star in German: A word-picture matching task. *Discourse Processes*, *60*, 294–319. doi: 10.1080/0163853X.2023.2199531.

Appendix A Primes Used in Experiment 1 and Experiment 2

Table A1: Feminine and masculine primes included in Experiment 1 and Experiment 2. For stereotypical primes, gender stereotypicality norms from Misersky et al. (2014) are listed as the estimated proportion of female members of the respective role

Feminine primes			**Masculine primes**		
Role Noun Primes		**Name Pair Primes**	**Role Noun Primes**		**Name Pair Primes**
Norwegian	***English Translation***		**Norwegian**	***English Translation***	
Advokatsekretærene (0.71)	*The law clerks*	Anette og Siri	Astronautene (0.20)	*The astronauts*	Geir og Kjetil
Aleneforeldrene (0.76)	*The single parents*	Anita og Silje	Bilmekanikerne (0.16)	*The car mechanics*	Håkon og Anders
Ballettdanserne (0.77)	*The ballet dancers*	Ann og Cathrine	Børsmeklerne (0.25)	*The stockbrokers*	Håvard og Hans
Barnepasserne (0.79)	*The nannies*	Anna og Marianne	Dataekspertene (0.21)	*The computer specialists*	Jan og Øystein
Barneskolelærerne (0.76)	*The primary school teachers*	Anne og Kristine	Drosjesjåførene (0.22)	*The taxi drivers*	Joakim og Erik
Barnevaktene (0.78)	*The babysitters*	Elin og Sandra	Fabrikkbestyrerne (0.25)	*The factory managers*	John og Vegard
Blomsterhandlerne (0.72)	*The florists*	Hege og Camilla	Fotballspillerne (0.22)	*The football players*	Jon og Andreas
Cheerleaderne (0.89)	*The cheerleaders*	Heidi og Marte	Fyrvokterne (0.24)	*The lighthouse keepers*	Jonas og Daniel
Danseinstruktørene (0.77)	*The dance instructors*	Ida og Karoline	Gruvearbeiderne (0.12)	*The miners*	Kenneth og Per
Fødselshjelperne (0.80)	*The birth attendants*	Ingrid og Tina	Lokomotivførerne (0.21)	*The engine-drivers*	Knut og Jørgen
Førskolelærerne (0.79)	*The nursery teachers*	Ingvild og Linn	Lydteknikerne (0.24)	*The sound engineers*	Kristian og Ole
Helsearbeiderne (0.76)	*The health visitors*	Julie og Stine	Mekanikerne (0.19)	*The mechanics*	Magnus og Espen
Husholderne (0.79)	*The housekeepers*	Linda og Malin	Motorsyklistene (0.23)	*The bikers*	Marius og Eirik
Håndleserne (0.75)	*The palm readers*	Line og Kristin	Pantelånerne (0.26)	*The pawnbrokers*	Mats og Øyvind
Kosmetikerne (0.85)	*The beauticians*	Mari og Marthe	Presidentene (0.13)	*The presidents*	Morten og Bjørn
Manikyristene (0.88)	*The manicurists*	Marie og Hanne	Privatdetektivene (0.26)	*The private detectives*	Petter og Lars
Mannekengene (0.72)	*The fashion models*	Marit og Lise	Programmererne (0.23)	*The programmers*	Robert og Rune
Omsorgsyterne (0.77)	*The caregivers*	Nina og Cecilie	Redningsarbeiderne (0.26)	*The rescue workers*	Simen og Thomas
Ridelærerne (0.72)	*The horse trainers*	Sara og Helene	Rørleggerne (0.14)	*The plumbers*	Stian og Henrik
Sekretærene (0.75)	*The secretaries*	Therese og Lene	Skogsforvalterne (0.22)	*The forest rangers*	Tom og Sindre
Sykepleierne (0.78)	*The nurses*	Tone og Monica	Treskjærerne (0.22)	*The wood carvers*	Tommy og Martin
Synkronsvømmerne (0.78)	*The synchronised swimmers*	Tonje og Hilde	Trommeslagerne (0.21)	*The drummers*	Tor og Fredrik
Tannpleierne (0.71)	*The dental hygienists*	Trine og Maria	Vektløfterne (0.20)	*The weightlifters*	Trond og André

Table A2: Gender-Balanced primes and Nonsense Text strings included in Experiment 2 only. For role noun primes, gender stereotypicality norms from Misersky et al. (2014) are listed as the estimated proportion of female members of the respective role. Name pair primes are presented in this table in line with Spatial arrangement 2 (male left)

Gender-Balanced Primes			Nonsense Text Strings	
Role Noun Primes		Name Pair Primes	Role Noun Primes	Name Pair Primes
Norwegian	***English Translation***			
Astrologene (0.50)	*The astrologers*	Andre og Nina	Dreleneriræ	Agtdes hi Ionic
Bankkassererne (0.52)	*The bank clerks*	Aslak og Sara	Ebersnpaernsa	Airsabl je Ollag
Butikkeierne (0.46)	*The boutique owners*	Christoffer og Line	Endernendbegarrisi	Amom tg Orymon
Campingturistene (0.45)	*The campers*	David og Cathrine	Ermdeylmeujmr	Anadan rs Teago
Demonstrantene (0.46)	*The protestors*	Gustav og Mona	Gosheejllrett	Anehtebh ol Selegia
Forfatterne (0.51)	*The authors*	Henrik og Kine	Junsiroeralt	Auran ri Aoldgo
Fotgjengerne (0.52)	*The pedestrians*	Jarl og Isabelle	Ketbuekirei	Fivelegg oj Nonisid
Hotellgjestene (0.46)	*The hotel guests*	Karl og Annette	Knæerjertres	Gamav ns Tuoog
Jobbsøkerne (0.49)	*The job seekers*	Kjetil og Christina	Laenreedorlfnee	Gego sr Lpie
Journalistene (0.53)	*The journalists*	Knut og Hanne	Mainklnkeiebre	Gitenan no Øysas
Jurymedlemmene (0.53)	*The jurors*	Kristoffer og Maria	Noerafeøtpteleorr	Goved ki Renridfei
Kinogjengerne (0.51)	*The cinema goers*	Lars og Cecile	Økelnrotiemorvof	Grueg ni Inord
Kiropraktorene (0.46)	*The chiropractors*	Magnus og Lisa	Ømorsnkrnveysnem	Horgne jo Naads
Nyhetsoppleserne (0.52)	*The news readers*	Ole og Tonje	Orntdsnemetra	Invesos ag Amra
Oseanografene (0.45)	*The oceanographers*	Patrik og Stine	Orshebeandtmnlelr	Mneo va Egarr
Psykiaterne (0.49)	*The psychiatrists*	Robert og Marit	Pervtenaitvekiedt	Nertrito sg Iaksvi
Rettspsykiaterne (0.45)	*The forensic psychologists*	Rune og Ingrid	Rarogselot	Noni ra Deagn
Rettsstenografene (0.52)	*The court reporters*	Sigmund og Silje	Ratoffrtee	Ogresa ic Elcl
Sektmedlemmene (0.52)	*The cult members*	Sindre og Kristine	Rikeenmenak	Opnes ic Geehlam
Slektningene (0.50)	*The relatives*	Stephen og Tone	Seotanækekvtdraerek	Senisert ki Goirdn
Telefonoperatørene (0.50)	*The telephone operators*	Thomas og Hilde	Tarpleånneen	Taijga ol Nkoerak
Tennisspillerne (0.49)	*The tennis players*	Trond og Linn	Tislnelolerfpabl	Tinses os Laga
Varamedlemmene (0.55)	*The deputies*	Øystein og Sanna	Uhrelhndoes	Tooge te Epdsrnr

Appendix B Model Comparisons for the Maximal Models and Models of Best Fit

Table B1: Model comparisons between maximal model and model of best fit for error rate analysis in Experiment 1

Aspect	Maximal Model	Final Model Role Noun Primes	Final Model Name Pair Primes
Fixed Factors: Main Effects	Face Pair Gender, Prime Gender	Face Pair Gender, Prime Gender	Face Pair Gender, Prime Gender
Fixed Factors: Interaction Effects	Face Pair Gender by Prime Gender		
Fixed Factors: Covariates	Age, Participant Gender, Spatial Allocation, Handedness, Trial Number		Trial Number
Random Intercepts	Participant, Specific Image, Specific Prime	Specific Image	Specific Prime
Random Slopes	Face Pair Gender by Participant, Prime Gender by Participant, Control Variable by Participant		
AIC	Role Noun Primes: 340 Name Pair Primes: 315	324	304
Errors prohibiting or limiting use of model	Role Noun Primes: Reached Singularity Name Pair Primes: Reached singularity		

Table B2: Model comparisons between maximal model and model of best fit for response time analysis in Experiment 1

Aspect	Maximal Model	Final Model Role Noun Primes	Final Model Name Pair Primes
Fixed Factors: Main Effects	Face Pair Gender, Prime Gender	Face Pair Gender, Prime Gender	Face Pair Gender, Prime Gender
Fixed Factors: Interaction Effects	Face Pair Gender by Prime Gender	Face Pair Gender by Prime Gender	Face Pair Gender by Prime Gender
Fixed Factors: Covariates	Age, Participant Gender, Spatial Allocation, Handedness, Trial Number		
Random Intercepts	Participant, Specific Image, Specific Prime	Specific Image, Specific Prime	Specific Image, Specific Prime
Random Slopes	Face Pair Gender by Participant, Prime Gender by Participant, Control Variable by Participant	Control Variable (logarithmically transformed) by Participant	Control Variable (logarithmically transformed) by Participant
AIC	Role Noun Primes: 18,274 Name Pair Primes: 18,328	17,669	17,716
Errors prohibiting or limiting use of model	Role Noun Primes: Failure to Converge Name Pair Primes: Failure to Converge		

Table B3: Model comparisons between maximal model and model of best fit for error rate analysis in Experiment 2

Aspect	Maximal Model	Final Model
Fixed Factors: Main Effects	Face Pair Gender, Prime Gender, Prime Type	Face Pair Gender, Prime Gender, Prime Type
Fixed Factors: Interaction Effects	All	
Fixed Factors: Covariates	Age, Participant Gender, Spatial Allocation, Handedness, Trial Number, Experimental Aim	Trial Number
Random Intercepts	Participant, Specific Image, Specific Prime	Specific Prime
Random Slopes	Face Pair Gender by Participant, Prime Gender by Participant, Prime Type by Participant, Control Variable by Participant	Control Variable by Participant
AIC	966	898
Errors prohibiting or limiting use of model	Reached Singularity	

Table B4: Model comparisons between maximal model and model of best fit for response time analysis in Experiment 2

Aspect	Maximal Model	Final Model
Fixed Factors: Main Effects	Face Pair Gender, Prime Gender, Prime Type	Face Pair Gender, Prime Gender, Prime Type
Fixed Factors: Interaction Effects	All	All
Fixed Factors: Covariates	Age, Participant Gender, Spatial Allocation, Handedness, Trial Number, Experimental Aim	
Random Intercepts	Participant, Specific Image, Specific Prime	
Random Slopes	Face Pair Gender by Participant, Prime Gender by Participant, Prime Type by Participant, Control Variable by Participant	Control Variable (logarithmically transformed) by Participant
AIC	57,285	57,684
Errors prohibiting or limiting the use of model	Failure to Converge	