

Social Psychological Bulletin

The Readability of the Non-Binary Gender Star in German: Evidence From a Lexical Decision Task

Lisa Zacharski¹ , Alexandra Kruppa¹ , Evelyn C. Ferstl¹ 

[1] Department of Psychology, Center for Cognitive Science, University of Freiburg, Freiburg, Germany.

Social Psychological Bulletin, 2025, Vol. 20, Article e13719, <https://doi.org/10.32872/spb.13719>

Received: 2024-01-14 • Accepted: 2024-09-25 • Published (VoR): 2025-06-02



Handling Editors: Carmen Cervone, Department of Developmental Psychology and Socialisation, University of Padova, Padova, Italy; Anne Maass, Department of Developmental Psychology and Socialisation, University of Padova, Padova, Italy; Division of Science, NYU Abu Dhabi, Abu Dhabi, United Arab Emirates; Jennifer Lewendon, Division of Science, NYU Abu Dhabi, Abu Dhabi, United Arab Emirates

Corresponding Author: Lisa Zacharski, University of Freiburg, Department of Psychology, Center for Cognitive Science, Hebelstraße 10, 79104 Freiburg im Breisgau, Germany. lisa.zacharski@cognition.uni-freiburg.de

Related: This article is part of the SPB Special Topic "The Gendered Language (R)Evolution: New Insights Into the Ever-Evolving Interaction Between Gender and Language", Guest Editors: Carmen Cervone, Jennifer Lewendon, & Anne Maass, Social Psychological Bulletin, 20, <https://doi.org/10.32872/spb.v20>

Supplementary Materials: Code, Data, Materials [see Index of Supplementary Materials]



Abstract

The non-binary gender star in German (e.g., Radfahrer*in - cyclist) is intended to inclusively address *all* genders, that is, persons identifying beyond a female-male dichotomy, as well as women and men. Critics of this gender-fair form claim that, because it is not in line with German orthography, it impedes the readability of texts. Experimental research on this claim is still scarce. Because word recognition is a crucial component of the reading process, we developed a lexical decision task to investigate lexical access to role nouns in star form with a student (Experiment 1: 97 participants, 18–29 years) and a non-student sample (Experiment 2: 80 participants, 30–80 years), thus taking interindividual differences into account. Our results are promising for proponents of the star form: First, we found that less than 3% of all participants rejected star nouns as German words. Second, amongst the remaining participants, students accepted star nouns as quickly and as often as feminine and masculine forms. In contrast, non-students accepted star nouns more slowly and less often. However, the non-students' initial difficulties in lexical access reflected in slower reaction times were overcome quickly over the course of the experiment thus suggesting that the readability of the gender star is a matter of familiarity and practice.



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

Keywords

gender-fair language, non-binary gender forms, visual word recognition, interindividual differences, psycholinguistics

Highlights

- The study is the first to use a lexical decision task to assess the readability of the nonbinary gender star in German.
- Word recognition of gender-inclusive words using the non-orthographic gender star is effortless for students, while initial difficulties among older, non-student participants disappear quickly.
- The gender star is an effective tool to reduce the male bias evoked by generically masculine forms without compromising readability.

The common traffic sign “Radfahrer bitte absteigen” (*Cyclists [masc.] please dismount*) illustrates the generic use of grammatically masculine role nouns in German: The feminine form *Radfahrerinnen* addresses female cyclists only, while the masculine form is supposed to address *all* cyclists. However, research has shown that using masculine forms whenever the gender of the referent(s) is unknown or irrelevant leads to a male bias (e.g., Gygax et al., 2021). A rather new, but popular gender-fair form intended to reduce this bias is the gender star (*Genderstern*) (Krome, 2020). Its proponents suggest that a sign displaying “Radfahrer*innen bitte absteigen” (*Cyclists [star] please dismount*) would address all people, that is, persons identifying beyond a female-male dichotomy, as well as women and men (Diewald & Steinhauer, 2017). In contrast, the Council for German Orthography¹ disapproves of the star for not aligning with German orthography and impeding the readability of texts (Rat für deutsche Rechtschreibung, 2021). Experimental research on this claim is still scarce. Because word recognition is a crucial component of the reading process (e.g., Coltheart, 2006), we developed a lexical decision task to investigate the readability of role nouns in star form with an implicit measure. Data from two experiments will be presented: Experiment 1 was conducted with a sample of students (18–30 years). Experiment 2 was conducted with a more heterogenous group of older non-students (30–80 years).

Gender-Fair Language in German

In English, a natural gender language (Gygax et al., 2019), most personal nouns are gender neutral. In contrast, in German, a grammatical gender language, all nouns carry

1) The Council for German Orthography (*Rat für deutsche Rechtschreibung*) is an intergovernmental body for Standard High German orthography. Its official spelling rules are the main reference for questions on this topic.

grammatical gender (masculine, feminine, neuter). Dependent forms, such as articles and pronouns, have to correspond. For the majority of role nouns, there are two forms: a masculine (e.g., *der Radfahrer* – *the (male) cyclist* [masc.]) and a feminine form. The latter is usually derived from the former by adding the feminine suffices *-in* [sing.] or *-innen* [plural] (e.g., *die Radfahrerin* – *the female cyclist* [fem.]) (Diewald & Steinhauer, 2017). As illustrated by the example above, feminine and masculine forms are used asymmetrically: Feminine forms only refer to women, while masculine forms can be used to refer to men (*specific use*) and to referents whose gender is unknown or irrelevant (*generic use*)². Since the 1980s, feminist linguists have argued that this asymmetry leads to an underrepresentation of women in mental representations (e.g., Pusch, 1984). Their claims found support in psycholinguistic studies showing that the generic use of masculine forms leads to a male bias. In contrast, gender-fair alternatives can increase the mental inclusion of women (e.g., Braun et al., 1998; Gygas et al., 2008; Körner et al., 2022; Sato et al., 2016; Stahlberg & Sczesny, 2001).

Two strategies of gender-fair language (GFL) can be distinguished (cf. Gabriel et al., 2018): Feminization makes the inclusion of women explicit (e.g., binary pair forms: *die Studentin* oder *der Student* – *the female student* [fem.] or *the male student* [masc.], or their abbreviations, e.g., the capital-I: *FlugbegleiterIn* – *stewardEss*). Neutralizations remove any clues to the referent's gender (e.g., nominalized participles: *Studierende* – *those who study*).

With the growing acknowledgement of the existence of gender identities beyond a female-male dichotomy, additional non-binary gender forms (GFs) have been introduced. They are intended to refer to *all* genders (Diewald & Steinhauer, 2017). The most popular non-binary GF and the one central to the public debate (Krome, 2020; Meuleneers, 2024) is the gender star. It has gained popularity particularly since 2018, when a third gender option (*diverse*) was introduced into the German personal status law in addition to the existing binary options (*female* and *male*). It is created by adding an asterisk between a role noun's stem and its feminine suffix (e.g., *Radfahrer*in* – *cyclist*). It intends to go beyond being a mere abbreviation for binary pair forms, such as capital-I forms, to also address persons identifying as non-binary. Recent psycholinguistic studies investigating the functionality of the star form showed that its use can successfully reduce the male bias evoked by masculine generics (e.g., Keith et al., 2022; Körner et al., 2022). One study moreover showed that it evokes well-balanced mental representations of *all* genders (Zacharski & Ferstl, 2023).

2) We will focus on the generic use of masculine *role nouns* and gender-fair alternatives intended to replace these. The search for gender-fair pronouns is beyond the scope of the study. The most prominent gender-fair pronouns are the singular *they* in English, e.g., Bradley (2020), and *hen* in Swedish, e.g., Renström et al. (2022). For German, there is no consensus on a gender-fair pronoun yet, see, e.g., Löhr (2022).

Grammatical gender is not the only relevant factor influencing gendered representations of person referents in German: Many role nouns are associated with semantic gender stereotypes encoding the readers' expectations of how likely it is that the referent is female or male (e.g., *female*: social workers, *male*: surgeons). This information from general world knowledge, which is related to the actual binary gender distribution within a specific group, might be altered when there is societal change (cf. Zacharski & Ferstl, 2023). Given that previous research has shown interactions between grammatical gender and stereotypical gender in German (e.g., Braun et al., 1998; Sato et al., 2016; Verweken et al., 2013), it is crucial to take gender stereotypes into account when conducting research on GFL.

Comprehensibility and Readability of Gender-Fair Forms

In order to build an appropriate mental representation in the first place, gender-fair alternatives have to be readable and comprehensible (Friedrich & Heise, 2019). While opponents of GFL have argued that use of binary pair forms and neutralizations that are in line with German orthography (in the following: *orthographic forms*) makes texts harder to read and understand (e.g., Gesellschaft für deutsche Sprache, 2020), experimental studies using self-report measures have not confirmed this claim (e.g., Blake & Klimmt, 2010; Friedrich & Heise, 2019; Pöschko & Prieler, 2018). Using eye-tracking, Steiger-Loerbroks and Von Stockhausen (2014) moreover found that gender-fair neutralizations required more processing effort than masculine forms only at early reading stages, but not at later ones.

The case is different for binary forms that are *not* in line with German orthography (in the following: *non-orthographic forms*): Pöschko and Prieler (2018) found significantly lower subjective readability ratings for texts using slash forms (e.g., *der/die Lehrer/in – the [masc.]/the [fem.] male teacher [masc.]/female teacher [fem. suffix]*). For texts using capital-I forms (e.g., *FlugbegleiterIn – stewardEss*), Blake and Klimmt (2010) reported longer total reading times than for texts using orthographic forms (masculine forms, pair forms, and neutralizations). The authors suggested that this might be due to the more complex and unusual shape as well as the rare occurrence of the former. Another open question is whether, as seems likely, frequent exposure to non-orthographic forms can reduce these processing difficulties (cf. Gabriel et al., 2018).

So far, only Friedrich et al. (2021) have investigated the readability of the star form using subjective comprehensibility ratings: In two different experiments, students were presented with texts in one of two variants (masculine/star). The text for the first experiment contained predominantly plural star forms (e.g., *die Spieler*innen – the players*), while the second text contained more singular star forms. For the first experiment, the ratings did not yield any difficulties for the star. However, in the second experiment, it significantly decreased comprehensibility, particularly due to increased sentence difficulty ratings. Interpreting these results, it should be noted that there is only one

definite plural article in German (*die*), which is used for feminine and masculine nouns. Hence, no adaptation of dependent articles is necessary when using plural star forms. In contrast, using singular star forms requires a more complex and very salient adaptation of the article (e.g., *der*die Spieler*in – the[masc.] *the[fem.] player[star]*)—the actual use of these more complex constructions is, however, very uncommon.

This study provided valuable first insights in the readability of the star, but has two limitations: First, its results are based on subjective ratings of participants. Hence, it cannot be ruled out that responses were influenced by factors such as social desirability, political correctness, or emotions towards GFL. In contrast, implicit measures enable investigation of word recognition at its earliest, pre-conscious stages³ thus avoiding potential influences of these factors. Second, its results are based on a homogenous, mostly female student sample and participants are likely to have held a positive attitude towards GFL (e.g., Jäckle, 2022; Zacharski, 2024). However, previous research showed that attitudes towards GFL might affect the processing of GFs: In their eye-tracking study, Steiger-Loerbroks and Von Stockhausen (2014) reported that a more positive attitude towards the generically intended masculine led to increased reading times for gender-fair alternatives. In their word-picture matching task, Stahlberg and Sczesny (2001) (Experiment 4) found that participants with a more positive attitude towards GFL showed faster response times for images of women following the gender-fair capital-I form than the masculine form. In contrast, no such effect was found in participants with a negative attitude towards GFL. Further research on the readability of the star is thus needed that uses implicit measures and takes potential attitudes effects into account.

Visual Word Recognition: A Fundamental Component of Text Comprehensibility

If we want to know whether the gender star makes reading a text more difficult, investigating how it affects the recognition of role nouns on the word level is a crucial step towards it. According to van Dijk and Kintsch (1983), text comprehension consists of building a mental representation of what the text is about. The comprehension processes that contribute to its production occur at the word, sentence, and text level and interact with the reader's world knowledge (Perfetti et al., 2005). Hence, "if we knew how people recognize whole words on the page, we [would] [...] know part of what we need to know in order to understand how people comprehend whole printed sentences" (Coltheart, 2006, p. 6). Reflecting this, when investigating the readability of the star, Friedrich et

3) Neuropsychological methods such as EEG and ERPs allow the investigation of the early stages of visual word recognition. Research on the exact time course of the retrieval of lexical and semantic information is, however, inconsistent: While many studies suggest that lexico-semantic processing takes place between 200–600ms (e.g., Grainger & Holcomb, 2009), some studies showed that competent readers retrieve lexical and semantic information already within 200ms after word onset (e.g., Hauk et al., 2012).

al. (2021) did not only ask participants to rate the comprehensibility of the text as a whole, but moreover assessed *word difficulty*, that is, how easily participants grasped the meanings of the text's words.

A well-established implicit measure to study word recognition is the lexical decision task (Coltheart et al., 2001; Rastle, 2016). In this paradigm, participants are presented with strings of letters. They have to decide as quickly as possible whether the stimulus shown is a word or not. Accuracy rates and reaction times (RTs) provide information on which factors influence the ease of word recognition. The most important lexical features are word length and word frequency: Shorter or more frequent words are recognized more quickly. Other factors include age of acquisition and familiarity: Words that are learned earlier in life and words that are more familiar are recognized faster (Coltheart et al., 2001; Rastle, 2016). Importantly, the paradigm even enables the detection of effects of within-word changes on word identification that are small in magnitude (e.g., case alternation: gArDeNeR, Coltheart & Freeman, 1974).

The Dual Route Cascaded model of visual word recognition (Coltheart et al., 2001) assumes that each word known by the reader is represented as an individual lexical entry in a mental lexicon. Words are recognized by mapping the identified letters on the correct entry (*lexical access*). This process allows the skilled reader to quickly and accurately read familiar words. Reading unknown words or pronounceable strings of letters (pseudowords), in contrast, requires the use of the non-lexical route (corresponding to spelling out the word) which is slower and less accurate.

For reading star nouns, there are two potential scenarios: First, owing to their non-orthographic form, they are not mapped on entries of the mental lexicon, but are processed via the non-lexical route. Second, star nouns are accessed via the lexical route, which is highly sensitive to word length, word frequency, and age of acquisition. Due to the insertion of the asterisk, star nouns are longer. Moreover, special character forms occur less frequently than feminine forms (Goldhahn et al., 2012), which are less frequent than masculine words (cf. Friedrich et al., 2021; Gabriel et al., 2018). Finally, owing to its relative recent introduction, participants will have first learnt the star form in adulthood, so that age of acquisition is higher for star nouns than for binary forms. Hence, in both scenarios, we expect slower reading times for star nouns. Interindividual differences, however, might influence the ease of recognition: First, as mentioned above, a positive attitude towards GFL might facilitate its processing. Second, owing to its recent introduction, younger participants will have become acquainted with it earlier in their life than older participants did. Moreover, many German universities recommend the use of the gender star in their guidelines for GFL (Schneider, 2022). Students might have thus encountered the gender star more frequently than non-students. Lexical access to star nouns might thus be easier for younger students than for older non-students.

The Present Study

The present study consists of two experiments employing the same lexical decision task to investigate whether the insertion of the asterisk impedes lexical access to role nouns. The first experiment was conducted with a sample of students (18–30 years). The second experiment was conducted with a more heterogeneous group of older non-students (30–80 years). In each experiment, we assessed attitudes towards GFL. This and comparisons between the samples allowed us to evaluate the influence of interindividual differences.

In the lexical decision task, 72 role nouns were used as experimental items. GF was varied *within* subjects (star/feminine/masculine) so that every participant saw 24 role nouns in each of the forms. The semantic stereotype of the role noun was controlled (Misersky et al., 2014). To conceal the purpose of the experiment, a large number of filler items was used (48 regular words, 120 pseudowords), leading to 240 trials in total. Participants had to decide as quickly as possible whether the string presented was a German word or not, and respond via keypress. Acceptance rates and RTs for experimental items were used as dependent variables.

Our study was guided by two main research questions (RQs) for which we derived specific hypotheses (Hs). The first question addresses the status of the gender star as a German word. The second one addresses the question of its readability.

Even though the star form is, by now, frequently used by media channels and institutions (Krome, 2020; Schneider, 2022), previous surveys showed that only about 25% of the population in Germany are in favor of its use (Jäckle, 2022: 21%; Welt am Sonntag, 2020: 26%). We thus wanted to know:

(RQ1) Which participants accept role nouns in star form as German words?

We expected that the polarization of the public debate (Meuleneers, 2024) would manifest itself in the distribution of acceptance rates of star nouns: While some participants accept star nouns as words, others would always reject them. Moreover, we expected that participants with a more positive attitude towards GFL and participants of the younger, student sample would be more likely to accept the star, thus leading to the following hypotheses:

(H1.a) Participants' mean acceptance rates for star nouns show a bimodal distribution.

(H1.b) Participants with a more positive attitude towards GFL are more likely to accept star nouns as words (*within* samples).

(H1.c) Participants of the younger, student sample are more likely to accept star nouns as German words than participants of the older, non-student sample (*between* samples).

Next, we were interested in whether and how the gender star affects word recognition for those participants who, in general, accept star nouns as German words:

(RQ2) Is lexical access to role nouns in star form more difficult than to feminine and masculine forms?

For this purpose, we had a closer look at RTs for accepted experimental items. Specifically, we formulated the following hypothesis:

(H2.a) The insertion of the asterisk impedes lexical access thus leading to slower RTs for star nouns compared to binary forms (controlling for word length).

Because more exposure to the star form is likely to facilitate lexical access, we expected the processing of star nouns to get easier over the course of the experiment:

(H2.b) RTs to star nouns decrease across time.

Moreover, we formulated the following hypotheses with regard to interindividual differences:

(H2.c) For participants with a more positive attitude towards GFL, recognition of star nouns is easier than for participants with a less positive attitude (*within* samples).

(H2.d) For participants of the younger, student sample, recognition of star nouns is easier than for participants of the older, non-student sample (*between* samples).

The wide range of participants' ages in the non-student sample furthermore allowed us to exploratorily test the influence of age on the ease of lexical access *within* this sample. Finally, it was an exploratory question whether gender stereotypes would influence the lexical decision process.

Experiment 1: Student Sample

In the first experiment, we tested university students to make it comparable with previous research (e.g., Friedrich et al., 2021). We focused on students younger than 30 years thus representing a group of typical students in Germany (Davies, 2023).

Method

Participants

Recruitment took place at the University of Freiburg, Germany. Only students between 18 and 29 years with sufficient knowledge of German (L1, or L2⁴ with > 10 years

of experience) were included in the dataset for analysis. Of the 124 participants who initially started the experiment, 27 were excluded (early drop-out: 5, insufficient German skills: 6, > 29 years: 6, non-students: 9). Moreover, one participant was excluded in the course of an outlier analysis (see Supplementary Materials, Zacharski, 2024S). Our final sample consisted of 97 students of psychology (74), cognitive science (21), and linguistics (2) ($M = 21.1$ years, $SD = 2.50$; non-binary: 2, female: 76, male: 18, unspecified: 1). All participants received course credit as compensation. An a priori power analysis using G*Power (Faul et al., 2007) showed that a sample of 73 participants would have been sufficient to yield a power of .80 (with $\alpha = .05$) for finding even small differences between GFs ($f = .15$).

Design and Materials

240 items were presented to each participant: 120 words and 120 pseudowords. 72 role nouns were used as experimental items. For the latter, GF was varied *within* subjects (star/feminine/masculine), so that every participant saw 24 role nouns in each form. The remaining items were fillers to distract participants from the goal of the study. Item types will be described in the following.

Words: Experimental Items – 72 role nouns that allowed feminine inflection were selected from Misersky et al. (2014), so that 24 nouns each were stereotypically female, neutral, and male. Word frequencies of masculine forms (Goldhahn et al., 2012) and word lengths of masculine forms were balanced over the three stereotypicality categories, one-way ANOVAs: word length: $F(2, 69) = 0.19$, $p = 0.83$; word frequency: $F(2, 69) = 1.67$, $p = 0.20$ (Table 1). For each role noun, three GFs were created: masculine, feminine (-in), star (-*in).

Words: Filler Items – 48 words (no role nouns; e.g., *Menschenrechte* – human rights) were used. 24 of these contained at least one special character (e.g., *H&M*) or majuscule (e.g., *iPhone*). Filler words (with and without special characters) were selected so that their word frequencies matched those of the experimental items (One-way ANOVA: $F(2, 117) = 0.35$, $p = 0.71$). Experimental items and filler items without special characters were equally long. However, words with special characters were significantly shorter, because longer ones are rare in German (One-way ANOVA: $F(2, 117) = 4.40$, $p = 0.01$) (Table 1).

4) L1 is a participant's first, L2 is a participant's second language.

Table 1
Descriptives of all Items Used as Stimuli

Stimulus type	Stimulus category	Gender stereotypicality	n (Stimuli)	Word length (masc. for exp. item)	Word frequency (masc. for exp. items)	Example
WORD	experimental item	female	24	10.71 (2.88), 7-17	15.71 (2.74), 10-23	Kosmetiker/Kosmetikerin/Kosmetiker*in (<i>beautician</i> [masc./fem./star])
WORD	experimental item	neutral	24	10.50 (3.44), 6-17	14.58 (3.05), 9-21	Biologe/Biologin/Biolog*in (<i>biologist</i> [masc./fem./star])
WORD	experimental item	male	24	10.17 (2.94), 5-17	14.50 (1.69), 12-18	Chirurg/Chirurgin/Chirurg*in (<i>surgeon</i> [masc./fem./star])
WORD	filler, no special character	—	24	10.75 (2.85), 4-16	14.42 (3.02), 7-19	Menschenrechte (<i>human rights</i>)
WORD	filler, with special character	—	24	8.50 (3.11), 3-16	14.58 (3.43), 10-22	iPhone
PSEUDOWORD	ending on <i>-er</i>	—	24	10.04 (2.93), 5-17	—	Welchzieter
PSEUDOWORD	ending on <i>-in</i>	—	24	12.29 (3.53), 7-19	—	Schrommerin
PSEUDOWORD	ending on <i>-in</i> , with * (asterisk) at random position	—	24	13.58 (2.96), 10-20	—	Flisc*rmin
PSEUDOWORD	no specific ending, no special character	—	24	10.71 (2.82), 4-16	—	Witschenbechte
PSEUDOWORD	no specific ending, with special character	—	24	8.42 (3.17), 3-16	—	Kassem-gofmalge

Note. Words: Experimental Items and Filler Items: Means, SDs, and ranges of word lengths and word frequencies by Stimulus Type. Word frequencies and word lengths of experimental are based on masculine forms. Pseudowords, all categories: Means, SDs and ranges of pseudoword lengths by Stimulus Type. Examples are given for both words and pseudowords.

Pseudowords — Based on every word (experimental and filler items), one pseudoword was generated using the multilingual pseudoword generator *Wuggy* (Version 0.1.7). The tool creates equally long pseudowords, i.e., nonwords that are in line with orthographic and phonological patterns of German (Keuleers & Brysbaert, 2010). Pseudowords based on experimental items carried the suffices that are typically associated with feminine (-in) or masculine gender (-er), or carried the feminine suffix -in and contained an asterisk at a random position. Pseudowords generated from fillers either contained no or at least one special character other than the asterisk (Table 1).

Experimental Lists — GF of experimental items was varied *within* participants, such that each participant saw 24 role nouns in each GF. The same 48 filler words and 120 pseudowords were used in each of the three lists. The entire experiment consisted of 240 trials.

To guarantee pseudo-randomization during presentation, 12 sub-lists of 20 items were created for each list. Each sub-list consisted of 10 words (2 of each type) and 10 pseudowords (2 of each type) and was presented in a randomized order. No list contained the pseudoword and the word used for its generation.

Questionnaire

A questionnaire with five items was used to assess attitudes towards GFL (Dietsche, 2020) (e.g., *In my opinion, more texts should be written in gender-fair language*; for all items see Supplementary Materials, Table A, Zacharski, 2024S). Items were rated on a Likert-scale from 1–5. Two items were reverse coded so that a higher score suggests a more positive attitude towards GFL. Before filling out the questionnaire, participants read a brief definition of the term ‘GFL’. The mean score was used for statistical analysis. With an internal consistency of Cronbach’s $\alpha = 0.81$, the scale proved to be reliable.

Presentation and Procedure

The experiment was implemented on *lab.js* (Henninger et al., 2023); the *JATOS* Server (Lange et al., 2015) was used for data collection. Participants received a link and completed the study at home on their computer. They were asked not to use smartphones/tablets, sit in a silent room, and make sure not to be disturbed. Although compliance to instructions cannot be directly measured in online-studies, their reliability has been confirmed (Schnoebelen & Kuperman, 2010).

Participants were told that the study investigates word comprehension, but were not informed about the goal of the study until after the experiment. They were instructed that they would be presented with letter strings and to decide as quickly as possible whether these are German words or not.

After giving consent and reading the instructions, all participants completed one practice block (8 trials). Then, each participant was presented with one of the experi-

mental lists. Per trial, one letter string was presented. Participants gave answers using their keyboard (D/yes, K/no or K/yes, D/no, counterbalanced across participants). Each stimulus was displayed until keypress, but not longer than 2,000ms, after which the next stimulus automatically appeared. Two 30s breaks were allowed (after trials 80 and 160). After finishing the experiment, participants filled in the attitudes-questionnaire and gave demographic information. Completing the study took about 20 minutes.

Data Analysis

All steps of the statistical analysis were conducted with *R* (Version 4.3.1) run on *R studio* (Posit Team, 2023; Version 2023.06.0).

Before the main statistical analysis, an outlier analysis was conducted to identify participants who did not perform the task appropriately, and an item check to identify words and pseudowords that elicited unpredicted responses (for details see Supplementary Materials, Zacharski, 2024S).

(RQ1): The statistical analysis for acceptance rates was based on experimental items only (6,566 datapoints). To test whether mean acceptance rates for star nouns show the expected bimodal distribution, *Hartigan's Dip Test for Unimodality* (Hartigan & Hartigan, 1985) was calculated using the *diptest*-package.

Next, a generalized mixed model (M1.1) was fitted using the *glmer*-function (*lme4*-package, Bates, Maechler, et al., 2015) with response type as the dependent variable. Yes-responses (acceptance) were coded as 1, no-responses (rejection) as 0. Based on our hypotheses, we chose a four-way interaction term for fixed effects including GF, gender stereotypicality, scaled attitude scores and scaled trial number. The parsimonious random effect structure (Bates, Kliegl, et al., 2015) including potential item effects (*1|Item*) and interindividual differences (*1|ID*) was controlled with the *rePCA*(-)-function (*lme4*-package) showing that these dimensions were sufficient to account for 100% of the explained variability. The *Anova*(-)-function (*car*-package) was used to produce type-III-Anova tables for fixed effects (for further details on contrast coding, post-hoc analyses, data visualization, and versions of *R*-packages see Supplementary Materials, Zacharski, 2024S).

(RQ2): In order to check how acceptance differed amongst participants generally accepting star nouns as words, we fitted the same *glmer*-Model (M1.2) for the sample reduced by participants who generally rejected star nouns (< 10% mean acceptance of the star; from now on: *rejectors*). For the analysis of RTs and to investigate lexical access, logarithmized yes-responses of all participants were used as the dependent variable (6,249 datapoints). We fitted a linear mixed effect model (M2) using the *lmer*-function (*lme4*-package). In addition to the fixed effect structure used for the *glmer*-Models (M1), we added scaled word lengths as an independent variable. We used the more complex, but still parsimonious random effect structure (*1|Item*) + (*1+Trial Nr+Word Length|ID*). The remaining procedures of statistical analysis were the same as for M1.

Results

Attitude Questionnaire

The mean attitude score $M = 3.79$ ($SD = 0.77$) suggested that, overall, participants held a positive attitude towards GFL (range = 1.80–5.00; scale: 1[negative] to 5[positive]). There were no significant differences in attitudes between female ($M = 3.81$, $SD = 0.74$, range 1.80–5.00) and male participants ($M = 3.72$, $SD = 0.83$, range = 2.00–4.8) (*Wilcoxon rank sum test*: $W = 649.5$, $n_1 = 76$, $n_2 = 18$, $p = 0.743$).

(RQ1) Acceptance of Role Nouns in Star Form as German Words

The statistical analysis for the complete sample (Table 2, M.1.1) yielded a significant main effect of GF. A priori contrast coding and post-hoc tests (Supplementary Materials, Tables C.1 & C.2, Zacharski, 2024S) showed that acceptance rates for star nouns are significantly lower than those for binary forms. Visualizing predicted probabilities for all participants (Figure 1, Plot A.1) showed a comparably large variance for the star. Having a closer look at the participants' mean acceptance rates for star nouns (Figure 1, Plot A.3), we found, in line with H1.a, a bimodal distribution (*Hartigan's Dip Test for Unimodality*: $D = 0.14$, $p < .001$). However, contrary to our expectations, there was only one *rejector* (100% rejection). The great majority of participants (99%) accepted star nouns as words (> 81% mean acceptance of the star).

In line with H1.b, the main effect of GF (Table 2, M.1.1) was qualified by a significant interaction between GF and attitude: Participants with a more positive attitude towards GFL were more likely to accept star nouns as words. This effect was due only to the one *rejector*, whose attitude (score = 2.40) was less positive than the sample's mean.

Moreover, the interaction between GF and trial number was significant. A priori contrast coding (Supplementary Materials, Tables C.1, Zacharski, 2024S) showed that this effect was driven by binary GFs only: While acceptance of feminine forms decreased across time, acceptance of masculine forms increased (see Supplementary Materials, Figure A, Zacharski, 2024S).

Table 2

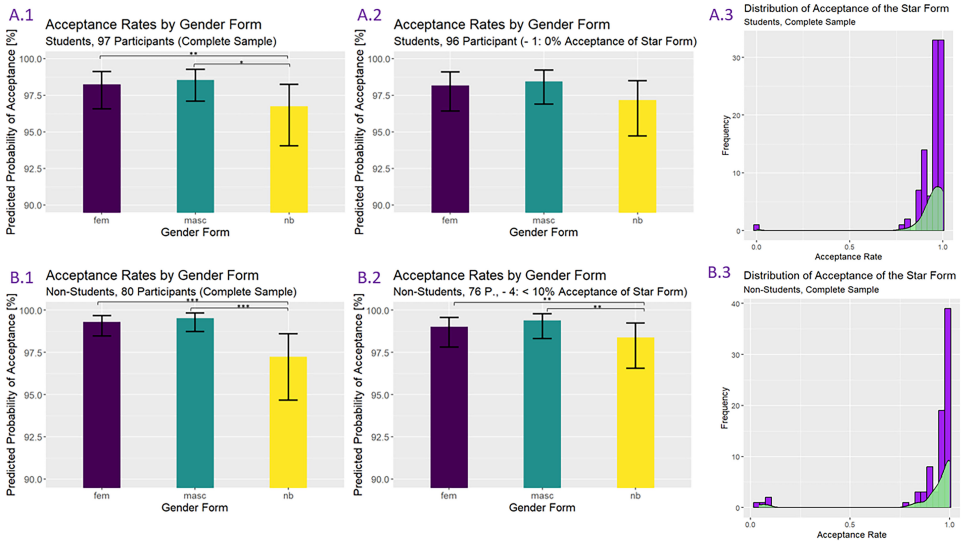
Results of Type-III Anova of *glmer*-Models of Acceptance Rates for Experimental Items for the Student Sample (M1.1, M1.2 – Experiment 1) and the Non-Student Sample (M3.1, M3.2 – Experiment 2)

Predictor	Student sample						Non-student sample					
	M1.1: Acceptance (97 participants)			M1.2: Acceptance (96 participants)			M3.1: Acceptance (80 participants)			M3.2: Acceptance (76 participants)		
	Chisq	df	p	Chisq	df	p	Chisq	df	p	Chisq	df	p
Intercept	401.49	1	<.001***	412.139	1	<.001***	340.82	1	<.001***	352.733	1	<.01***
Gender Form	13.361	2	.001**	3.391	2	.184	62.536	2	<.001***	14.784	2	.001***
Gender Stereotypicality	2.938	2	.230	3.805	2	.149	2.226	2	.329	1.501	2	.472
Trial Nr	0.794	1	.373	0.901	1	.342	3.678	1	.055	1.773	1	.183
Attitude	0.002	1	.960	0.544	1	.461	0.131	1	.718	0.019	1	.889
Gender Form:Gender Stereotypicality	5.648	4	.227	6.710	4	.152	2.277	4	.685	2.739	4	.602
Gender Form: Trial Nr	6.633	2	.036*	5.650	2	.059^	2.809	2	.245	2.045	2	.360
Gender Stereotypicality: Trial Nr	0.288	2	.866	0.737	2	.692	1.464	2	.481	1.064	2	.588
Gender Form: Attitude	11.112	2	.004**	3.902	2	.142	35.409	2	<.001***	0.153	2	.926
Gender Stereotypicality: Attitude	2.024	2	.364	1.198	2	.549	1.863	2	.394	2.910	2	.233
Trial Nr: Attitude	0.448	1	.503	0.592	1	.441	0.009	1	.926	0.182	1	.670
Gender Form: Gender Stereotypicality: Trial Nr	0.898	4	.925	0.375	4	.984	11.673	4	.020*	10.053	4	.040*
Gender Form: Gender Stereotypicality: Attitude	2.960	4	.565	7.196	4	.126	7.183	4	.127	6.183	4	.186
Gender Form: Trial Nr: Attitude	0.877	2	.645	0.504	2	.777	0.850	2	.654	0.594	2	.743
Gender Stereotypicality: Trial Nr: Attitude	2.715	2	.257	3.223	2	.200	0.521	2	.771	0.271	2	.873
Gender Form: Gender Stereotypicality: Trial Nr: Attitude	5.940	4	.204	4.972	4	.290	4.665	4	.323	4.950	4	.293

*** $p < .001$. ** $p < .01$. * $p < .05$. ^ $p < .1$.

Figure 1

Acceptance Rates by Gender Form and Distribution of Acceptance Rates of the Star Form for the Student and the Non-Student Sample



Note. Predicted Probability of Acceptance based on the Generalized Mixed Effect Models for the Student Sample (Plot A.1 & A.2) and the Non-Student Sample (Plot B.1 & B.2): Comparison of Complete (left) and Reduced Sample (right) within each of the samples. Error bars show SEs; SEs and significance levels are taken from post-hoc analysis. Histogram of the expected bimodal distribution of mean acceptance rates for the participants of the Student (Plot A.3) and the Non-Student Sample (Plot B.3). Abbreviations: fem = feminine; masc = masculine; nb = non-binary gender star.

*** $p < .001$. ** $p < .01$. * $p < .05$.

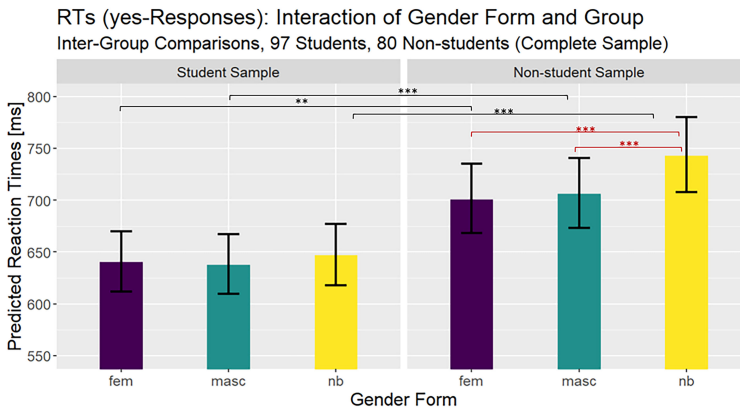
(RQ2) Lexical Access to Role Nouns in Star Form

Acceptance Rates — The statistical analysis of acceptance rates for the sample reduced by the *rejector* (Table 2, M1.2; Figure 1, Plot A.2) showed no main effect of GF.

Reaction Times — Predicted RTs of yes-responses are visualized in Figure 2 (left panel). As expected, the statistical analysis (Table 3, M2) yielded significant main effects of trial number and word length: Participants got faster across time and RTs to shorter words were faster. Importantly, and not in line with H2.a, there was no significant main effect of GF: RTs for star nouns were as fast as RTs for binary forms and do thus not suggest any difficulties in lexical access.

Figure 2

Reaction Times by Gender Form for the Student and the Non-Student Sample



Note. Predicted Reaction Times of yes-Responses of Experimental Item based on the Linear Mixed Effect Models for the Student Sample (left) and the Non-Student Sample (right). Error bars show SEs; SEs and significance levels are taken from post-hoc analysis. Significance levels for within-group models: red; significance levels for inter-group comparisons: black. fem = feminine; masc = masculine; nb = non-binary gender star.

*** $p < .001$. ** $p < .01$.

Table 3

Results of Type-III Anova of lmer-Model of Reaction Times (Yes-Responses) for Experimental Items for the Student Sample (M2 – Experiment 1) and the Non-Student Sample (M4 – Experiment 2)

Predictor	Student sample			Non-student sample		
	M2: Reaction times (yes-responses) (97 participants)			M4: Reaction times (yes-responses) (80 participants)		
	Chisq	df	p	Chisq	df	p
Intercept	89,516.429	1	< .001***	134,282.059	1	< .001***
Gender Form	3.626	2	.163	113.831	2	< .001***
Gender Stereotypicality	1.444	2	.486	0.979	2	.613
Trial Nr	18.256	1	< .001***	5.133	1	.023 *
Attitude	1.619	1	.203	0.540	1	.462
Word Length	9.968	1	.002**	12.088	1	.001***
Gender Form:Gender Stereotypicality	3.354	4	.500	9.664	4	.046*
Gender Form:Trial Nr	0.659	2	.719	6.559	2	.038*
Gender Stereotypicality:Trial Nr	2.197	2	.333	0.324	2	.851
Gender Form:Attitude	3.648	2	.161	2.430	2	.297
Gender Stereotypicality:Attitude	3.742	2	.154	1.200	2	.549

Predictor	Student sample			Non-student sample		
	M2: Reaction times (yes-responses) (97 participants)			M4: Reaction times (yes-responses) (80 participants)		
	Chisq	df	p	Chisq	df	p
Trial Nr:Attitude	2.487	1	.115	0.001	1	.976
Gender Form:Gender	5.922	4	.205	2.452	4	.653
Stereotypicality:Trial Nr						
Gender Form:Gender	11.589	4	.021*	2.246	4	.691
Stereotypicality:Attitude						
Gender Form:Trial Nr:Attitude	0.249	2	.883	0.364	2	.833
Gender Stereotypicality:Trial Nr:Attitude	0.542	2	.763	2.784	2	.249
Nr:Attitude						
Gender Form:Gender	0.432	4	.98	2.193	4	.700
Stereotypicality:Trial Nr:Attitude						

*** $p < .001$. ** $p < .01$. * $p < .05$. ^ $p < .1$.

However, we found a significant interaction between GF, gender stereotypicality, and attitude towards GFL: Participants with a more positive attitude responded faster to stereotypically female nouns in star form than in binary forms. In contrast, they showed slower RTs to stereotypically male nouns in star form than in binary forms (see Supplementary Materials, Figure B, Zacharski, 2024S).

Summary

In line with hypothesis H1.a, we found the expected bimodal distribution of acceptance rates. Notably, however, only one of 97 students generally rejected star nouns as German words. In line with H1.b, we found that the general acceptance of the gender star might be influenced by attitudes towards GFL—however, as there was only one *rejector*, this finding is not reliable. For participants who generally accepted star nouns as words, there was no significant effect of GF on RTs. Hypothesis H2.a could thus not be confirmed. Consequently, H2.b and H2.c were not relevant.

Experiment 1 is subject to two limitations: First, we tested a homogenous sample of young students, who held a rather positive attitude towards GFL. Second, the majority of participants identified as female. Thus, Experiment 2 was conducted with a sample of older non-students, that was balanced with regard to binary gender identities.

Experiment 2: Non-Student Sample

In Experiment 2, we replicated the study described with a sample of non-students varying in age (30–80 years) and academic background. We stuck as closely to the initial design as possible, but used a more detailed questionnaire to assess attitudes towards GFL. All aspects in which Experiment 2 differed from the first one will be described in the following.

Method

Participants

84 non-student participants over the age of 30 were recruited via *Prolific*. Two participants currently enrolled at a university and one participant with insufficient knowledge of German were excluded. One participant was excluded in the course of an outlier analysis (see Supplementary Materials, [Zacharski, 2024S](#)). Our final sample consisted of 80 non-students (non-binary: 1, female: 32, male: 47) between 30–80 years (age distribution: Supplementary Materials, Table B, [Zacharski, 2024S](#)). 45 participants had an academic background, while 35 had not studied at a university. All participants received financial compensation in line with the recommendations given by *Prolific*.

Design and Materials

To avoid technical difficulties that occurred during the first experiment, two filler items were added after the start of the experiment and each of the two breaks.

Questionnaire

The 32-item *ABNBL*-questionnaire ([Zacharski, 2024](#)) designed to assess attitudes towards non-binary GFL in German was used. All items were rated on a Likert-scale from 0–9 and a higher score indicated a more positive attitude towards GFL. Following [Zacharski \(2024\)](#), participants read a brief definition of the term ‘non-binary gender identity’ before the questionnaire. The mean item score was used for statistical analysis.

Presentation and Procedure

The experiment was implemented on PCIBex ([Zehr & Schwarz, 2022](#)). Participants were redirected to the study via *Prolific*. They replied via keypress (Y/yes, N/no⁵). The procedure for the experiment was the same as in the first experiment, but participants went through two practice blocks (6 trials each)—one with, and one without feedback.

5) We changed the response keys to a more intuitive combination. The distance between the keys on the German keyboard is the same as in the first experiment.

Data Analysis

(RQ1): The acceptance rates (M3.1) for experimental items (5,664 datapoints) were analyzed as in M1.1 (Experiment 1).

(RQ2): The model fit for RTs (M4) was based on yes-responses to experimental items of all participants (5,423 datapoints), as in M2 (Experiment 1).

We fit an additional exploratory model to test the influence of age on lexical access *within* this sample.⁶ The model specification (*M4_age*, Supplementary Materials, Tables K.1 & K.2, Zacharski, 2024S) was the same as for M4, except that *Age* was added to the fixed effects term.

Finally, for intergroup comparisons and to test H2.d, the processed datasets for both age groups were combined. The two models fitted (acceptance rates for reduced samples: M5; RTs of yes-responses for full samples: M6) were the same as for within-group analyses, except that, because different attitude-scales were used in the experiments, the attitude-variable was replaced by the factor *Group* (*Students* vs. *Non-Students*) (see Supplementary Materials, Tables F-K, Zacharski, 2024S).

Results

Attitude Questionnaire

The mean attitude score was $M = 4.26$ ($SD = 1.73$, range = 0.94–8.16; scale: 0 [negative] to 9 [positive]), with female participants having significantly more positive attitudes ($M = 4.97$, $SD = 1.63$, range = 2.38–8.16) than male participants ($M = 3.75$, $SD = 1.62$, range = 0.94–7.41) (*Wilcoxon rank sum test*: $W = 1056.50$, $n_1 = 32$, $n_2 = 47$, $p = 0.002$).

(RQ1) Acceptance of Role Nouns in Star Form as German Words

The statistical analysis of acceptance rates for the complete sample (Table 2, M3.1) yielded a significant main effect of GF. A priori contrast coding and post hoc analysis (Supplementary Materials, Tables F.1 & F.2, Zacharski, 2024S) showed that acceptance of star nouns was significantly lower than for binary forms. For the complete sample (Figure 1, Plot B.1), there was a comparably large variance for the star form. The participants' mean acceptance rates for star nouns (Figure 1, Plot B.3) showed, in line with H1.a, the expected bimodal distribution (*Hartigan's Dip Test for Unimodality*: $D = 0.11$, $p < .001$). However, in line with H1.c, compared to just one *rejector* in the student sample, more people, four in total, rejected star nouns in the majority of cases (less than 10% accepted). Still, as the major mode of the distribution shows, the great majority of participants (95%) accepted star nouns as German words (> 79% mean acceptance of star).

As in the student sample, and in line with H1.b, the main effect of GF in the complete sample was qualified by a significant interaction between GF and attitude towards GFL:

6) We thank an anonymous reviewer for this suggestion.

All four *rejectors* held a rather negative attitude towards GFL ($M = 2.0$, $SD = 0.72$, range = 1.47–3.06). Interestingly, they all identified as male.

Moreover, we found a significant three-way interaction between GF, gender stereotypicality, and trial number. A priori contrast-coding (Supplementary Materials, Table F.1, Zacharski, 2024S) showed that this effect was driven by binary GFs: For stereotypically female role nouns, acceptance of feminine forms increased while acceptance of masculine forms decreased over time. Thus, over the course of the experiment, a consistency effect between stereotypes and GF for stereotypically female nouns emerged. For stereotypically neutral nouns, acceptance of feminine forms decreased while acceptance of masculine forms increased (see Supplementary Materials, Figure C, Zacharski, 2024S). However, overall, the acceptance of binary forms was very high and the differences that emerged were rather small.

(RQ2) Lexical Access to Role Nouns in Star Form

Acceptance Rates — In contrast to the student sample, the statistical analysis of acceptance rates for the sample reduced by *rejectors* (Table 2, M3.2; Figure 1, Plot B.2) yielded a main effect of GF. A priori contrast coding and post hoc analysis (Supplementary Materials, Tables G.1 & G.2, Zacharski, 2024S) showed that this effect was due to higher rejection rates for star nouns compared to binary forms. This effect was, however, not qualified by an interaction with attitudes towards GFL as in model M3.1. Only the interaction effect of GF, gender stereotypicality, and trial number driven by binary forms described above remained significant.

Reaction Times — Predicted RTs of yes-responses for the different GFs are visualized in Figure 2 (right panel). Analogous to the student sample, the statistical analysis (Table 3, M4) yielded a main effect of trial number and of word length. In contrast to the student sample, however, it moreover yielded a significant main effect of GF: A priori contrasts and post-hoc analysis (Supplementary Materials, Tables H.1 & H.2, Zacharski, 2024S; Figure 2, right panel, significance levels in red) showed that RTs for star nouns were significantly slower than for binary forms. This finding is in line with H2.a and suggests difficulties in lexical access for star nouns.

In line with H2.b, this effect was qualified by a significant interaction of GF and trial number: RTs to star nouns decreased faster than those for binary forms. That is, there was an adaptation effect for star nouns over the course of the experiment—up to the point where initial differences between the three GFs disappeared (Figure 3, Plot A).

Contrary to H.2c, there was no significant interaction between GF and attitudes towards GFL. Hence, for the subgroup of people who generally accepted the gender star, the attitude score was unrelated to the performance in the lexical decision task.

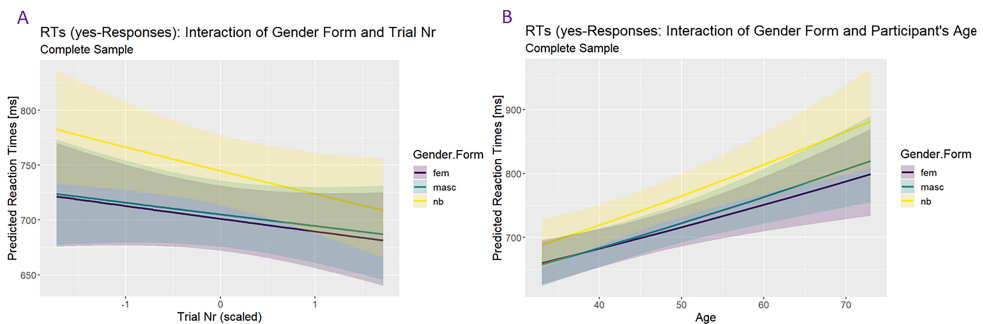
A priori contrast coding (Supplementary Materials, Tables H.1, Zacharski, 2024S) showed that the interaction of GF and gender stereotypicality was due to binary GFs

only, suggesting, again, consistency effects: For stereotypically female words, RTs for feminine forms were faster than for masculine forms, and vice versa for stereotypically male words (Supplementary Materials, Figure D, Zacharski, 2024S).

Finally, note that in the exploratory model *M4_age* (Supplementary Materials, Table K.1, Zacharski, 2024S), the main effect of GF was qualified by an interaction with age. A priori contrast coding (Supplementary Materials, Table K.2, Zacharski, 2024S) showed that, while RTs generally increased with age, this increase was stronger for the star compared to binary forms (Figure 3, Plot B).

Figure 3

Reaction Times: Interaction Plots for Gender Form and A: Trial Nr; B: Participant's Age for the Non-Student Sample



Note. Plot A: Predicted Reaction Times (yes-Responses) by Gender Form and Trial Number, Model M4.1 (Table 3) for the Non-Student Sample. Plot B: Predicted Reaction Times (yes-Responses) by Gender Form and Age, Model M4.2 (Table 3) for the Non-Student Sample. fem = feminine; masc = masculine; nb = non-binary gender star.

Inter-Group Comparisons – The description of the results for the two samples confirmed the expectation that students have less difficulties processing the rather recently introduced gender star. Although the experiments were conducted separately, a direct statistical comparison was carried out to uncover group differences (Table 4, M5 & M6): In line with the observations from M1.1 and M.3.1, for acceptance rates of the reduced samples without *rejectors* (Table 4, M5), there was a main effect of *Group* qualified by a significant interaction with gender form. A priori contrast coding and post-hoc analysis (Supplementary Materials, Table I.1 & I.2, Zacharski, 2024S) showed that, while for binary GFs, non-students had significantly higher acceptance rates than students (*feminine*: non-students: $M = 97.73$, $SD = 3.36$, students: $M = 95.96\%$, $SD = 4.15$, *masculine*: non-students: $M = 97.66$, $SD = 3.10$, students: $M = 95.48\%$, $SD = 5.90$), no significant difference was found for star forms (non-students: $M = 96.22$, $SD = 5.03$, students: $M = 95.07\%$, $SD = 4.92$). The statistical analysis for RTs (Table 4, M6) yielded a significant main effect of *Group*: Independent of GF, students responded faster than non-students (Figure 2, significance levels in black). A priori contrast coding for the significant interaction

of GF and *Group* (Supplementary Materials, Table J, Zacharski, 2024S) confirmed the observation that, while there was no significant effect of GF on RTs in the student sample, GF had a significant effect on lexical access in the non-student sample.

Table 4

Inter-Group Comparison: Results of Type-III Anova of glmer-Models of Acceptance Rates (M5) and of lmer-Model of Reaction Times (Yes-Responses) (M6) for Experimental Items

Predictor	Inter-Group comparisons					
	M5: Acceptance rates (students and non-students, reduced sample, 172 participants)			M6: Reaction times (yes- responses) (students and non-students, complete sample, 177 participants)		
	Chisq	df	p	Chisq	df	p
Gender Form	18.268	2	< .001***	57.167	2	< .001***
Gender Stereotypicality	2.680	2	.262	1.292	2	.524
Trial Nr	5.595	1	.018*	56.679	1	< .001***
Group	17.449	1	< .001***	22.983	1	< .001***
Word length	—	—	—	13.520	1	< .001***
Gender Form:Gender Stereotypicality	7.251	4	.123	8.388	4	.078 [^]
Gender Form:Trial Nr	6.332	2	.042*	4.937	2	.085 [^]
Gender Stereotypicality:Trial Nr	1.504	2	.472	1.199	2	.549
Gender Form:Group	5.390	2	.068 [^]	55.057	2	< .001***
Gender Stereotypicality:Group	1.616	2	.446	2.209	2	.331
Trial Nr:Group	1.926	1	.165	2.955	1	.086 [^]
Gender Form:Gender Stereotypicality:Trial Nr	7.767	4	.100	4.517	4	.340
Gender Form:Gender Stereotypicality:Group	2.480	4	.648	4.802	4	.308
Gender Form:Trial Nr:Group	1.182	2	.554	1.355	2	.508
Gender Stereotypicality:Trial Nr:Group	2.302	2	.316	2.421	2	.298
Gender Form:Gender Stereotypicality:Trial Nr:Group	9.405	4	.052 [^]	3.657	4	.454

*** $p < .001$. ** $p < .01$. * $p < .05$. [^] $p < .1$.

Summary

With regard to (RQ1) and in line with hypothesis H1.a, we found the expected bimodal distribution of acceptance rates: Four of the 80 non-students generally rejected star nouns as German words. In line with H1.b, they held a rather negative attitude towards GFL. In line with H1.c, the percentage of *rejectors* was higher in the older, non-student sample than in the younger, student sample.

With regard to (RQ2) and in line with H2.a, participants of the non-student sample who generally accepted star nouns as words, that is, the majority of participants (95%), showed significantly slower RTs for star nouns than for feminine and masculine forms. However, in line with H2.b, RTs for star nouns decreased more strongly over the course

of the experiment than for binary forms suggesting that the processing of the star became less effortful over time. Contrary to H2.c, attitudes towards GFL did not predict the performance in the lexical decision task. However, in line with H2.d, comparisons between the two samples showed that processing was easier for the younger, student sample than for the older, non-student sample. In line with that, *within* the non-student sample, a higher age was accompanied by slower RTs to star nouns compared to binary forms. This suggests that, in addition to the (non-)student status, one factor driving the processing differences *between* the student and the non-student sample might be age.

General Discussion

The non-binary gender star in German (e.g., Radfahrer*in – *cyclist*) is the most popular gender-fair alternative for generically intended masculine role nouns to refer to persons of *all* genders, that is, persons identifying beyond a female-male dichotomy, as well as women and men (Krome, 2020). Opponents of its use argue that its non-orthographic form impedes the readability of texts (e.g., *Rat für deutsche Rechtschreibung*, 2021). Experimental research on this claim is rare. Because visual word recognition is a crucial component of the reading process (Coltheart, 2006), we developed a lexical decision task to test whether lexical access to singular role nouns is more difficult to star nouns compared to the orthographic and more common feminine and masculine forms. To the best of our knowledge, this is the first study investigating the readability of the star with implicit measures and, therefore, an important contribution to the current debate on GFL in Germany. In order to account for interindividual differences, we tested not only a homogenous student sample (Experiment 1), but also a heterogenous sample of non-students varying in age as well as academic background, which is a further strength of our study (cf. Jones, 2010).

Previous surveys on the acceptance of the gender star in Germany showed that only one quarter of the German population is in favor of its use (e.g., Jäckle, 2022). Based on these findings, we expected a bimodal distribution of the acceptance of star nouns as German words. While this hypothesis was confirmed, the number of *rejectors* was surprisingly small: only one student and four non-students, that is, merely 2.8% of all participants rejected star nouns as words in almost all of the cases. As expected, participants of the younger, student sample were more likely to generally accept star nouns as German words. Moreover, general rejection of star nouns as words was predicted by attitudes towards GFL: Participants with a less positive attitude towards GFL were more likely to be amongst the *rejectors*.

A closer inspection of the influence of GF on RTs for those participants who generally accept star nouns as words, sheds light on whether lexical access to star nouns is more difficult than to the more common binary forms. For the student sample, we found no evidence for difficulties in the recognition of role nouns in star form. These findings are

in line with previous findings by Friedrich et al. (2021), who reported that the use of the gender star did not increase subjectively perceived word difficulty in a student sample. The case was different for the non-student sample: Participants showed significantly slower RTs to star nouns than to binary forms. These findings suggest difficulties in lexical access due to the insertion of the non-orthographic asterisk. However, processing of the star became easier over the course of the experiment: While RTs generally decreased across time, that is, independent of GF, the decrease was significantly steeper for star nouns—up to the point where RTs were comparably fast for all GFs. This suggests an adaptation effect for the non-student sample. Thus, while the initial processing of star nouns seems to be harder, their processing becomes less effortful with time (cf. Gabriel et al., 2018)—even over the short course of the experiment. Interestingly, and not in line with our expectations, attitudes towards GFL did not predict the success of lexical access to star nouns. However, participants' age might be a predictor: First, participants of the younger, student sample had significantly less difficulties than participants of the older, non-student sample. Second, an additional exploratory analysis *within* the non-student sample showed that RTs increased with age, and this increase was largest for star nouns. A possible explanation for the crucial role participants' age seems to play for successful lexical access—in line with theories on visual word recognition (Coltheart et al., 2001; Rastle, 2016)—is the fact that the gender star has been in the public for less than a decade. Consequently, older people first encountered star nouns as adults and age of acquisition is higher. Moreover, in its guidelines for GFL, the University of Freiburg explicitly suggests replacing the generically intended masculine with gender-fair alternatives such as the gender star (cf. Schneider, 2022). Thus, students are likely to be more familiar with the star than non-students: Even though the use of GFL in the media and other public domains also increases, the usage of GFL is, in general, less common and less systematic in informal contexts (cf. Gabriel et al., 2018). A gender-fair form useful to further investigate the role of age of acquisition and familiarity is the capital-I form, as this non-orthographic abbreviation of the pair form has already been in use since the 1980s. Comparing word recognition of the capital-I and the gender star form in older readers might allow us to differentiate more thoroughly between the influences of orthography, familiarity, and age of acquisition.

Another interesting finding is that gender stereotypes are activated already on the word level. In line with previous studies (e.g., Sato et al., 2016; Zacharski & Ferstl, 2023), we found consistency effects for the binary forms in the non-student sample: RTs were faster when the semantic stereotype matched the grammatical gender of the role noun. Thus, semantics are activated in the processing of different GFs—even if the meaning of role nouns is irrelevant for the participants' task. A more subtle semantic influence was found for students with a positive attitude towards GFL. Here, RTs to star nouns decreased more slowly over the duration of the experiment when the role noun was stereotypically male, as compared to stereotypically female and neutral nouns.

Further research is needed to investigate potential interactions between attitudes, gender stereotypicality, and non-binary forms such as the gender star.

Even though one strength of our study is that a lexical decision task allows processes on the word level to be tested early on and without potential influences of text context, word recognition is only one component of the reading process. Future studies should investigate the processing of the gender star on the text level using implicit measures such as eye-tracking. It will be moreover interesting to consider further interindividual differences. In particular, in order to find out more about the influence of participants' gender identity on the processing of GFL, we have to take *all* genders into account—particularly those identifying beyond the binary. Furthermore, future studies should test how different forms of GFL affect persons with reading disorders, or L2-learners.

Conclusions

The experimental design was well-suited to test the influence of the within-word insertion of a special character in non-binary gender-fair forms on lexical access to role nouns, thus providing a highly valuable addition to studies based on self-report questionnaires. RTs showed differences in lexical access to the gender star compared to binary forms for the older, non-student sample, but not for the younger, student sample. While age—in addition to the (non-)student status of participants—predicted the ease of lexical access to star nouns, subjective attitude ratings did not. The worry that the gender star is more difficult to read is thus not completely unwarranted. However, our findings showed that initial difficulties can be overcome the more a person is confronted with gender-fair alternatives. In our opinion, the results presented are thus promising for proponents of the star: When the gender star becomes more established, traffic signs displaying “Radfahrer*innen bitte absteigen” (*Cyclists [star] please dismount*) might decrease the male bias—while still guaranteeing readability.

Funding: This research was funded by the German Research Foundation (DFG) (project number: 456835372/Ferstl, FE 474/5-1).

Acknowledgments: The first experiment was conducted by Alexandra Kruppa in partial fulfilment of the requirements for the M.Sc. degree in psychology. We are indebted to Julius Fenn for supporting the programming and data processing of the first experiment. Moreover, we are thankful for the inspirational exchange with Helga Kotthoff, Damaris Nübling, Hannah Bröder, and Paul Meuleneers within the DFG-project *Gender related practices in person reference: Discourse, grammar, cognition* and with Lars Konieczny. Finally, we wish to thank our student assistants Sarah Kapp and Tim Sudermann.

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: *Lisa Zacharski*—Methodology | Theoretical background | Formal analysis | Writing – original draft | Writing – review & editing. *Alexandra Kruppa*—Conceptualization | Methodology. *Evelyn C. Ferstl*—Conceptualization | Methodology | Theoretical background | Supervision | Funding acquisition | Writing – review & editing.

Ethics Statement: The study was conducted in line with the ethical standards required by the Declaration of Helsinki and did not involve vulnerable participants or pose any risks. Informed consent has been obtained from all respondents prior to their participation in the study.

Data Availability: For this article, data, a codebook, and the code used for data analysis are available (see *Zacharski, 2024S*).

Supplementary Materials

Stimulus materials, raw data, a codebook, and the code used for data analysis are available on figshare (see *Zacharski, 2024S*).

Index of Supplementary Materials

Zacharski, L. (2024S). *Supplementary materials to "The readability of the non-binary gender star in German: Evidence from a lexical decision task"* [Materials, data, codebook, code]. figshare. <https://doi.org/10.6084/m9.figshare.24995900>

References

- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv. <https://doi.org/10.48550/arXiv.1506.04967>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Blake, C., & Klimmt, C. (2010). Geschlechtergerechte Formulierungen in Nachrichtentexten. *Publizistik*, 55(3), 289–304. <https://doi.org/10.1007/s11616-010-0093-2>
- Bradley, E. D. (2020). The influence of linguistic and social attitudes on grammaticality judgments of singular ‘they’. *Language Sciences*, 78, Article 101272. <https://doi.org/10.1016/j.langsci.2020.101272>
- Braun, F., Gottburgsen, A., Sczesny, S., & Stahlberg, D. (1998). Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift Für Germanistische Linguistik*, 26(3), 265–283. <https://doi.org/10.1515/zfagl.1998.26.3.265>
- Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, 4(1), 5–17. <https://doi.org/10.1080/13603110600574322>
- Coltheart, M., & Freeman, R. (1974). Case alternation impairs word identification. *Bulletin of the Psychonomic Society*, 3, 102–104. <https://doi.org/10.3758/BF03333407>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. <https://doi.org/10.1037//0033-295X.108.1.204>
- Davies, K. (2023). *Average age of first degree university graduates in Germany from 2003 to 2022 (in years)*. Statista. <https://www.statista.com/statistics/584325/first-degree-university-graduates-age-germany/>
- Dietsche, L. (2020). *Disambiguierung des generischen Maskulinums: Eine Eye-Tracking-Studie zum Einfluss des grammatischen Geschlechts auf die Interpretation von Personenbezeichnungen mithilfe des Visual-World-Paradigmas* [Master’s thesis]. Albert-Ludwigs-Universität Freiburg, Freiburg.
- Diewald, G., & Steinhauer, A. (2017). *Richtig gendern. Wie Sie angemessen und verständlich schreiben*. Dudenverlag.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Friedrich, M. C. G., Drößler, V., Oberlehberg, N., & Heise, E. (2021). The Influence of the Gender Asterisk (“Gendersternchen”) on Comprehensibility and Interest. *Frontiers in Psychology*, 12, Article 760062. <https://doi.org/10.3389/fpsyg.2021.760062>
- Friedrich, M. C. G., & Heise, E. (2019). Does the Use of Gender-Fair Language Influence the Comprehensibility of Texts? *Swiss Journal of Psychology*, 78(1-2), 51–60. <https://doi.org/10.1024/1421-0185/a000223>
- Gabriel, U., Gyax, P. M., & Kuhn, E. A. (2018). Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations*, 21(5), 844–858. <https://doi.org/10.1177/1368430218771742>
- Gesellschaft für deutsche Sprache. (2020). *Leitlinien der GfdS zu den Möglichkeiten des Genderings*. <https://gfdS.de/standpunkt-der-gfdS-zu-einer-geschlechtergerechten-sprache/>
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig corpora collection: From 100 to 200 languages. *Proceedings of the 8th International*

Language Resources and Evaluation (LREC'12).

https://corpora.uni-leipzig.de/de?corpusId=deu_news_2020

- Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, 3(1), 128–156. <https://doi.org/10.1111/j.1749-818X.2008.00121.x>
- Gygax, P., Elmiger, D., Zufferey, S., Garnham, A., Sczesny, S., Von Stockhausen, L., Braun, F., & Oakhill, J. (2019). A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in Psychology*, 10, Article 1604. <https://doi.org/10.3389/fpsyg.2019.01604>
- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, 23(3), 464–485. <https://doi.org/10.1080/01690960701702035>
- Gygax, P., Sato, S., Öttl, A., & Gabriel, U. (2021). The masculine form in grammatically gendered languages and its multiple interpretations: A challenge for our cognitive system. *Language Sciences*, 83, Article 101328. <https://doi.org/10.1016/j.langsci.2020.101328>
- Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *Annals of Statistics*, 13(1). Advance online publication. <https://doi.org/10.1214/aos/1176346577>
- Hauk, O., Coutout, C., Holden, A., & Chen, Y. (2012). The time-course of single-word reading: Evidence from fast behavioral and brain responses. *NeuroImage*, 60(2), 1462–1477. <https://doi.org/10.1016/j.neuroimage.2012.01.061>
- Henninger, F., Shevchenko, Y., Mertens, U., Kieslich, P. J., & Hilbig, B. E. (2023). *lab.js: A free, open, online experiment builder* [Computer software]. Zenodo.
- Jäckle, S. (2022). Per aspera ad astra – Eine politikwissenschaftliche Analyse der Akzeptanz des Gendersterns in der deutschen Bevölkerung auf Basis einer Online-Umfrage. *Politische Vierteljahresschrift*, 63(3), 469–497. <https://doi.org/10.1007/s11615-022-00380-z>
- Jones, D. (2010). A WEIRD view of human nature skews psychologists' studies. *Science*, 328, 1627. <https://doi.org/10.1126/science.328.5986.1627>
- Keith, N., Hartwig, K., & Richter, T. (2022). Ladies first or ladies last: Do masculine generics evoke a reduced and later retrieval of female exemplars? *Collabra. Psychology*, 8(1), Article 32964. <https://doi.org/10.1525/collabra.32964>
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633. <https://doi.org/10.3758/BRM.42.3.627>
- 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. <https://doi.org/10.1177/0261927X221080181>
- Krome, S. (2020). Zwischen gesellschaftlichem Diskurs und Rechtschreibnormierung: Geschlechtergerechte Schreibung als Herausforderung für gelungene Textrealisation. *Der Sprachdienst*, 64(1-2), 31–45.

- Lange, K., Kühn, S., & Filevich, E. (2015). Correction: "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLoS One*, *10*(7), Article e0134073. <https://doi.org/10.1371/journal.pone.0134073>
- Löhr, R. (2022). „Ich denke, es ist sehr wichtig, dass sich so viele Menschen wie möglich repräsentiert fühlen“: Gendergerechte Sprache aus der Sicht nicht-binärer Personen. In G. Diewald & D. Nübling (Eds.), *Linguistik - Impulse et Tendenzen: Band 95. Genus - Sexus - Gender* (pp. 349–379). De Gruyter. <https://doi.org/10.1515/9783110746396-012>
- Meuleneers, P. (2024). On the 'invention' of the Gendersprache in German media discourse. In F. Pfalzgraf (Ed.), *Language and social life: Vol. 31. Public attitudes towards gender-inclusive language. A multilingual perspective* (pp. 159–182). De Gruyter Mouton.
- Misersky, J., Gygas, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., Chiarini, T., Englund, K., Hanulíková, A., Ottl, A., Valdrova, J., Von Stockhausen, L., & 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. <https://doi.org/10.3758/s13428-013-0409-z>
- Perfetti, C., Landi, N., & Oakhill, J. (2005). Chapter 13. The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell Publishing. <https://doi.org/10.1002/9780470757642.ch13>
- Pöschko, H., & Prieler, V. (2018). Zur Verständlichkeit und Lesbarkeit von geschlechtergerecht formulierten Schulbuchtexten. *Zeitschrift Für Bildungsforschung*, *8*(1), 5–18. <https://doi.org/10.1007/s35834-017-0195-2>
- Posit Team. (2023). *RStudio: Integrated development environment for R* [Computer software]. PBC. <http://www.posit.co/>
- Pusch, L. F. (1984). *Das Deutsche als Männersprache*. Suhrkamp Verlag.
- Rastle, K. (2016). Visual word recognition. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 255–264). Academic Press/Elsevier.
- Rat für deutsche Rechtschreibung. (2021). *Geschlechtergerechte Schreibung: Empfehlungen vom 26.03.2021*. https://www.rechtschreibrat.com/DOX/rfdr_PM_2021-03-26_Geschlechtergerechte_Schreibung.pdf
- Renström, E. A., Lindqvist, A., & Sendén, M. G. (2022). The multiple meanings of the gender-inclusive pronoun hen: Predicting attitudes and use. *European Journal of Social Psychology*, *52*(1), 71–90. <https://doi.org/10.1002/ejsp.2816>
- Sato, S., Gygas, P. M., & Gabriel, U. (2016). Gauging the impact of gender grammaticization in different languages: application of a linguistic-visual paradigm. *Frontiers in Psychology*, *7*, Article 140. <https://doi.org/10.3389/fpsyg.2016.00140>
- Schneider, J. G. (2022). Gendern in institutionellen Leitfäden. Im Spannungsfeld von Indexikalität und grammatischen Erfordernissen. In M. Hennig & R. Niemann (Eds.), *Ratgeber in der spätmodernen Gesellschaft. Ansätze einer linguistischen Ratgeberforschung* (pp. 233–261). Stauffenburg.

- Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, 43(4), 441–464. <https://doi.org/10.2298/PSI1004441S>
- Stahlberg, D., & Sczesny, S. (2001). Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen. *Psychologische Rundschau*, 52(3), 131–140. <https://doi.org/10.1026//0033-3042.52.3.131>
- Steiger-Loerbroks, V., & Von Stockhausen, L. (2014). Mental representations of gender-fair nouns in German legal language: An eye-movement and questionnaire-based study. *Linguistische Berichte*, 237, 57–80. https://doi.org/10.46771/2366077500237_4
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- Vervecken, D., Hannover, B., & Wolter, I. (2013). Changing (S)expectations: How gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *Journal of Vocational Behavior*, 82(3), 208–220. <https://doi.org/10.1016/j.jvb.2013.01.008>
- Welt am Sonntag. (2020). *Bundesweite infratest dimap Umfrage: Vorbehalte gegenüber genderneutraler Sprache*. <https://www.infratest-dimap.de/umfragen-analysen/bundesweit/umfragen/aktuell/vorbehalte-gegenueber-genderneutraler-sprache/>
- Zacharski, L. (2024). Using pair forms, criticizing the gender star—Attitudes towards binary and non-binary gender-inclusive language in German. In F. Pfalzgraf (Ed.), *Language and social life: Vol. 31. Public attitudes towards gender-inclusive language. A multilingual perspective* (pp. 209–242). De Gruyter Mouton.
- 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(4-5), 294–319. <https://doi.org/10.1080/0163853X.2023.2199531>
- Zehr, J., & Schwarz, F. (2022). *PennController for internet based experiments (IBEX)*. Advance online publication. <https://doi.org/10.17605/OSF.IO/MD832>

PSPS
Polish
Social
Psychological
Society

Social Psychological Bulletin (SPB)
is an official journal of the Polish
Social Psychological Society
(PSPS).



leibniz-psychology.org

PsychOpen GOLD is a publishing
service by Leibniz Institute for
Psychology (ZPID), Germany.