The affective and linguistic roots of gender stereotypes

Does grammatical gender influence the affective meanings attached to social categories in French and German?

Diego Dametto

Potsdam University of Applied Sciences

Luc Vieira

Université Paris Cité

Tania Noël

University of Liège

Tobias Schröder

Potsdam University of Applied Sciences

Cristina Aelenei

Université Paris Cité

Christophe Blaison

Université Paris Cité

Corresponding Author:

Diego Dametto

Institute for Urban Futures

Potsdam University of Applied Sciences

Kiepenheuerallee 5

14469 Potsdam, Germany

Email: diego.dametto@fh-potsdam.de

Funding

This work was supported by the ANR (Agence Nationale de la Recherche, grant # ANR-20-FRAL-0008) granted to Christophe Blaison, the DFG (Deutsche Forschungsgemeinschaft, grant # SCHR1282/6-1) granted to Tobias Schröder, as part of the collaborative French-German funding program for the Humanities and Social Sciences (FRAL). The study was also supported by the IdEx Université Paris Cité, ANR-18-IDEX-0001, granted to Christophe Blaison.

Data Availability Statement

The data that support the findings of this study are openly available in PsychArchives at https://doi.org/10.23668/psycharchives.14237.

Informed Consent

In the data collection of the study, participants provided their informed consent at the beginning of the online questionnaire.

Ethics Statement

The study reported in this manuscript involves human participants and was approved by the ethical board of Université Paris Cité (N° 2021-75).

Conflict of Interest Statement

All authors declare that they have no conflicts of interest.

Abstract

This paper compares the affective meanings attributed to feminine and masculine versions of the same social categories. We collected semantic differential ratings for 192 pairs of masculine and feminine social categories from two representative samples of n = 700 French and n = 700 German participants. We investigated how lexical and grammatical gender influences the affective meaning of social categories on Osgood's dimensions of evaluation, potency, and activity, as well as how differences in affective meaning are related to known gender stereotypes. Our results suggest that gender does produce only small effects. However, they are not randomly distributed and consistent with well-known gender stereotypes. The results enrich social psychological research on gender as well as the cultural debate on the use of gender-fair language in grammatically gendered languages by providing a more nuanced assessment of the affective roots of gender stereotypes conveyed by language.

Keywords

gendered language, affective meanings, EPA dimensions, generic masculine, gender fair language

Introduction

Symbolic interactionism (Mead, 1934; Goffman, 1959) sees language as key to socialization and the maintenance of social order (Heise et al., 2015). Beyond practical knowledge conveyed through denotative meanings (Berger & Luckmann, 1966), language carries affective meanings shared among members of the same culture (Heise, 2010; Ambrasat et al., 2014; Dametto et al., 2023) and captured by the cross-cultural dimensions Evaluation-Potency-Activity (EPA; Osgood et al., 1957; Osgood et al., 1975). In fact, culture can be seen as the totality of socially transmitted and intersubjectively shared affective meanings that constrain and guide social interaction (Heise, 2007). Through everyday encounters and interactions, this affective knowledge is continuously enacted and, according to Affect Control Theory (ACT), underlies the impression formation process by providing the backbone of situational appraisal (Heise, 2007). In this way, the affective meanings encapsulated in the language are inextricably intertwined with stereotypes (Schröder et al., 2013). The overall research goal of this paper is to understand how the use of the feminine versus masculine version of various social categories relates to affective meanings attributed to these social categories[1].

Such a goal is of particular importance in the context of grammatically gendered languages (Sczesny et al., 2016). Contrary to English, in gendered languages like French and German, gender is salient, promoting and reinforcing gender associations and stereotypes encapsulated in semantics (Gygax et al., 2019; Garg et al., 2018; Lewis & Lupyan, 2020; Charlesworth et al., 2021). Usage of gender-fair language can reduce gender stereotyping by counteracting the representativity bias elicited by the generic

masculine in French (Vervecken et al., 2015) and German (Irmen & Roßberg, 2004; Horvath et al., 2016). Nevertheless, in the context of professional roles, researchers have identified a potential negative impact of the feminine form on women. This is due to the activation of stereotypical representations, which may result in less favorable judgments (Formanowicz et al., 2013) and lower salary estimates (Horvath et al., 2016). The present study contributes to this debate by examining how affective meanings ascribed to gendered social categories vary across semantic categories and affective space as defined by the EPA dimensions. The analysis is based on EPA ratings for 192 social categories in both genders, totaling 384 items, collected from 700 French and 700 German participants, representative of their countries by age, gender, and location.

Theoretical background

Osgood and colleagues identified and validated three key dimensions for affective meanings across cultures (1957, 1975). The Evaluation dimension (E) assesses a social category's goodness vs. evilness, Potency (P) its strength and power, and Activity (A) its dynamism vs. passivity. Studies have applied the EPA framework to link masculine/feminine traits (Langford & MacKinnon, 2000), roles (Kroska & Cason, 2019), and tasks (Kroska, 2003) with their affective meanings. However, these studies, conducted in English, capture gender-driven affective meanings indirectly through stereotypes (Kroska, 2003; Kroska & Cason, 2019) or perceived gender associations (Langford & MacKinnon, 2000). In contrast, to the best of our knowledge no EPA studies have been conducted in gendered languages like French and German where separate masculine and

feminine forms (e.g., der Arzt vs. die Ärztin) directly invoke and highlight gender, making it salient (Ridgeway, 2011).

In the following we present three pre-registered hypotheses[2] about the way gender is expected to be reflected in the affective meanings associated with social categories along the EPA dimensions. To do so, we draw on social psychological work based on attitudes (Eagly & Mladinic, 1989; Eagly et al., 1991) or on the "warmth/competence" framework of social perception (Glick et al., 2000; Fiske et al., 2002; Cuddy et al., 2008; Martin & Slepian, 2021).

Evaluation. Eagly and colleagues (Eagly & Mladinic, 1989; Eagly et al., 1991; Eagly & Mladinic, 1994) shows that women's as well as men's attitudes are more positive toward women than toward men, i.e., the "women are wonderful" effect. Similarly, using the Ambivalent Sexism Inventory (Glick & Fiske, 1997), studies reveal a bias for viewing women in traditional roles more favorably than men (Glick et al., 2000; 2004). This difference in attitudes toward gender reflects the segregation of roles whose evolutionary roots compelled women and men to assume communal and agentic traits, respectively (Eagly & Wood, 1999). This separation of positive and agentic characteristics among genders, also seen in other marginalized groups (Conway et al., 1996; Judd et al., 2005; Cuddy et al., 2008), implies that viewing lower-status groups more positively may compensate for societal structures that regard them as less capable in terms of agency (Ridgeway, 2019; Yzerbyt, 2018). Research using the Stereotype Content Model (Fiske et al., 2002) and ACT (Langford & MacKinnon 2000; Kroska 2001) supports women's positive evaluation. Thus, we hypothesize (HpE) that *the more positive evaluation*

ascribed to women in general will be systematically reflected in a more positive evaluation ascribed to the feminine versions of social categories.

Potency. Status and power are fundamental for understanding social interactions and their macro-social effects (Kemper & Collins, 1990). Repeatedly observing gender-segregated roles, especially men in power, leads to inferring power based on gender (Eagly & Steffen, 1984). From both an evolutionary (Schwartz & Rubel-Lifschitz, 2009) and a social role perspective (Eagly & Wood, 1999), such segregation presses men to engage in power- and status-seeking behaviors, fuelling expectations about men as more powerful and agentic. Research on the Stereotype Content Model finds men seen as "bad but bold", with their competence and agency counterbalanced by negative views (Glick et al., 2004). Within ACT, Heise (2007) found masculine forms rated as more potent in family-related social categories (e.g., mother-father). Langford & MacKinnon (2000) reported that stereotypically masculine traits were consistently rated as more potent compared to feminine traits. Therefore, we formulate our hypothesis as follows: *social categories presented in the feminine form will mostly be perceived as less potent compared to the corresponding masculine version of the same social categories* (HpP).

Activity. The effect of gender is less clear for activity. Men are perceived as more competent and agentic (Glick et al., 2004; Ridgeway, 2011), potentially raising the activity ratings of masculine categories. Conversely, feminine roles, seen as more communal, might imply greater dynamism and interpersonal agency, thus higher activity levels (Kroska, 2001). However, considering the strong correlation between potency and activity in social categories (Osgood et al., 1957; Cuddy et al., 2008; Rogers et al., 2013), we

hypothesize (HpA) that *social categories presented in the feminine form will be mostly perceived as less active than the corresponding masculine social categories*.

Cross-dimensional patterns

So far, we have proposed three one-dimensional hypotheses. According to these, feminine social categories should be rated as more positive (HpE), less potent (HpP), and less active (HpA) than their masculine counterparts. However, gender discrimination affects women in multiple ways (Ellemers, 2018; Barreto & Doyle, 2023), suggesting interdependence between the hypotheses. To understand how single-dimensional effects merge into cross-dimensional patterns, we refer to Figure 1's matrix. Given the high overlap between potency and activity, we focus on evaluation and potency. The *x*-axis of the matrix is defined by differences on the evaluation dimension between the feminine and masculine versions of the same social category, while the *y*-axis is defined by differences on the potency dimension.

Evolutionary and social role perspectives hold that society accepts and encourages dominance and power-seeking in men, but not in women. Conversely, society prescribes communality more strongly for women than for men (Eagly & Wood, 1999; Prentice & Carranza, 2002; Schwartz & Rubel-Lifschitz, 2009). As the more positive evaluation ascribed to the feminine versions of social categories (HpE) may compensate for the lower potency ascribed to them (HpP), many social categories can be expected to confirm both HpE and HpP, which corresponds to the benevolent sexism stereotype (Glick & Fiske, 1997).

Social categories that refute both hypotheses will sit in the lower right quadrant. Research demonstrates that when women are in powerful roles, their assertiveness is often exaggerated and perceived as dominance, which has a negative impact on their likability (Rudman, 1998; Eagly & Karau, 2002; Rudman & Fairchild, 2004; Rudman et al., 2012a; Williams & Tiedens, 2016). Therefore, social categories that contradict HpP are also expected to violate HpE, this way mimicking the case of hostile sexism (Glick & Fiske, 1997).

The other two quadrants represent cases confirming one hypothesis but not the other. The lower left quadrant shows feminine categories rated more positively (HpE confirmed) and more potent (HpP disconfirmed) than masculine ones. Prior research indicates this applies to stereotypically feminine roles interpreted by men, as in the cases of passive (Costrich et al., 1975) and modest men (Moss-Racusin et al., 2010), nurturing professors (Meltzer & McNulty, 2011), and male primary school teachers (Moss-Racusin & Johnson, 2016). Here, violating masculinity norms makes masculine categories seem less potent than feminine ones, without positive evaluation compensation (Rudman et al., 2012b).

Finally, In the upper right quadrant, feminine categories are viewed more negatively (HpE disconfirmed) and less potent (HpP confirmed). The glass escalator effect (Williams, 1992) exemplifies this, where men gain from benevolent attitudes by advancing in traditionally feminine roles (Meltzer & McNulty 2011; Gaunt, 2013).

Methods and materials

The analysis presented below is based on EPA ratings collected online in 2022 from n = 700 German and n = 700 French participants who rated the stimuli in German and French, respectively. We defined quotas with respect to gender, age, and region to obtain a stratified representation of both countries for each word set. Respondents were recruited through a commercial access panel (Bilendi S.A.) and received monetary compensation for their participation. The current results are the fruit of a secondary analysis run on a data set – the "French-German EPA dictionary" (Dametto et al., 2024) – openly available to researchers.

Stimuli construction

The social categories contained in the "French-German EPA dictionary" were sampled from the 929 social categories enlisted in the "2015 USA Combined Surveyor Dictionary" (Smith-Lovin et al., 2016), with the goal of covering the entire EPA space. Cluster analysis divided the US affective dictionary into five clusters, from which 188 categories were chosen[3]. The social categories were then translated into French and German gender-specific versions, e.g. a male artist: *ein Künstler* (German), *un artiste* (French); a female artist: *eine Künstlerin* (German), *une artiste* (French) through a back-translation approach. Additionally, four categories related to sexuality and gender identity (androgynous, homophobic, transgender, asexual) were included[4]. Therefore, the final dictionary consisted of 192 social category pairs, for a total of 384 gendered social categories. Social categories were split into ten subsets (37-40 words each), ensuring a gender balance. Each

subset was rated by a representative sample of 70 French and 70 German participants, totaling 140 participants, following a balanced incomplete block design.

Research design

Our research, rooted in ACT's classic rating procedure (Heise, 2007), aims to compare EPA affective meanings between gendered social categories (Heise, 2010). Our method expands Langford and MacKinnon's (2000) approach, exploring how gender affects EPA profiles in grammatically gendered languages. Additionally, our method aligns with Goldberg's (1968) paradigm, where only the target gender changes while assessing social categories on EPA dimensions (Rudman et al., 2012b). Crucially, we employed a between-subjects design, ensuring no participant rated both masculine (e.g., ein Student, un étudiant) and feminine (e.g., eine Studentin, une étudiante) versions of a category. Hence, the methodology avoids a comparison bias (Biernat, 2005) and decreases any demand characteristics (Orne, 1962; Corneille & Lush, 2023).

Rating procedure and semantic differential scales

Respondents rated social categories one at a time on three EPA dimension scales, with no option to skip, using bipolar semantic differentials. In the Supplementary Materials, the instructions as well as the rating instruments are presented. Participants knew the study aimed to evaluate affective perceptions of social categories, without clues on gender differences. We randomized the order of the identities and the direction of the EPA scales across participants. The bipolar anchors for the EPA dimensions have long been established in German (Schröder, 2011) and have recently been defined in French

(Dametto et al., 2023). The contrasting anchors of the bipolar scales are as follows: for evaluation *angenehm*/*plaisant* [pleasant], *freundlich*/*agréable* [friendly], *gut*/*bien* [good], versus *unangenehm*/*déplaisant* [unpleasant], *unfreundlich*/*désagréable* [unfriendly]. *schlecht*/*mal* [bad]; for potency: *kraftvoll*/*puissant* [powerful], *stark*/*fort* [strong], *groß*/*grand* [large] versus *zart*/*délicat* [gentle], *schwach*/*faible* [weak], *klein*/*petit* [small]; for activity: *schnell*/*rapide* [quick], *lebhaft*/*vif* [lively], *geräuschvoll*/*bruyant* [noisy] versus *langsam*/*lent* [slow], *träge*/*inerte* [inert], *still*/*silencieux* [quiet]. We used the following verbal labels to mark the 9-point scale: *etwas*/*un peu* [slightly], *ziemlich*/*plutôt* [quite], *sehr*/*très* [very], *äußerst*/*extrêmement* [extremely], with neutral/neutre marking the center.

Results

We first tested our three single-dimensional hypotheses and, secondly, explored cross-dimensional interactions. The article's Supplementary Materials offer a preliminary analysis on whether respondents' gender and language affect perceptions of gendered social categories. Findings indicate no interaction between the perceiver's gender and the social category's gender, for both German and French participants. Thus, while men and women tend to produce slightly different affective ratings (MacKinnon & Keating; 1989; Chapman et al., 2022), the difference between the affective meanings attributed to feminine and masculine versions of social categories is not affected by the gender of the perceiver. Therefore, we merged the ratings from both genders and both cultures.

Hypothesis testing: single-dimensional analysis

To test our hypotheses, we provide our results at two levels of analysis. (I) Cohen's *d* effect sizes were calculated for the EPA dimensions of each pair of social categories in order to quantify the differences in mean affective ratings between the masculine and feminine versions of the social categories in relation to the overall variability of the ratings[5]. A positive Cohen's *d* indicates the masculine version scored higher on a dimension, while a negative value indicates the feminine form was rated higher. Table 1 lists the strongest positive and negative Cohen's *d* values and social categories where grammatical gender doesn't influence affective ratings. We interpreted effect sizes using thresholds from Lovakov and Agadullina (2021), assessing how gender impacts affective perceptions across no, small, moderate, or strong effect levels for each hypothesis. Because variation in effect size among our 192 identities could be due to type II statistical error, we analyzed their distribution to assess consistency and direction.

(II) Participant-level analysis included a mixed-effects regression model, regressing each EPA dimension ratings (dependent variable) on social category gender (independent variable). The gender of the social category was contrast coded with weights equal to -0.5 (feminine) and +0.5 (masculine). Given the balanced incomplete block design's random allocation of participants and social categories to word subsets, we used crossed random effects for both.

Evaluation. Our hypothesis (HpE) was that feminine social categories would receive more positive evaluations. Table 2 shows mostly negative effect sizes, indicating feminine versions were indeed often rated more positively. However, gender's effect was

mainly null or small, exceeding a moderate threshold in just 14 of 192 cases. Female versions were rated clearly more positively for many negatively connotated social categories (e.g. vandal, mafioso, daredevil, robber, ex-convict). Next we focused on the distribution of effect sizes for all social categories in our dataset (see Supplementary Materials Figure S2). A one sample *t*-test revealed that the mean effect size (M = -.09, SD =0.16) was significantly different from zero, $t(191) = -8.417$, $p < .001$. In addition, the distribution was left-skewed (γ = -.41). Both results show that the gender effect is robust, with feminine categories, on average, rated more positively than masculine ones. In the same line, the mixed-model (Table 3) showed that participants generally rated masculine categories less positively (b = -.16, $t(192) = -11.74$, $p < .001$, 95% CI [-.19; -.13]). Thus, feminine versions of social categories receive higher average evaluation ratings, though the effect is moderate or large only in a few cases.

Potency. Following our hypothesis (HpP), we expected that masculine categories would be seen as more potent than their feminine counterparts. Table 2 shows that the effect sizes are not systematically positive: only in relatively few cases does gender have at least a small effect on potency, such that masculine versions are rated more potent than the feminine version. Mainly, these instances involve physical jobs (garbage collector, butcher, boxer), authority figures (soldier, police agent), seduction roles (top-model, playboy), and leadership (manager, capitalist). Against our hypothesis, 24 categories, including sexual identities (gay, androgynous, asexual, homosexual, queer, bisexual) and those linked to passivity (bum, malingerer) and creativity (artist, musician), are seen as more potent when interpreted by women, with effect sizes from small to moderate. A one-

sample *t*-test rejected the null hypothesis (M = .03, SD = .17) with $t(191) = 2.70$, $p = .007$, indicating masculine forms were rated more potent on average than feminine. Furthermore, the potency effect sizes were right-skewed (γ = .48; refer to Supplementary Materials Figure S3). In line with HpP, participant-level mixed-model (Table 4) reveals a significant yet small effect of social category gender on potency ratings (b = .05, $t(192) = 3.75$, $p < .001$, 95% CI [.02; .08]). Thus, the second hypothesis appears to be only marginally supported. Participants do rate the masculine versions of social categories as slightly more potent than the corresponding feminine version. Yet, the average effect size is close to zero, and only slightly more social categories were rated as more potent in the masculine version than would be expected by chance. Nevertheless, the distribution is right-skewed, suggesting that in some cases grammatical gender does bias potency ratings in favor of the masculine version.

Activity. Contrary to our hypothesis of feminine forms being seen as less active, effect sizes are mostly negative indicating that social categories were often viewed as more active when interpreted by women (Table 2). Among those cases, intellectual professions (judge, scientist, white collar worker, computer programmer, physician, priest, artist) stand out. The mean effect size (M = -.05, SD = .15) is significantly different from zero, $t(191) = -5.15$, $p < .001$, and slightly right skewed (γ = .19; see Supplementary Materials Figure S4). The mixed model confirmed that the masculine forms were rated less active (b = -.09, $t(192) = -6.73$, $p < .001$, 95% CI [-.12; -.07], see Table 5). Thus, the results did not support our third hypothesis, but showed the opposite pattern.

Exploratory analysis of cross-dimensional effects

Though gender's effects on social category perceptions were generally small, we used linear regression models to examine under which conditions the effect of gender is larger (Figures 2.1, 2.2, 2.3, Table 6). In each model, Cohen's *d* effect size for an EPA dimension served as the dependent variable. Independent variables included average EPA ratings (*Avg. E*, *Avg. P*, *Avg. A*) aggregated over both masculine and feminine forms. To identify which form was rated higher, three variables were contrast coded (*HigGenE*, *HigGenP*, *HigGenA*) with -0.5 for feminine and +0.5 for masculine. We also explored interaction effects between average EPA means and contrast variables (*Avg. E * HigGenE*, *Avg. P * HigGenP*, *Avg. A * HigGenA*), examining how the average ratings and higher ratings by gender interacted on each dimension.

The first model assessing the gender effect on evaluation (R^2 = .490, $F(9, 182)$ = 21.37, $p < .001$) shows a significant impact of *HigGenE* (b = -.24, $t(182)$ = -12.424, $p <$ .001, $\eta 2$ = .46, 95% *CI* [-.28; -.20]), indicating stronger gender effect sizes when feminine categories score higher in evaluation than masculine ones. Thus, the latter case is more likely due to chance. Nonetheless, effect size is unaffected by the average ratings on evaluation, potency, and activity, showing stability across the affective space. There is a significant but small interaction between average potency and *HigGenP* (b = -.04, $t(182)$ = -2.367, p = .019, $\eta 2$ = .03, 95% *CI* [-.08; -.01]) indicating that in case of social categories that were rated more potent in the feminine version, the higher the potency rating, the larger the effect size on evaluation.

The second model investigates the determinants of the effect size on potency (R^2 = .624, $F(9, 182) = 36.18$, $p < .001$) shows a significant effect of *HigGenP* (b = -.22, $t(182)$ = -11.751, $p < .001$, $\eta 2$ = .53, 95% *CI* [-.25; -.18]), indicating weaker gender effect sizes when feminine categories score higher in potency than masculine ones. *HigGenA* also shows a significant, though weaker, effect (b = -.05, $t(182)$ = -2.473, p = .014, $\eta 2$ = .06, 95% *CI* [-.10; -.01]). Moreover, average potency affects potency effect size. The more powerful the social category, the larger the potency difference in favor of the masculine version (= .03, $t(182)$ = 1.845, p = .067, $\eta 2$ = .02, 95% *CI* [.00; .06]). Thus, the absolute gender effect size is larger at the extremes, but in the opposite direction: it is larger and positive – indicating that the masculine versions are seen as more potent – when the social categories are potent. In contrast, it is larger and negative – indicating that the feminine versions are seens as more potent – when social categories are weak (i.e. the *Avg. P* is negative). Average evaluation also affects potency effect size (b = -.01, $t(182)$ = -1.872, p = .063, $\eta 2$ = .01, 95% *CI* [-.02; .00]), with negative categories showing a trend towards masculine form being seen as more potent, and positive categories trending towards feminine form being seen as more potent.

Finally, the third model examining gender effect on activity (R^2 = .656, $F(9, 182)$ = 38.63, $p < .001$) shows a significant effect of *HigGenA* (b = -.23, $t(182)$ = -11.946, $p <$ .001, $\eta 2$ = .57, 95% *CI* [-.27; -.19]), indicating stronger gender effect sizes when feminine categories score higher in activity than masculine ones. *HigGenP* also shows a significant, though weaker, effect (b = -.04, $t(182)$ = -2.636, p = .009, $\eta 2$ = .07, 95% *CI* [-.07; -.01]). Remarkably, average activity has a significant positive effect (b = .03, $t(182)$ = 2.346, p =

.020, $\eta 2$ = .03, 95% *CI* [.00; .06]). Thus, the more active the category, the greater the effect size, trending toward seeing the masculine form as more active. Lastly, average potency rating has a negative significant effect (b = -.03, $t(182)$ = -1.988, p = .048, $\eta 2$ = .02, 95% *CI* [-.06; .00]). Thus, the more potent the category, the smaller the effect size, trending toward seeing the feminine form as more active.

In summary, the analyses confirmed two out of the three hypotheses on the effect of gender on the affective meanings attributed to social categories. Our effect size analysis shows that while gender impacts on EPA dimensions are small, they reveal significant patterns. As expected, feminine versions of social categories tend to be rated more positively overall (*HpE*). This gender effect is constant along the entire range of the evaluation dimension. The second hypothesis (*HpP*) on gender's impact on potency was only partly confirmed: the masculine version of a social category is not perceived as systematically more potent than its feminine counterpart. Especially when social categories are rated as neutral on the potency dimension, the gender effect appears to be randomly distributed around zero. By contrast, the gender effect is stronger when social categories are either very potent or very impotent. Lastly, our hypothesis for activity was not supported (HpA), with feminine categories rated more active and gender effect sizes increasing oppositely at activity extremes. Contrary to our expectation, feminine social categories were rated as slightly more active, and we found a slight tendency for the intensity of the effect sizes to become larger – in the opposite direction of our expectation – at the extremes of the activity range.

Exploratory semantic analysis

Using Figure 1's theoretical matrix, we examine the combined impact of evaluation and potency rating differences between gendered social categories. Figure 3 locates social categories according to the effect size on the evaluation (*x*-axis) and potency (*y*-axis) dimensions. To identify categories that either support (upper left quadrant) or refute (lower right quadrant) both HpE and HpP hypotheses, we focus on 118 of 192 categories with effect sizes on evaluation or potency exceeding the small effect threshold (0.15) as defined by Lovakov and Agadullina (2021).

With feminine categories often rated more positively and slightly more often as less potent than masculine ones, many confirm both HpE and HpP, landing in the upper left quadrant. Moreover, many categories aligned with HpE but not HpP, with fewer cases in the remaining quadrants (Table 7). To validate our approach, we inspected the quadrants to detect meaningful semantic patterns, which will be commented on in the general discussion.

HpE and HpP confirmed (n = 55). Feminine versions of traditionally threatening (bully, gangster, mafioso, outlaw, prisoner, racist, pusher, robber, kidnapper, ex-convict, sadist, shoplifter) or somewhat intimidating (boxer, bouncer, bodyguard, neurotic, psychotic, psychiatric patient, punk) categories, along with leadership roles (banker, consultant, manager, governor, lobbyist, boss), are seen as less potent but more positive.

HpE and HpP disconfirmed (n = 16). Categories seen as less positive yet more powerful in feminine form include marginalized roles (e.g., homeless person, beggar) and sexual identities (e.g., asexual, transgender, queer, gay).

HpE confirmed and HpP disconfirmed (n = 36). In feminine form, categories like entrepreneur, politician, army officer, attorney, prosecutor attorney, judge, and scientist are viewed as both more positive and potent.

HpE disconfirmed and HpP confirmed (n = 11). We could not identify any clear semantic patterns related to the social identities located in this quadrant.

General discussion

This study explored how grammatical gender influences the affective meaning in terms of Evaluation, Potency and Activity connotations attached to social categories in French and German. Affective meanings are fundamental for the maintenance of social order (Heise et al., 2015) because they act as vectors for disseminating norms and values. Affective meanings shape social interactions by providing the fundamental sentiments for forming impressions (Heise, 2007) and building expectations (Ridgeway & Correll, 2006). By doing so they not only mirror stereotypes but contribute to their sedimentation within a culture (Schröder et al., 2013). This research, therefore, plays a pivotal role in understanding the subtle yet profound ways in which language and perception interlock to perpetuate gender norms.

Overall, while the direct impact of gender on affective meanings appears small at first glance, a closer look reveals nuanced, gender-specific patterns in how feminine and masculine social categories were perceived.

Evaluation. We found consistent evidence in support of our first hypothesis regarding the evaluation dimension. Social categories were often seen more positively in

feminine forms, suggesting a "women are wonderful" effect embedded in French and German affective meanings (Eagly & Mladinic, 1994; Langford & MacKinnon, 2000). This trend is consistent along the entire evaluation dimension, applying to good, neutral, and bad categories alike. Notably, we might expect that the feminine form of criminal and deviant social categories would be judged more harshly than the masculine one, due to higher evaluation standards for women (Biernat & Kobrynowicz, 1997). However, there is no such contrast effect, and these feminine social categories are assimilated to the reference standard, suggesting – so to speak – that women cannot be so evil (Biernat, 2005). These findings align with gender stereotype congruence theories (Rudman, 1998; Eagly & Karau, 2002), suggesting traits like communality and goodness are prescribed rather than inherent to gender hierarchy (Prentice & Carranza, 2002; Rudman et al., 2012a), thus preserving positive evaluations despite violations.

Potency. The hypothesis predicting masculine categories as more potent than feminine was partly confirmed. Our results align with more recent research arguing that the perceived competence gap between genders is narrowing (Ebert et al., 2014; Eagly et al., 2020). The analysis we presented shows a more nuanced effect of gender on potency, influenced by (I) the average potency rating and (II) gender salience. In the former case (I), gender differences are larger when either weak or strong social categories are considered, but smaller or even absent when social categories are neutral on the potency dimension. Our findings limit the "men are bold" stereotype (Glick et al., 2004) to potent categories (e.g. banker, manager, millionaire, governor, police agent, boss). Conversely, masculine forms of weak categories (e.g., coward, beggar, looser) are seen as less potent

than their feminine counterparts. One explanation is that traits like agency, dominance, and assertiveness – underpinning potency – are central to gender hierarchy, proscribing men from showing weakness (Rudman, 1998; Eagly & Karau, 2002; Prentice & Carranza, 2002; Rudman et al., 2012b; Krauth-Gruber & Aelenei, 2022). By violating the potency proscription, weak masculine social categories are perceived as undermining the gender hierarchy, which cause them to be contrasted away from the masculine stereotype (Biernat & Kobrynowicz, 1997; Biernat, 2005). From this perspective, the evaluation and potency dimensions operate differently.

However, gender still affects neutral potency-rated social categories when (II) semantics make gender salient (Ridgeway, 2011). Social categories related to sexual identities and behaviors (e.g., queer, homosexual, asexual) are seen as less potent in masculine forms, suggesting that masculine social categories that openly challenge the gender hierarchy are contrasted away from the gender stereotype. Potency is thus closely linked to gender (Eagly & Steffen, 1984; Martin & Slepian, 2021): Gender differences emerge when potency is salient, and potency differences emerge when social categories make the gender structure of society salient.

Activity. Feminine social categories were consistently rated more active. Contrary to our hypothesis, feminine categories were rated more active, despite strong correlation between potency and activity (r = .84). Differences in average ratings for potency (masculine M = .59, feminine M = .53) and activity (masculine M = .64, feminine M = .73) illustrate distinct gender interactions with these dimensions. Following Kroska (2001), one explanation may be that feminine social categories seem more communal, and thus more

dynamic and interpersonally agentic, and active. In addition, Freeland and Hoey (2018) suggest potency, indicating symbolic power, better reflects gender status differences than activity. This suggests that activity levels attributed to women in professional roles (e.g. scientists, programmers, physicians) reflect their perceived effort towards achieving these positions (Tiedemann, 2000; Ellemers, 2018).

In sum, despite small effects, clear patterns emerge showing the influence of gender on the affective meanings attributed to social categories. These patterns show how affective knowledge enacted in everyday interactions is intertwined with gender stereotypes, consistent with theories proposing that social order follows from affective meanings (Heise et al., 2015).

Exploratory analysis

This paper also explores the link between social categories' affective meanings and established gender discrimination patterns. Though gender's effect on affective perception is small, cumulative small biases can result in significant societal impacts (Payne et al., 2017). Analysis shows that social categories mostly confirm HpE and HpP, indicating benevolent sexism (Glick et al., 2000). Among these, we found many deviant and criminal categories. The differing affective views on negative social categories may stem from men's higher criminal statistics (Spjeldnes & Goodkind, 2009; Statistisches Bundesamt, 2022a; 2022b; Ministère de la justice, 2022), likely causing women's crimes to seem more affectively dissonant (Boyle, 2023) and less severe (Rachlinski & Wistrich, 2021).

Our results align with findings regarding marginalized and sexual social categories that violate both HpE and HpP. Sexual categories are subject to gender-specific contrast

effects and, as hypothesized in previous research (Rudman & Fairchild, 2004; Rudman et al., 2012b), are subject to overt hostile sexism. They are systematically rated as more powerful and usually less positive when interpreted by women rather than men. Thus, by reversing the prescription for women to be 'wonderful' and for men to be 'bold', these social categories violate both HpE and HpP.

Yet, our findings only partially align with prior research on bias against female leaders (Eagly & Karau, 2002; Rudman et al., 2012a). Our data shows some leadership roles (banker, consultant, manager, governor, lobbyist, boss) support both HpE and HpP, while others (entrepreneur, politician, army officer, attorney, prosecutor attorney, judge, scientist) confirm HpE but not HpP. Just one category (minister) contradicts both HpE and HpP, as would theoretically be expected for leadership roles. A possible explanation may be that in "business and work" contexts (MacKinnon & Heise, 2010), negatively viewed leadership roles (e.g., banker, lobbyist) rate more positive in feminine forms, as women are perceived less able to fulfill these 'tough' roles' demands. Contrarily, state institution leaders (e.g., judge, prosecutor attorney) and positive leaders (e.g., entrepreneur, scientist) in the feminine form are less likely to conflict with role expectations (Eagly & Karau, 2002). Our findings align with Bye et al. (2022), noting no perceived competence gap in academic, political, and leadership roles across genders. This view resonates with theories positing competence and potency's high context dependency (Martin & Slepian, 2021), meaning ascribed potency to women and men varies with institutional leadership demands.

Limitations

Our study, using a gender, age, and region-representative sample from Germany and France, yields findings specific to these nations, not extendable to all grammatical gender languages. Furthermore, our analysis, with a statistical power of .80 to detect effect sizes of d = .25 at .001 significance level (see Supplementary Materials "Power Analysis"), shows few cases of moderate or large gender effects. This way, we can confidently make the argument that there are consistent similarities in the affective meanings attributed to the masculine and feminine forms of social categories. Nevertheless, our analysis shows that gender has small effects on the social categories we studied, and that these effects appear to be consistent with previous theorizing and documented patterns of gender discrimination. However, more power is needed to definitively assess the magnitude and direction of these small effects.

Contributions

The present study contributes to the social psychological but also cultural discussion on gender, especially the public debate about the use of gendered forms in grammatical gender languages. While the usage of the generic masculine form leads to overrepresentation of men against women (Irmen & Roßberg, 2004; Vervecken et al., 2015; Sczesny et al., 2016), researchers also noted the negative impact of gendered language on women (Formanowicz et al., 2013, Horvath et al., 2016). In this context, our results provide encouraging insights: the effect of gender seems to be mostly small. Mostly, the affective meanings attached to gendered social categories resemble a *t*-shaped rather than a bimodal distribution (Figure 4). Whether the increased use of feminine forms

will lead to a drifting apart of affective perceptions of gendered social categories, or to a merging of affective meanings previously biased by masculine overrepresentation, is an open question that should be explored in future research.

Conclusions

This research investigated the interaction between gender and affective meanings attached to social categories. Our approach merged Heise's cultural survey method (2007) with Goldberg's paradigm (1968) to test three hypotheses concerning the differences in affective meanings between feminine and masculine versions of social categories. Overall, the differences between gender forms were small. Feminine forms were generally seen as more positive, in part less potent, and somewhat more active. Moreover, differences on the evaluation and potency dimensions combine to reproduce well-known patterns of gender biases. Our research provides a detailed examination of how gender stereotypes are embedded in the affective connotations within French and German linguistic contexts, contributing to discussions on gender language and equality.

Notes

[1] Aligned with social psychology, we employ "social categories" to denote the fundamental labels society uses to classify individuals (Haslam et al., 2000). Within the Affect Control Theory (ACT) framework, MacKinnon and Heise (2010) and Smith-Lovin (2007) embrace a definition akin to that of Thoits and Virshup (1997, 106-107), who characterize social identities as "socially-constructed and meaningful categories acknowledged by individuals as representative of themselves or their group." This paper maintains the use of "social category." Consequently, we include professional roles (e.g., insurance agent, baker), identities derived from psychological or emotional traits (e.g., hothead, paranoid), and behaviors (e.g., criminal, miser, know-it-all) under the umbrella of social categories. The term gendered social categories will refer to the grammatically or lexically gender-specific versions of these social categories.

[2] The pre-registration is available at https://osf.io/fq9a2/?view_only=4249aefea99245368180e8c387553345. All data and the analysis script are available at: https://osf.io/jxhs2/?view_only=7df2fc30b5c543619a4781ad5fd1f6b1. Note that the present hypotheses are pre-registered in the second part of the pre-registration form. The verification of the first hypothesis was presented elsewhere (Dametto et al., 2023).

[3] A more precise description of this approach is provided in the Supplementary Information to Dametto and colleagues (2023).

[4] While acknowledging the effort of categories like androgynous, queer, or transgender to transcend gender binaries, interactions often revert to grammatical or semantic gender

assignments. Thus, we translated non-binary identities to German and French, maintaining gender markers: "ein androgyner Mann," "eine androgyne Frau," "un homme androgyne," "une femme androgyne" for androgynous; "ein transgender Mann," "eine transgender Frau," "un transgenre," "une transgenre" for transgender; and "ein queer Mann," "eine queer Frau," "un homme queer," "une femme queer" for queer.

[5] In the pre-registration, we planned to assess differences based on mean differences. This analysis is now proposed in the Supplementary Materials (“Statistical significance analysis”) to this article. In the main article, we decided to use Cohen's d effect sizes as they also take into account the variability in the data.

References

Ambrasat, Jens, Christian Von Scheve, Markus Conrad, Schauenburg Gesche, and Tobias Schröder. 2014. ‘‘Consensus and Stratification in the Affective Meaning of Human Sociality.’’ *Proceedings of the National Academy of Sciences of the United States of America* 111(22):8001–8006. doi:10.1073/pnas.1313321111.

Barreto, Manuela, and David M. Doyle. 2023. Benevolent and hostile sexism in a shifting global context. *Nature reviews psychology*, 2(2), 98-111. doi: 10.1038/s44159-022-00136-x.

Berger, Peter L., and Thomas Luckmann [1966] 1991. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. London: Penguin.

Biernat, Monica. 2005. *Standards and expectancies: Contrast and assimilation in judgments of self and others*. New York:Routledge Taylor & Francis Group.

Biernat, Monica, and Diane Kobrynowicz. 1997. Gender-and race-based standards of competence: lower minimum standards but higher ability standards for devalued groups. *Journal of personality and social psychology*, 72(3), 544. doi: 10.1037//0022-3514.72.3.544.

Boyle, Kaitlin M. 2023. Doing Gender, Avoiding Crime: The Gendered Meaning of Criminal Behavior and the Gender Gap in Offending in the United States. *Social Psychology Quarterly*. doi: 10.1177/01902725231167845.

Bye, Hege H., Vera V. Solianik, Martine Five, & Mehri S. Agai. 2022. Stereotypes of women and men across gender subgroups. *Frontiers in Psychology*, 13, 881418. doi: 10.3389/fpsyg.2022.881418.

Chapman, Robert M., Margaret N. Gardner, and Megan Lyons. 2022. Gender differences in emotional connotative meaning of words measured by Osgood's semantic differential techniques in young adults. *Humanities and Social Sciences Communications*, 9(1), 1-11. doi: 10.1057/s41599-022-01126-3.

Charlesworth, Tessa E. S., Victor Yang, Thomas C. Mann, Benedek Kurdi, and Mahzarin R. Banaji. 2021. Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychological Science*, 32(2), 218-240. doi: 10.1177/0956797620963619.

Conway, Michael, Teresa M. Pizzamiglio, and Lauren Mount. 1996. Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, 71(1), 25–38. doi: 10.1037/0022-3514.71.1.25.

Corneille, Olivier, and Peter Lush. 2023. Sixty years after Orne's american psychologist article: a conceptual framework for subjective experiences elicited by demand characteristics. *Personality and Social Psychology Review*, 27(1), 83-101. doi: 10.1177/10888683221104368.

Costrich, Norma, Joan Feinstein, Louise Kidder, Jeanne Marecek, and Linda Pascale. 1975. When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of experimental social psychology*, 11(6), 520-530. doi: 10.1016/0022-1031(75)90003-7.

Cuddy, Amy J. C., Susan T. Fiske, and Peter Glick. 2008. Warmth and competence as universal dimensions of social perception: The stereotype content model and the

BIAS map. *Advances in experimental social psychology*, 40, 61-149. doi: 10.1016/S0065-2601(07)00002-0.

Dametto, Diego, Luc Vieira, Tobias Schröder, and Christophe Blaison. 2023. Charlemagne's Legacy: A Consensus Analysis of Affective Meanings in French and German Culture. *Social Psychology Quarterly*. doi: 10.1177/01902725231205855.

Dametto, Diego, Luc Vieira, Christophe Blaison, and Tobias Schröder. 2024. *French German EPA dictionary [Data set]*. PsychArchives. doi: 10.23668/psycharchives.13567.

Eagly, Alice H., and Steven J. Karau. 2002. Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3), 573. doi: 10.1037//0033-295X.109.3.573.

Eagly, Alice H., and Antonio Mladinic. 1989. Gender stereotypes and attitudes toward women and men. *Personality and social psychology bulletin*, 15(4), 543-558. doi: 10.1177/0146167289154008.

Eagly, Alice H., and Antonio Mladinic. 1994. Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European review of social psychology*, 5(1), 1-35. doi: 10.1080/14792779543000002.

Eagly, Alice H., Antonio Mladinic, and Stacey Otto. 1991. Are women evaluated more favorably than men? An analysis of attitudes, beliefs, and emotions. *Psychology of women quarterly*, 15(2), 203-216. doi: 10.1111/j.1471-6402.1991.tb00792.x.

Eagly, Alice H., Christa Nater, David I. Miller, Michèle Kaufmann, & Sabine Sczesny. 2020. Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American psychologist*, 75(3), 301. doi: 10.1037/amp0000494.

Eagly, Alice H., and Valerie J. Steffen. 1984. Gender stereotypes stem from the distribution of women and men into social roles. *Journal of personality and social psychology*, 46(4), 735. doi: 10.1037/0022-3514.46.4.735.

Eagly, Alice H., and Wendy Wood. 1999. The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American psychologist*, 54(6), 408. doi: 10.1037/0003-066X.54.6.408.

Ebert, Irena D., Melanie C. Steffens, and Alexandra Kroth. 2014. Warm, but maybe not so competent?—Contemporary implicit stereotypes of women and men in Germany. *Sex roles*, 70, 359-375. doi: 10.1007/s11199-014-0369-5.

Ellemers, Naomi. 2018. Gender stereotypes. *Annual review of psychology*, 69, 275-298. doi: 10.1146/annurev-psych-122216-011719.

Fiske, Susan, Amy Cuddy, Peter Glick, and Xu Jun. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902. doi: 10.1037/0022-3514.82.6.878.

Formanowicz, Magdalena, Sylwia Bedynska, Aleksandra Cisłak, Friederike Braun, and Sabine Sczesny. 2013. Side effects of gender-fair language: How feminine job

titles influence the evaluation of female applicants. *European Journal of Social Psychology*, 43(1), 62-71. doi: 10.1002/ejsp.1924.

Freeland, Robert E., and Jesse Hoey. 2018. The structure of deference: Modeling occupational status using affect control theory. *American Sociological Review*, 83(2), 243-277. doi: 10.1177/0003122418761857.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635-E3644. doi: 10.1073/pnas.1720347115.

Gaunt, Ruth. 2013. Breadwinning moms, caregiving dads: Double standard in social judgments of gender norm violators. *Journal of family issues*, 34(1), 3-24. doi: 10.1177/0192513X12438686.

Glick, Peter, Susan T. Fiske. 1997. Hostile and benevolent sexism: Measuring ambivalent sexist attitudes toward women. *Psychology of women quarterly*, 21(1), 119-135. doi: 10.1111/j.1471-6402.1997.tb00104.x.

Glick, Peter, Susan T. Fiske, Antonio Mladinic, Jose L. Saiz, Dominic Abrams, Barbara Masser, Bolanle Adetoun, Johnstone E. Osagie, Adebowale Akande, Amos Alao, Annetje Brunner, Tineke M. Willemsen, Kettie Chipeta, Benoit Dardenne, Ap Dijksterhuis, Daniel Wigboldus, Thomas Eckes, Iris Six-Materna, Francisca Exp6sito, Miguel Moya, Margaret Foddy, Hyun-Jeong Kim, Maria Lameiras, Maria Jose Sotelo, Angelica Mucchi-Faina, Myrna Romani, Nuray Sakalh, Bola Udegbe, Mariko Yamamoto, Miyoko Ui, Maria Cristina Ferreira, and Wilson Lopez Lopez. 2000. Beyond prejudice as simple antipathy: Hostile and benevolent

sexism across cultures. *Journal of Personality and Social Psychology*, 79(5), 763–775. doi: 10.1037/0022-3514.79.5.763.

Glick, Peter. Lameiras, Maria Fiske, Susan T. Eckes, Thomas Masser, Barbara Volpato, Chiara Manganelli, Anna Maria Pek, Jolynn C. X. Huang, Li-li Sakalli-Uğurlu, Nuray Castro, Yolanda Rodriguez D'Avila Pereira, Maria Luiza Willemsen, Tineke M. Brunner, Annetje Six-Materna, and Robin Wells. 2004. Bad but Bold: Ambivalent Attitudes Toward Men Predict Gender Inequality in 16 Nations. *Journal of Personality and Social Psychology*, 86(5), 713–728. doi: 10.1037/0022-3514.86.5.713.

Goffman, Erving. 1959. *The presentation of self in everyday life*. Garden City: Double Day.

Goldberg, Philip. 1968. Are women prejudiced against women? *Trans-action*, 5(5), 28-30. doi: 10.1007/BF03180445.

Gygax, Pascal Mark, Daniel Elmiger, Sandrine Zufferey, Alan Garnham, Sabine Sczesny, Lisa Von Stockhausen, Frederike Braun, and Jane Oakhill. 2019. A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in Psychology* 10, 1604. doi: 10.3389/fpsyg.2019.01604.

Haslam, Nick, Louis Rothschild, and Donald Ernst. 2000. Essentialist beliefs about social categories. *British Journal of social psychology*, 39(1), 113-127. doi: 10.1348/014466600164363.

Heise, David R. 2007. *Expressive Order: Confirming Sentiments in Social Actions*. Boston: Springer.

Heise, David R. 2010. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. Hoboken: John Wiley & Sons.

Heise, David R., Neil J. MacKinnon, and Scholl Wolfgang. 2015. ‘‘Identities, Roles, and Social Institutions.’’ Pp. 165–88 in *Order on the Edge of Chaos*, edited by E. J. Lawler, S. R. Thye and J. Yoon. New York: Cambridge University Press.

Horvath, Lisa K., Elisa F. Merkel, Anne Maass, and Sabine Sczesny. 2016. Does gender-fair language pay off? The social perception of professions from a cross-linguistic perspective. *Frontiers in psychology*, 6, 2018. doi: 10.3389/fpsyg.2015.02018.

Irmen, Lisa, and Nadia Roßberg. 2004. Gender markedness of language: The impact of grammatical and nonlinguistic information on the mental representation of person information. *Journal of Language and Social Psychology*, 23(3), 272-307. doi: 10.1177/0261927X04266810.

Judd, Charles M., Laurie James-Hawkins, Vincent Yzerbyt, and Kashima Kashima. 2005. Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6), 899–913. doi: 10.1037/0022-3514.89.6.899.

Kemper, Theodore D., and Randall Collins. 1990. Dimensions of microinteraction. *American Journal of Sociology*, 96(1), 32-68. doi: 10.1086/229492.

Krauth-Gruber, Silvia, and Cristina Aelenei. 2022. An intersectional approach to gender and social class stereotypes. Retrieved [09.11.2023] from:

https://www.researchgate.net/publication/360823243_An_intersectional_approach_to_gender_and_social_class_stereotypes.

Kroska, Amy. 2001. Do we have consensus? Examining the relationship between gender ideology and role meanings. *Social Psychology Quarterly*, 18-40. doi: 10.2307/3090148.

Kroska, Amy. 2003. Investigating gender differences in the meaning of household chores and child care. *Journal of marriage and family*, 65(2), 456-473. doi: 10.1111/j.1741-3737.2003.00456.x.

Kroska, Amy, and Trent C. Cason. 2019. The gender gap in business leadership: Exploring an affect control theory explanation. *Social Psychology Quarterly*, 82(1), 75-97. doi: 10.1177/0190272518806292.

Langford, Tom, and Neil J. MacKinnon. 2000. The affective bases for the gendering of traits: Comparing the United States and Canada. *Social Psychology Quarterly*, 34-48. doi: 10.2307/2695879.

Lewis, Molly, and Gary Lupyan. 2020. Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature human behaviour*, 4(10), 1021-1028. doi: 10.1038/s41562-020-0918-6.

Lovakov, Andrey, and Elena R. Agadullina. 2021. Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, 51(3), 485-504. doi: 10.1002/ejsp.2752.

MacKinnon, Neil J., and David R. Heise. 2010. *Self, identity, and social institutions*. New York: Palgrave MacMillan.

MacKinnon, Neil J., and Leo Keating. 1989. The Structure of Emotions: Canada - United States Comparisons. *Social Psychology Quarterly* 52(1):70–83. doi:10. 2307/2786905.

Martin, Ashley E., and Michael L. Slepian. 2021. The primacy of gender: Gendered cognition underlies the big two dimensions of social cognition. *Perspectives on Psychological Science*, 16(6), 1143-1158. doi: 10.1177/1745691620904961.

Mead, George H. 1934. *Mind, self, and society* (Vol. 111). Chicago: University of Chicago press.

Meltzer, Andrea L., and James K. McNulty. 2011. Contrast effects of stereotypes:"Nurturing" male professors are evaluated more positively than "nurturing" female professors. *The Journal of Men's Studies*, 19(1), 57-64. doi: 10.3149/jms.1901.57.

Ministère de la justice. 2022. *Séries statistiques des personnes placées sous main de justice 1980 - 2022*. Retrieved [09.11.2023] from: https://www.justice.gouv.fr/sites/default/files/migrations/portail/art_pix/PPSMJ_2022_vf.pdf.

Moss-Racusin, Corinne A., and Elizabeth R. Johnson. 2016. Backlash against male elementary educators. *Journal of Applied Social Psychology*, 46(7), 379-393. doi: 10.1111/jasp.12366.

Moss-Racusin, Corinne A., Julie Phelan, and Laurie A. Rudman. 2010. When men break the gender rules: status incongruity and backlash against modest men. *Psychology of Men & Masculinity*, 11(2), 140. doi: 10.1037/a0018093.

Orne, Martin T. [1962] 2009. “Demand characteristics and the concept of quasi-controls”. Pp. 110-137 in: *Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow’s classic books*, edited by Robert Rosenthal and Ralph L. Rosnow. New York: Oxford University Press.

Osgood, Charles E., William H. May, Samuel Miron Murray, and Murray S. Miron. 1975. *Cross-Cultural Universals of Affective Meaning*. Urbana: University of Illinois Press.

Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.

Payne, Keith B., Heidi A. Vuletich, and Kristjen B. Lundberg. 2017. The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28(4), 233-248. doi: 10.1080/1047840X.2017.1335568.

Prentice, Deborah A., and Erica Carranza. 2002. What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of women quarterly*, 26(4), 269-281. doi: 10.1111/1471-6402.t01-1-00066.

Rachlinski, Jeffrey J., and Andrew J. Wistrich. 2021. Benevolent sexism in judges. *San Diego Law Review*, 58, 101. Retrieved [09.11.2023] from: https://ssrn.com/abstract=3852914.

Ridgeway, Cecilia L. 2011. *Framed by gender: How gender inequality persists in the modern world*. New York: Oxford University Press.

Ridgeway, Cecilia L. 2019. *Status: Why Is It Everywhere? Why Does It Matter?* New York: Russell Sage Foundation.

Ridgeway, Cecilia L., and Shelley J. Correll. 2006. Consensus and the creation of status beliefs. *Social Forces*, 85(1), 431-453. doi: 10.1353/sof.2006.0139.

Rogers, Kimberly B., Tobias Schröder, and Wolfgang Scholl. 2013. The affective structure of stereotype content: Behavior and emotion in intergroup context. *Social Psychology Quarterly*, 76(2), 125-150. doi: 10.1177/0190272513480191.

Rudman, Laurie A. 1998. Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology*, 74(3), 629. doi: 10.1037//0022-3514.74.3.629.

Rudman, Laurie A., and Kimberly Fairchild. 2004. Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *Journal of personality and social psychology*, 87(2), 157. doi:10.1037/0022-3514.87.2.157.

Rudman, Laurie A., Corinne A. Moss-Racusin, Julie E. Phelan, and Sanne Nauts. 2012a. Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of experimental social psychology*, 48(1), 165-179. doi: 10.1016/j.jesp.2011.10.008.

Rudman, Laurie A., Corinne A. Moss-Racusin, Peter Glick, and Julie E. Phelan. 2012b. Reactions to vanguards: Advances in backlash theory. *Advances in experimental social psychology*, 45, 167-227. doi: 10.1016/B978-0-12-394286-9.00004-4.

Schröder, Tobias, Kimberly B. Rogers, Shuichirou Ike, Julija Mell, and Wolfgang Scholl. 2013. ‘‘Affective Meanings of Stereotyped Social Groups in Cross-Cultural

Comparison.'' *Group Processes & Intergroup Relations* 16(6):717–33. doi:10.1177/ 1368430213491788.

Schwartz, Shalom H., and Tammy Rubel-Lifschitz. 2009. Cross-national variation in the size of sex differences in values: effects of gender equality. *Journal of personality and social psychology*, 97(1), 171. doi: 10.1037/a0015546.

Sczesny, Sabine, Magda Formanowicz, and Franziska Moser. 2016. Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in psychology* 7, 25. doi: 10.3389/fpsyg.2016.00025.

Smith-Lovin, Lynn. 2007. The strength of weak identities: Social structural sources of self, situation and emotional experience. *Social psychology quarterly*, 70(2), 106-124. doi: 10.1177/019027250707000203.

Smith-Lovin, Lynn, Dawn T. Robinson, Bryan C. Cannon, Jesse K. Clark, Robert Freeland, Jonathan H. Morgan, and Kimberly B. Rogers. 2016. ''Mean Affective Ratings of 929 Identities, 814 Behaviors, and 660 Modifiers by University of Georgia and Duke University Undergraduates and by Community Members in Durham, NC, in 2012-2014.'' University of Georgia. http:// research.franklin.uga.edu/act/.

Spjeldnes, Solveig, and Sara Goodkind. 2009. Gender differences and offender reentry: A review of the literature. *Journal of Offender Rehabilitation*, 48(4), 314-335. doi: 10.1080/10509670902850812.

Statistisches Bundesamt. 2022a. *Anzahl der rechtskräftig verurteilten Personen in Deutschland nach Geschlecht von 2011 bis 2021*. Retrieved [09.11.2023] from:

https://de.statista.com/statistik/daten/studie/1068769/umfrage/rechtskraeftig-verurteilte-personen-in-deutschland-nach-geschlecht/.

Statistisches Bundesamt. 2022b. *Anzahl der Strafgefangenen in den Justizvollzugsanstalten in Deutschland nach Geschlecht und Alter am 31. März 2022 (geschlossener und offener Vollzug)*. Retrieved [09.11.2023] from: https://de.statista.com/statistik/daten/studie/1353153/umfrage/strafgefangene-in-deutschland-nach-geschlecht-und-alter/.

Tiedemann, Joachim. 2000. Gender-related beliefs of teachers in elementary school mathematics. *Educational studies in Mathematics*, 41(2), 191-207. doi: 10.1023/A:1003953801526.

Thoits, Peggy A., and Lauren Virshup. 1997. Me's and we's: Forms and functions of social identities. Pp. 106-33 in *Self and identity: Fundamental issues*, edited by R. D. Ashmore, and L. Jussim. New York: Oxford University Press.

Vervecken, Ddries, Pascal M. Gygax, Ute Gabriel, Matthias Guillod, and Bettina Hannover. 2015. Warm-hearted businessmen, competitive housewives? Effects of gender-fair language on adolescents' perceptions of occupations. *Frontiers in psychology*, 6, 1437. doi: 10.3389/fpsyg.2015.01437.

Williams, Christine L. 1992. The glass escalator: Hidden advantages for men in the "female" professions. *Social problems*, 39(3), 253-267. doi: 10.2307/3096961.

Williams, Melissa J., and Larissa Tiedens. 2016. The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, 142(2), 165. doi: 10.1037/bul0000039.

Yzerbyt, Vincent. 2018. The dimensional compensation model: Reality and strategic constraints on warmth and competence in intergroup perceptions. In: *Agency and Communion in Social Psychology*, edited by A. Abele and B. Wojciszke, pp. 126-141. London: Routledge

Tables

Table 1: Top 3 most positive, negative, and neutral Cohen's *d* effect sizes by dimension

Dimension	*Strongest negative effects*	*Effect size close to zero*	*Strongest positive effects*
Evaluation	vandal, *d*=-.69 mafioso, *d*=-.50 pornographer, *d*= -.49 & prosecuting attorney, *d*= -.49	receptionist, *d*=0 flight attendant, *d*=0 scoundrel, *d*=0	bully=.31 garbage collector, *d*=.24 beggar, *d*=.23 & transgender, *d*=.23
Potency	gay, *d*=-.42 androgyn person, *d*=-.38 asexual, *d*=-.37	conformist, *d*=0 retailer, *d*=0 receptionist, *d*=0	top-model, *d*=.64 playboy, *d*=.54 mafioso, *d*=.50
Activity	janitor, *d*=-.38 truck driver, *d*=-.38 judge, *d*=-.38	flight attendant, *d*=0 homeless, *d*=0 laid off worker, *d*=0	garbage collector, *d*=.32 millionaire, *d*=.31 playboy, *d*=.31
	← Feminine form rated higher		**Masculine form rated higher →**

Table 2: Number of Cohen's *d* effect sizes by dimension, direction, and intensity

	Evaluation		*Potency*		*Activity*	
Effect size	*F d<0*	*M d>0*	*F d<0*	*M d>0*	*F d<0*	*M d>0*
no effect *abs(d)* < .15	81	37	68	56	85	41
small .15 < *abs(d)* < .36	49	11	21	38	44	18
moderate .36 < *abs(d)* < .65	13	0	3	6	4	0
large *abs(d)* > .65	1	0	0	0	0	0

Note: F → Feminine version of the social category rated higher on given dimension; M → Masculine version of the social category rated higher on given dimension. The magnitude of the effect sizes is based on Lovakov and Agadullina (2021).

Table 3: Mixed-effects regression coefficients for social categories' evaluation

Effects	*Estimate*	*SE*	*t*	*95% CI*	
				LL	*UL*
(Intercept)	-.05	.01	-.39	-.28	.19
Gender of the social category: *Masculine form*	-.16***	.01	-11.74	-.19	-.13
Participants: 1,400; Social category pairs: 192					

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1. SE = standard error; t = *t*-statistic; CI = confidence interval; LL = lower limit; UL = upper limit.

Table 4: Mixed-effects regression coefficients for social categories' potency

Effects	*Estimate*	*SE*	*t*	*95% CI*	
				LL	*UL*
(Intercept)	.56	.07	7.69	.42	.70
Gender of the social category: *Masculine form*	.05***	.01	3.75	.02	.08
Participants: 1,400; Social category pairs: 192					

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1. SE = standard error; t = *t*-statistic; CI = confidence interval; LL = lower limit; UL = upper limit.

Table 5: Mixed-effects regression coefficients for social categories' activity

Effects	*Estimate*	*SE*	*t*	*95% CI*	
				LL	*UL*
(Intercept)	.69	.07	9.89	.55	.82
Gender of the social category: *Masculine form*	-.09***	.01	-6.73	-.12	-.07
Participants: 1,400; Social category pairs: 192					

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1. SE = standard error; t = *t*-statistic; CI = confidence interval; LL = lower limit; UL = upper limit.

Table 6: Regression analysis with average ratings, and "highest gender" as independent variable, and the gender effect size as dependent variable

	Evaluation						*Potency*						*Activity*					
Effects	*Est*	*SE*	*95% CI*		*p*		*Est*	*SE*	*95% CI*		*p*		*Est*	*SE*	*95% CI*		*p*	
			LL	*UL*					*LL*	*UL*					*LL*	*UL*		
(Intercept)	-.03	.01	-.05	-.00	.025	*	.01	.01	.01	.04	.175	***	-.01	.01	-.03	.00	.124	
Avg. E	.00	.00	-.01	.02	.379		-.01	.01	-.02	.00	.063	.	.00	.00	-.01	.01	.628	
Avg. P	-.02	.02	-.05	.02	.342		.03	.02	.00	.06	.066	.	-.03	.01	-.06	.00	.048	*
Avg. A	.00	.02	-.04	.03	.762		.00	.02	-.03	.03	.989		.03	.01	.00	.06	.020	*
HigGenE Feminine	-.24	.02	-.28	-.20	.000	***	.01	.02	-.02	.05	.402		-.02	.01	-.05	.01	.161	
HigGenP Feminine	-.01	.01	-.05	.03	.583		-.22	.02	-.25	-.18	.000	***	-.03	.02	-.08	.01	.009	**
HigGenA Feminine	.00	.01	-.05	.04	.908		-.05	.02	-.10	-.01	.014	*	-.23	.02	-.27	-.19	.000	***
*Avg. E * HigGenE* Avg. E * *Feminine*	.01	.01	-.01	.03	.424		.02	.01	.00	.04	.054	.	-.01	.01	-.03	.01	.389	
*Avg. P * HigGenP* Avg. P * *Feminine*	-.04	.02	-.08	.01	.019	*	-.02	.02	-.05	.01	.257		-.01	.01	-.04	.01	.341	
*Avg. A * HigGenA* Avg. A * *Feminine*	.02	.02	-.02	.06	.434		.01	.02	-.04	.03	.734		.00	.02	-.03	.03	.970	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1. Est. = Estimate; SE = standard error; CI = confidence interval; LL = lower limit; UL = upper limit; p = p-value.

Table 7: Social categories with at least a small effect size (abs(*d*) >= 0.15) on the evaluation or potency dimension sorted by HpE and HpP.

Quadrant	*N*	*Freq.*
HpE and HpP confirmed	54	.46
HpE and HpP disconfirmed	15	.13
HpE disconfirmed and HpP confirmed	12	.10
HpE confirmed and HpP disconfirmed	37	.31

Note: n = 118 out of 192 social categories with abs(d) on the evaluation dimension or abs(d) on the potency dimension larger than 0.15. N = Number of cases; Freq. = relative frequency.

Figures

	Masculine category lower on Evaluation compared to the feminine category	Masculine category higher on Evaluation compared to the feminine category
Masculine category higher on Potency compared to the feminine category	*HpE and HpP confirmed*	*HpE disconfirmed and HpP confirmed*
Masculine category lower on Potency compared to the feminine category	*HpE confirmed and HpP disconfirmed*	*HpE and HpP disconfirmed*

Figure 1: Stereotypes matrix.

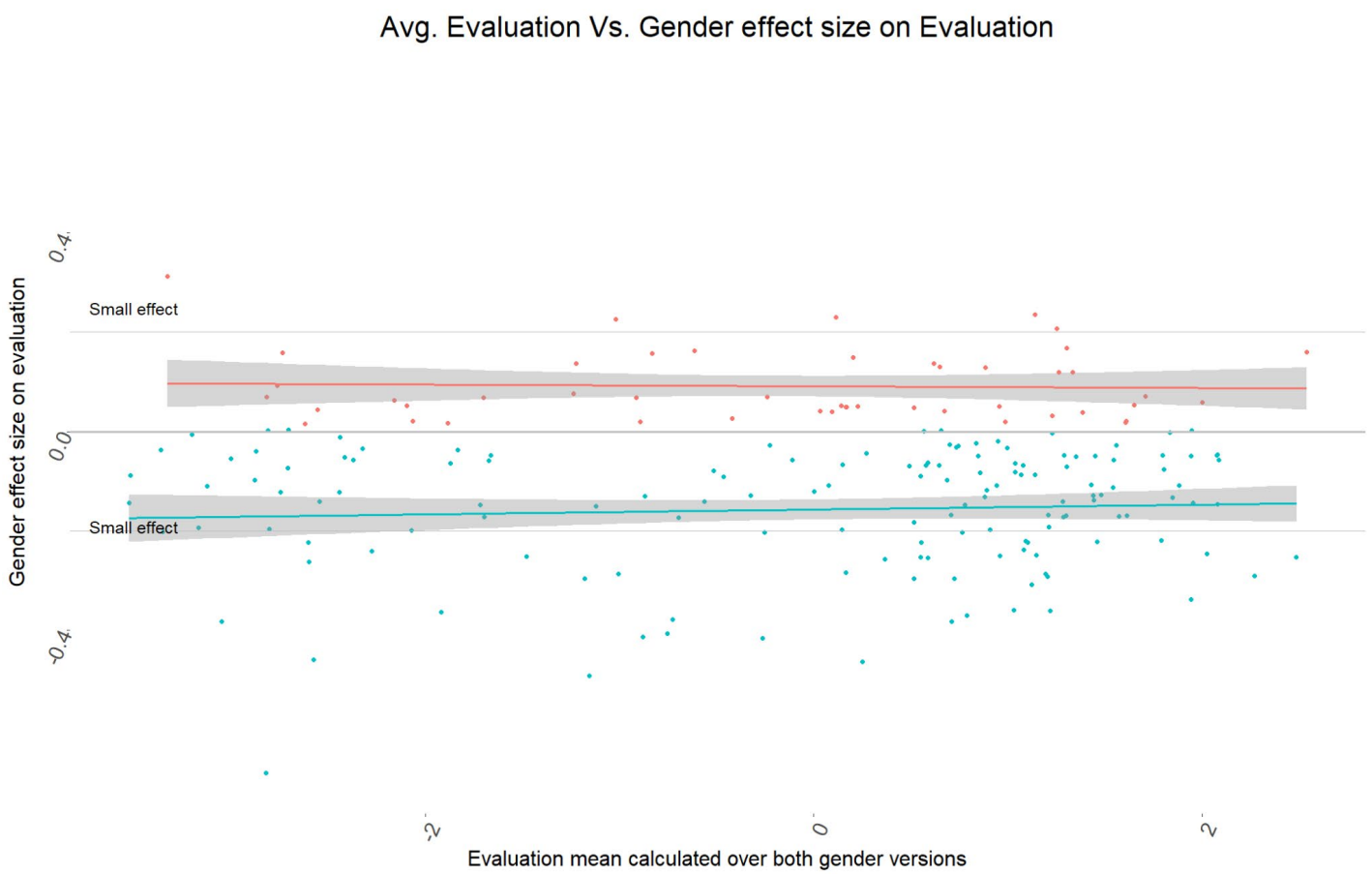


Figure 2.1: Scatter plots of the average evaluation calculated using the ratings for the masculine and feminine version (*x*-axis) and gender effect size on the evaluation dimension (*y*-axis). The red regression line represents the social categories whose masculine version is rated higher; the blue line represents the social categories whose feminine version is rated higher. The gray lines mark the threshold for small effects defined by Lovakov and Agadullina (2021).

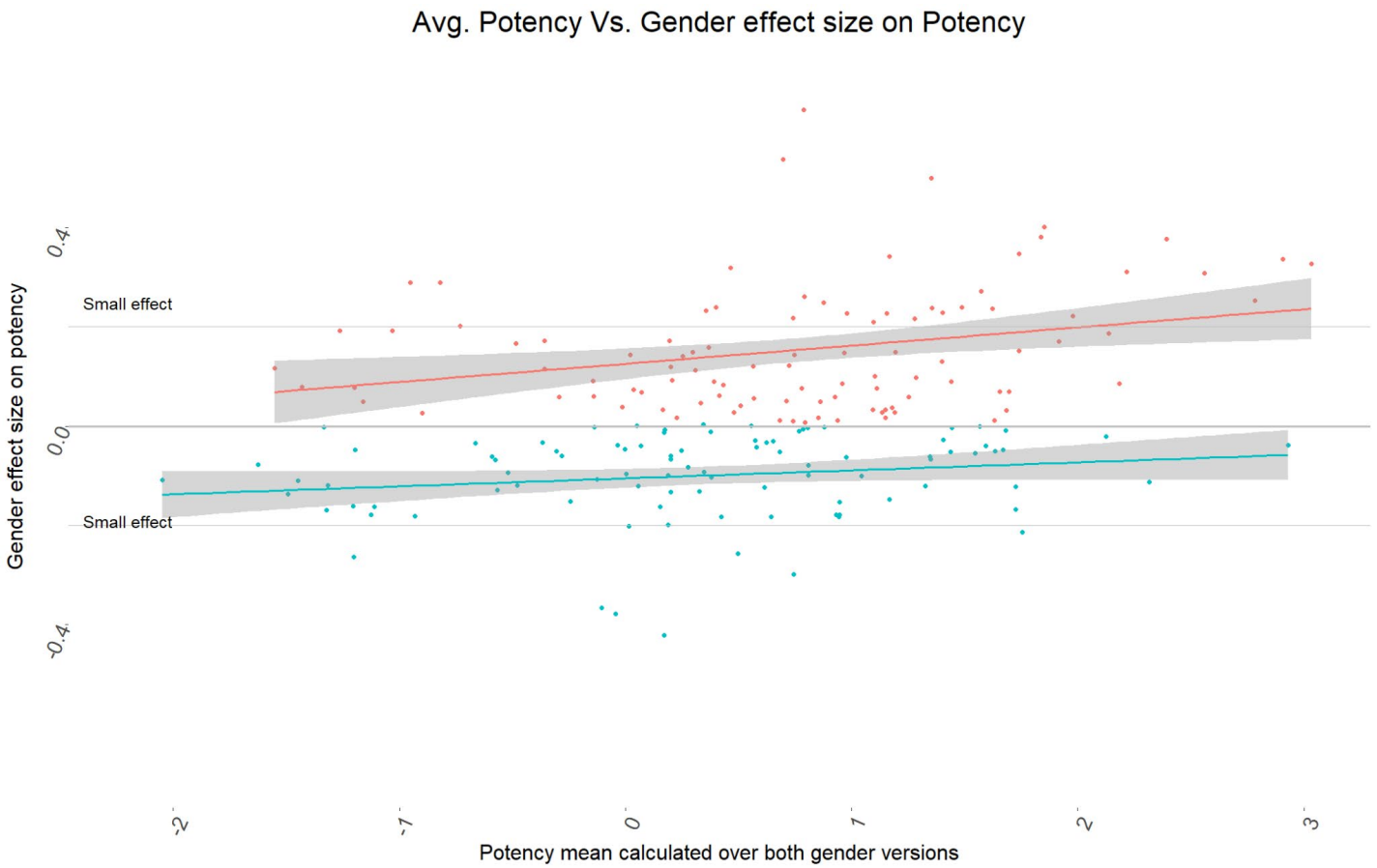


Figure 2.2: Scatter plots of the average potency calculated using the ratings for the masculine and feminine version (*x*-axis) and gender effect size on the potency dimension (*y*-axis). The red regression line represents the social categories whose masculine version is rated higher; the blue line represents the social categories whose feminine version is rated higher. The gray lines mark the threshold for small effects defined by Lovakov and Agadullina (2021).

Figure 2.2: Scatter plots of the average potency calculated using the ratings for the masculine and feminine version (*x*-axis) and gender effect size on the potency dimension (*y*-axis). The red regression line represents the social categories whose masculine version is rated higher; the blue line represents the social categories whose feminine version is rated higher. The gray lines mark the threshold for small effects defined by Lovakov and Agadullina (2021).

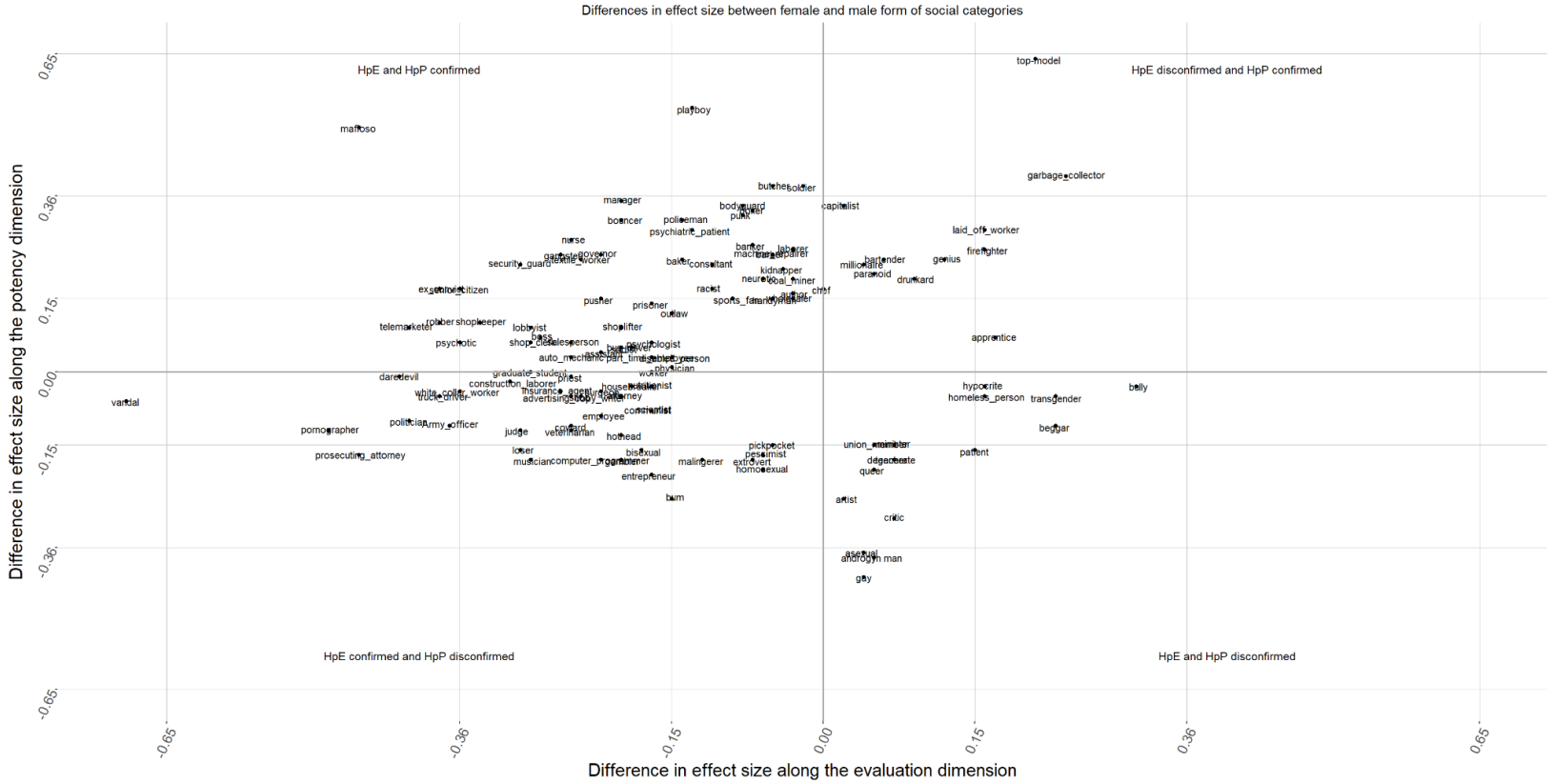


Figure 3: Social categories plotted according to the effect size of gender along the evaluation (*x*-axis) and potency (*y*-axis) dimension. For instance, the social category mafioso is located in the upper left quadrant because the feminine version is rated as more positive (negative effect of gender on evaluation; HpE confirmed), as well as more potent (positive effect of gender on potency; HpP confirmed) than the corresponding masculine version. Effect sizes are interpreted following the meta-analysis conducted by Lovakov and Agadullina (2021).

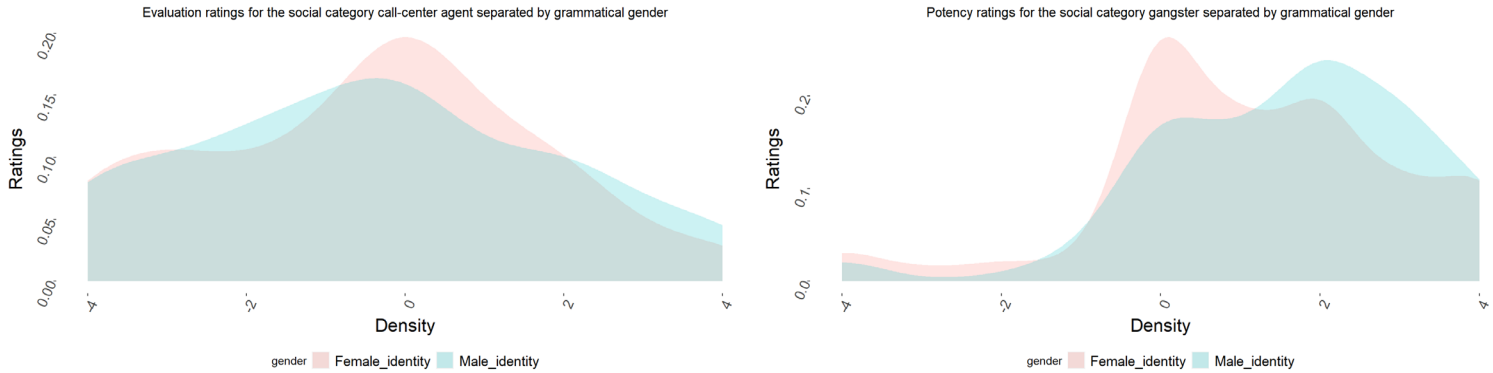


Figure 4: On the left, the density plot of the ratings for a "call center agent". On the right, the density plot of the ratings for the potency of a "gangster". In both cases, the ratings for the female version are colored in red and the ratings for the male version are colored in blue. The social category "call center agent" represents the case where there are no differences in effect size and the distribution is *t* shaped. The social category "gangster" represents an extreme case where there are significant differences in effect size and the distribution is bimodal.

Supplementary information to:

The affective roots of gender stereotypes

Does grammatical gender influence the affective meanings attached to social categories in French and German?

Analysis by gender and language of respondents

Researchers have demonstrated that gender stereotypes are extensively shared across language cultures and – at least partially – between genders (Glick et al., 2000). As a preliminary step in our research project, we wanted to assess whether gender and language of the participant have a moderating effect on the affective perception of gendered social categories. The following analysis is structured as follows: firstly, we performed regression analysis to investigate whether the gender and language culture of the participant interacts with the gender of the social categories. Secondly, we compared ratings provided by female and male participants and used Cohen's *d* effect size to assess the magnitude of difference as well as to identify social categories that are consistently judged differently between male and female participants.

Due to the balanced incomplete block design typical for EPA surveys (Heise, 2007), we computed mixed effects regression models – one for each EPA dimension – with fixed

effects for gender (*GndPart*) and language culture (*LCPart*) of the participant as well as for the gender of the rated social category (*GndSocCat*) which was contrast coded. The models include crossed random slopes for participants and social categories. Three interaction terms were also included. First, the effect of perceiver's gender on affective perception may vary across cultures (*GndSocCat* * *LCPart*). Second, the gender of the social category may activate ingroup and outgroup biases (*GndSocCat* * *GndPart*) or – third – culture-specific affective patterns (*GndSocCat* * *LCPart*). Table S1 and Figure S1 report the results of the three models, which show that the gender of the participants affects the perception of social category on the EPA dimensions. Given the negative sign of the coefficients, male participants perceive social categories as more neutral along all three EPA dimensions than female participants (for evaluation: b = -.10, $t(192)$ = -2.04, $p < .05$, 95% *CI* [-.20; .00]; for potency: b = -.13, $t(192)$ = -2.81, $p < .05$, 95% *CI* [-.22; -.04]; for activity: b = -.14, $t(192)$ = -3.04, $p < .05$, 95% *CI* [-.23; -.05]). Participants' country also has a significant effect on affective perceptions of evaluation (b = .13, $t(192)$ = -2.78, $p < .01$, 95% *CI* [.04; .22]) and potency (b = .15, $t(192)$ = 3.46, $p < .01$, 95% *CI* [.07; .24]), but not on perceptions of activity. Thus, German participants tend to evaluate social categories as more positive and potent. The gender of the social category does not influence affective perceptions, neither as a fixed term nor as a term that interacts with the gender of the participant. Only the interaction between the gender and the language culture of the participants is significant suggesting that German men tend to rate social categories as more neutral on the potency (b = -.15, $t(192)$ = -2.43, $p < .05$, 95% *CI* [-.28; -.03]) and activity (b = -.14, $t(192)$ = -2.17, $p < .05$, 95% *CI* [-.27; -.01]) dimensions. Hence, while

German and France as well as male and female participants do rate social categories differently, such differences do not significantly interact with the gender of the social category and are consequently irrelevant for the main purpose of our research.

Figure S1: Differences in affective meanings for evaluation, potency, and activity by gender of participants (GndPart), language culture of participants (LCPart), and gender of the social category rated (GndSocCat). Horizontal Bars correspond to interaction-term coefficients of the corresponding mixed-effects models (Table S1). Number of participants = 1.400; number of social categories = 192. Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05.

As a next step, we calculated Cohen's *d* by comparing the ratings given by female and male respondents for the same social category. A negative effect size indicates higher ratings by female participants, while a positive effect size indicates the opposite. In this way, we aim to identify social categories that elicit greater gender differences in affective perceptions. Given the significant effect of language culture on ratings and potency, we computed Cohen's *d* separately by country and compared the results to detect cross-national differences and consistencies. As Table S2 shows, in most cases the gender of the participant has small or no effect on the affective perception. Moreover, female participants tend to rate social categories as more positive, potent, and active than male participants. This pattern is consistent in both countries.

Lastly, we were led by the hypothesis that differences in affective perception between men and women are mainly related to specific social categories that make gender salient, i.e. in the case of sexual social categories (Smith-Lovin 1987; Heise, 2007; Ridgeway, 2011). In this regard, we find anecdotal support for this hypothesis, especially with regard to the evaluation dimension. Considering the French data, the top five social categories by effect size are (masculine) worker, d = -1.02; (feminine) sadist, d = .77; (feminine) playboy, d = .77; (masculine) pornographer, d = .76; (masculine) homosexual, d = .76. A similar pattern can be observed for Germany: (masculine) welfare recipient, d = -.81; (masculine) queer, d = -.73; (masculine) transgender, d = -.71; (feminine) queer, d = -.69; (feminine) laborer, d = -.69. For both countries, at least three out of the top five social categories eliciting differences in affective perception between genders are either related to sexuality (pornographer, playboy, sadist) or are sexual social categories (queer, transgender, homosexual).

In summary, we propose the following answer to the question of whether the gender of the perceiver interacts with the affective perception of gendered social categories. Our results show that the effect of the perceiver's gender on the affective meanings attributed to gendered

social categories tends to be null or small, roughly corresponding to the effect size reported in Hyde's (2005) meta-analysis. Despite the consistent similarities, there are also patterns of dissimilarity that persist in both countries. First, we found some anecdotal evidence that when gender becomes semantically salient – as in the case of sexuality or sexual social categories – the effect size becomes larger. Second, the gender of the perceiver leads to greater differences in affective meaning produced by men and women when social categories are evaluated far from the neutral anchor (MacKinnon & Keatings; 1989; Chapman et al., 2022). Nevertheless, and most importantly for our research project, the gender of the perceiver does not interact with the gender of the social category, neither for German nor for French participants. Following this finding, we developed our analysis merging ratings provided by male and female as well as French and German participants.

Distributions of effect sizes by dimension

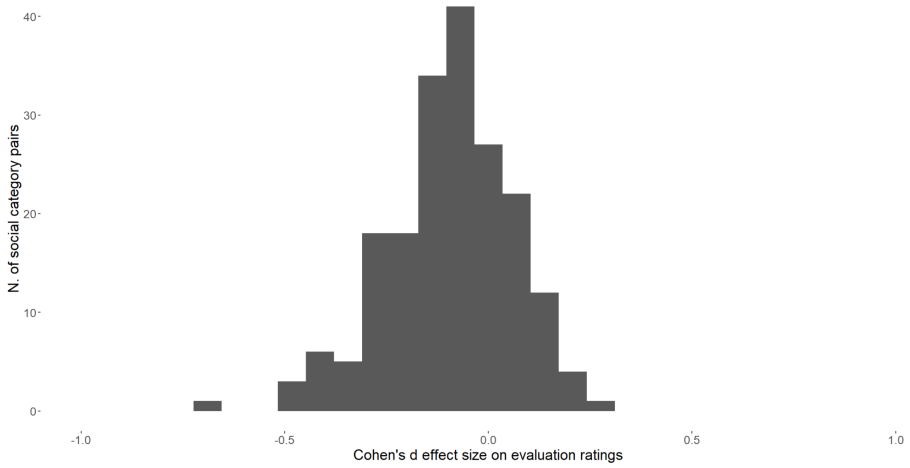


Figure S2: Distribution of Cohen's d effect sizes of gender on the evaluation dimension.

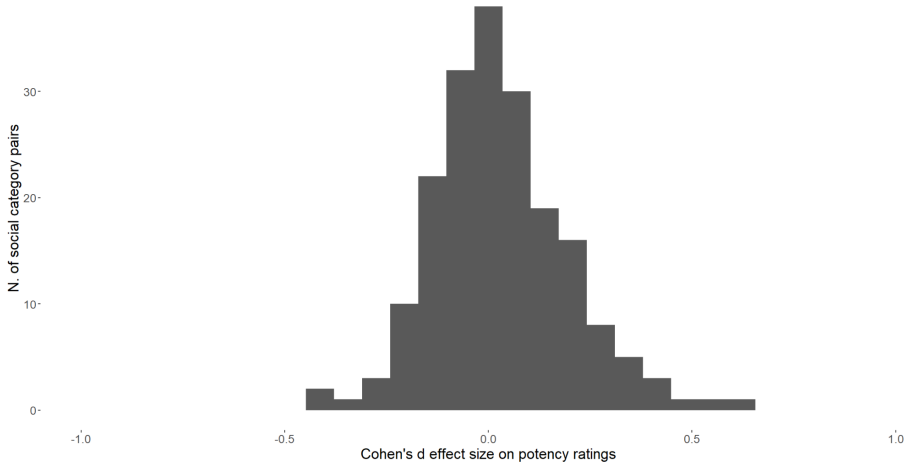


Figure S3: Distribution of Cohen's d effect sizes of gender on the potency dimension.

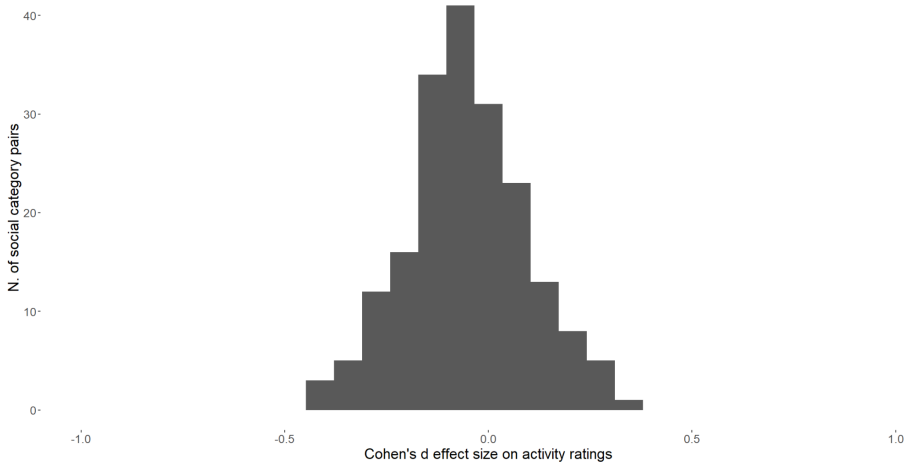


Figure S4: Distribution of Cohen's d effect sizes of gender on the activity dimension.

Statistical significance analysis

In the following, we integrate the effect size based analysis presented in the main article with an analysis of statistical significance. To this purpose, we performed for each EPA dimension 192 two-sample unpaired Student *t*-tests assessing whether the null hypothesis that the two means are equal should be rejected. We considered two levels of significance: the canonical 0.05 level and a very conservative alpha value (.0003) obtained after a Bonferroni correction. Moreover, we also considered absolute differences between the means.

Evaluation. We advanced the hypothesis (HpE) that the more positive evaluation ascribed to women will be systematically reflected in a more positive evaluation ascribed to the feminine versions of social categories. As shown in Table S3, when looking at absolute differences, feminine social categories are consistently rated more positively than their masculine counterparts (141 of 192 social categories, 73%). However, few differences are significant at the 0.05 level (42 of 192, 21%) and almost none after Bonferroni correction (5 of 192, 3%). When pairs of social categories are statistically different, feminine versions are rated as more positive in 38 out of 42 cases. Although there are some small differences, this pattern of results also holds when only the masculine and feminine subsamples of respondents are considered. Overall, these results confirm the effect size based analysis: few pairs of social categories are statistically different. Nonetheless, when differences are statistically significant the feminine versions of a social category are mostly evaluated more positively.

Potency. Following our hypothesis (HpP), we expected that the masculine version of a social category would be rated as more potent than the same social category interpreted by a woman. Our data (Table S4) show that 103 out of 192 (54%) masculine social categories are rated as more potent than the corresponding feminine version. This number is only slightly higher than what would be expected by chance. In terms of ratings, only a few differences are significant at the 0.05 level (34 out of 192, 18%) and almost none after Bonferroni correction (4 out of 192, 2%). However, when statistically different pairs of social categories are considered, masculine versions are rated higher in 29 out of 34 cases. Thus, the second hypothesis appears to be only marginally supported: feminine versions of a social category are not systematically rated as less potent, and the differences are mostly not large

enough to reach statistical significance. However, when looking at statistically significant differences, there is a slight tendency for masculine social categories to be rated as more powerful.

Activity. We hypothesized (HpA) that social categories presented in the feminine form will be perceived as less active than the corresponding masculine social category. According to Table S5, feminine social categories are instead considered to be more active than their masculine counterparts in 129 out of 192 cases. Again, only a few differences are significant at the 0.05 level (30 out of 192, 16%) and none after Bonferroni correction. Nevertheless, when statistically different pairs of social categories are considered, feminine versions are rated higher on the activity dimension in about 70% of cases (21 of 30, 70%). Thus, because in most cases women interpreting social categories have been rated as more active, our third hypothesis is not supported.

Overall, the analysis of the differences in mean ratings between the masculine and feminine versions of the social categories in our dictionary confirms the results of the analysis based on effect sizes. The absolute differences confirm that feminine social categories are mostly judged as more positive (HpE), in slightly more cases as less potent (HpP), and – contrary to our hypothesis (HpA) – mostly as more active than their masculine counterparts. However, in most cases the differences are not large enough to reach statistical significance. Thus, the gender of social categories leads to statistically significant differences in affective meanings only in relatively few cases.

Power analysis

The analysis presented in the main article is based on EPA ratings collected from n = 700 German and n = 700 French participants. Since we organized the social categories into 10 sets of words that are identical between French and German, each social category was rated by n = 70 German and n = 70 French participants. Therefore, the average ratings presented in the study are based on n = 140 French and German participants. Since we compared the ratings given in French and German for the masculine (n = 140) with the ratings given for the feminine version of the social categories (n = 140), we have a total sample size of n = 280 participants organized in two groups. Under these conditions, we have statistical power of .80 to detect an effect size of d = .25 at a highly reliable significance level of .001. A Cohen's d effect size of .25 is above the threshold for small and below the one for moderate effects in social psychology derived empirically by Lovakov and Agadullina (2021).

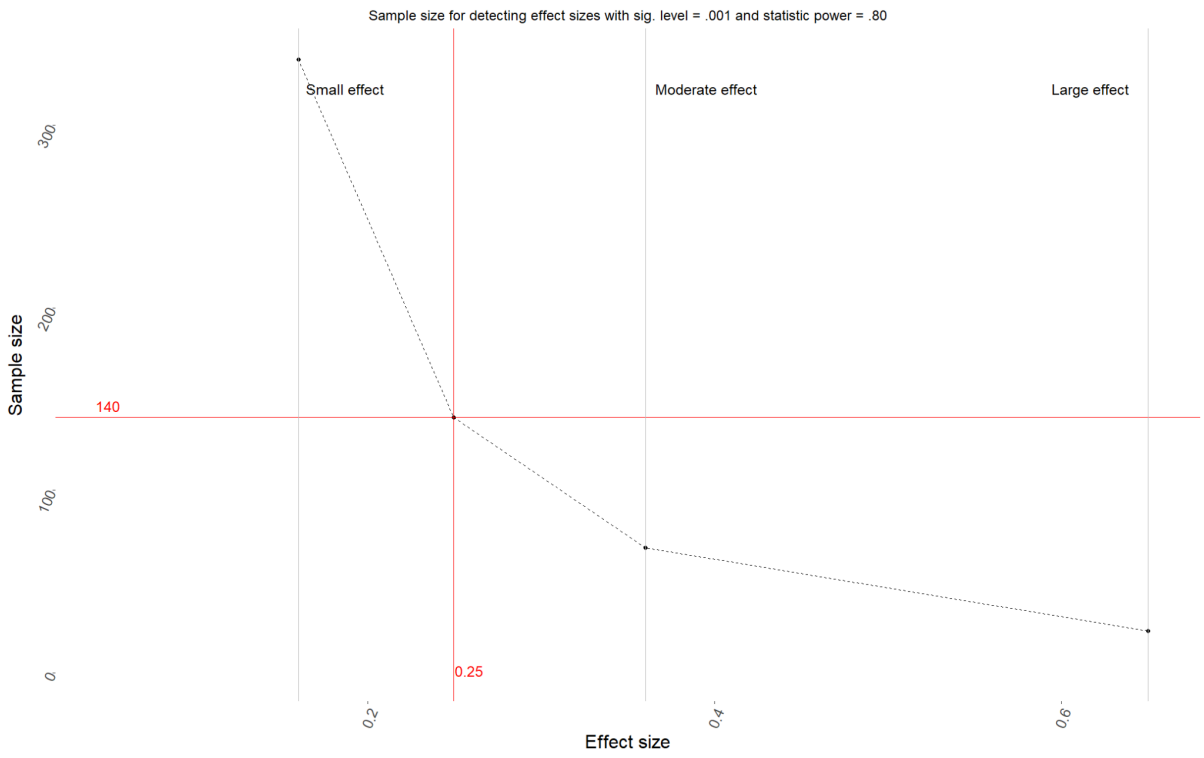


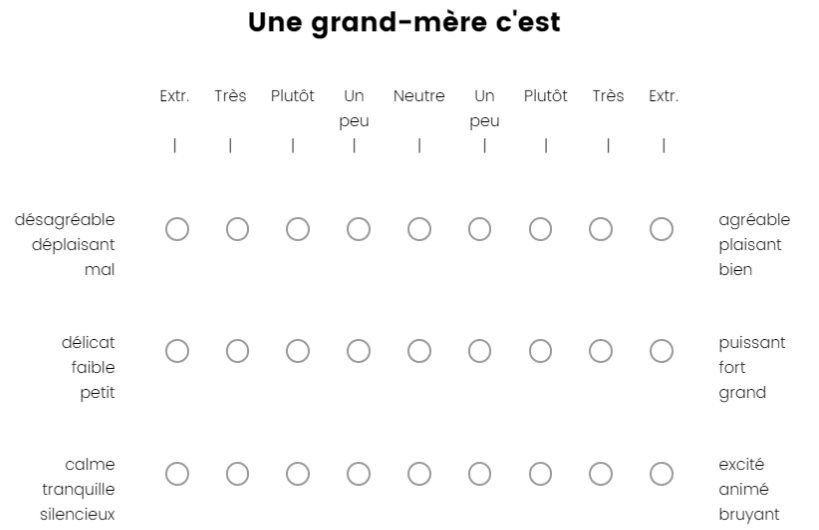
Figure S4: Effect size (*x*-axis) and sample size (*y*-axis). The threshold for small, moderate and large effects are defined following Lovakov & Agadullina (2021). The red line intersecting the *y*-axis indicates the sample size of this study. The red line intersecting the *x*-axis indicates the resulting effect size that we can detect.

Instructions - France

Voici un exemple avec le terme "une grand-mère".

Il vous est demandé d'évaluer ce terme sur 3 dimensions :

- désagréable/déplaisant/mal **vs.** agréable/plaisant/bien
- délicat/faible/petit **vs.** puissant/fort/grand
- calme/tranquille/silencieux **vs**. excité/animé/bruyant



Pour donner votre réponse sur ces 3 dimensions, vous allez pouvoir utiliser toutes les nuances de l'échelle proposée :

- Extrêmement (abrégé Extr.)
- Très
- Plutôt
- Un peu
- Neutre

Nous vous demandons de répondre spontanément, sans trop réfléchir.

Pour voir la suite des explications, cliquez sur la flèche en bas de l'écran.

Instructions - Germany

Hier ist ein Beispiel mit dem Begriff "eine Großmutter".

Wir bitten Sie, dieses Wort auf den folgenden drei Dimensionen zu bewerten:

- "unangenehm/unfreundlich/schlecht" **vs.** "angenehm/freundlich/gut",
- "zart/schwach/klein" **vs.** "kraftvoll/stark/groß" und
- "ruhig/träge/still" **vs.** "bewegt/lebhaft/geräuschvoll".

Für Ihre Antwort können Sie alle Abstufungen der vorgeschlagenen Skala verwenden:

- "äußerst",
- "sehr",
- "ziemlich",
- "etwas" oder
- "neutral".

Wir bitten Sie, spontan zu antworten, ohne zu viel nachzudenken.

Bitte beachten Sie, dass sich die Richtung der Skalen ändern kann!

Klicken Sie auf "Weiter", um mit der Anleitung fortzufahren.

Rating instrument - France

Un mécanicien automobile c'est

	Extr.	Très	Plutôt	Un peu	Neutre	Un peu	Plutôt	Très	Extr.	
agréable plaisant bien	○	○	○	○	○	○	○	○	○	désagréable déplaisant mal
puissant fort grand	○	○	○	○	○	○	○	○	○	délicat faible petit
excité animé bruyant	○	○	○	○	○	○	○	○	○	calme tranquille silencieux

Rating instrument - Germany

eine Kapitalistin ist

	Äußerst	Sehr	Ziemlich	Etwas	Neutral	Etwas	Ziemlich	Sehr	Äußerst	
unangenehm unfreundlich schlecht	○	○	○	○	○	○	○	○	○	angenehm freundlich gut
zart schwach klein	○	○	○	○	○	○	○	○	○	kraftvoll stark groß
langsam träge still	○	○	○	○	○	○	○	○	○	schnell lebhaft geräuschvol

References

Chapman, Robert M., Margaret N. Gardner, and Megan Lyons. 2022. Gender differences in emotional connotative meaning of words measured by Osgood's semantic differential techniques in young adults. *Humanities and Social Sciences Communications*, 9(1), 1-11. doi: 10.1057/s41599-022-01126-3.

Glick, Peter, Susan T. Fiske, Antonio Mladinic, Jose L. Saiz, Dominic Abrams, Barbara Masser, Bolanle Adetoun, Johnstone E. Osagie, Adebowale Akande, Amos Alao, Annetje Brunner, Tineke M. Willemsen, Kettie Chipeta, Benoit Dardenne, Ap Dijksterhuis, Daniel Wigboldus, Thomas Eckes, Iris Six-Materna, Francisca Exp6sito, Miguel Moya, Margaret Foddy, Hyun-Jeong Kim, Maria Lameiras, Maria Jose Sotelo, Angelica Mucchi-Faina, Myrna Romani, Nuray Sakalh, Bola Udegbe, Mariko Yamamoto, Miyoko Ui, Maria Cristina Ferreira, and Wilson Lopez Lopez. 2000. Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology*, 79(5), 763–775. doi: 10.1037/0022-3514.79.5.763.

Heise, David R. 2007. *Expressive Order: Confirming Sentiments in Social Actions*. Boston, MA: Springer.

Hyde, Janet S.2005. The gender similarities hypothesis. *American psychologist*, 60(6), 581. doi: 10.1037/0003-066X.60.6.581.

Lovakov, Andrey, and Elena R. Agadullina. 2021. Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, 51(3), 485-504. doi: 10.1002/ejsp.2752.

MacKinnon, Neil J., and Leo Keating. 1989. The Structure of Emotions: Canada - United States Comparisons. *Social Psychology Quarterly* 52(1):70–83. doi:10. 2307/2786905.

Ridgeway, Cecilia L. 2011. *Framed by gender: How gender inequality persists in the modern world*. New York: Oxford University Press.

Smith-Lovin, Lynn. 1987. Impressions from events. *Journal of mathematical sociology*, 13(1,2), 35-70. doi: 10.1080/0022250X.1987.9990026.

.

Tables

Table S1: Mixed-effects regressions coefficients for social categories' EPA ratings

	Model 1: Evaluation					*Model 2: Potency*					*Model 3: Activity*				
Effects	*Est*	*SE*	*t*	*95% CI*		*Est*	*SE*	*t*	*95% CI*		*Est*	*SE*	*t*	*95% CI*	
				LL	*UL*				*LL*	*UL*				*LL*	*UL*
(Intercept)	.05	.12	-.40	-.29	.19	.58	.08	7.50	.43	.74	.78	.07	10.47	.63	.93
GndPart *Male*	-.10*	.05	-2.04	-.20	.00	-.13*	.05	-2.81	-.22	-.04	-.14**	.05	-3.04	-.23	-.05
LCPart *Germany*	.13*	.05	2.78	.02	.21	.15**	.05	3.46	.07	.24	.01	.05	.32	-.07	.10
GndSocCat *Masculine social category*	-.18***	.02	-7.58	-.22	.-13	.04	.02	1.55	-.01	.08	-.08**	.02	-3.23	-.12	.03
GndSocCat * LCPart *Masculine social category * Germany*	.03	.03	1.24	-.02	.09	.05	.03	1.88	.00	.10	-.03	.03	-.91	-.08	.03
GndPart * LCPart *Male * Germany*	-.06	.07	-.90	-.19	.07	-.15*	.06	-2.43	-.28	-.03	-.14*	.06	-2.17	-.27	-.01
GndSocCat * GndPart *Masculine social category * Male*	.00	.03	.04	-.05	.05	-.02	.03	-.80	-.07	.03	-.01	.03	-.28	-.06	.05

Participants: 1.400; Social categories: 192. Note: Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05. Est. = Estimate; SE = standard error; t = *t*-statistic; CI = confidence interval; LL = lower limit; UL = upper limit.

Table S2: Cohen's *d* effect size (1988) of gender of perceiver by dimension, direction and country for n = 384 gendered social categories

	France						*Germany*					
	Evaluation		*Potency*		*Activity*		*Evaluation*		*Potency*		*Activity*	
Effect size	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*
no effect	80	77	84	69	82	86	89	68	76	46	91	42
small	76	61	101	53	88	57	98	50	117	37	107	39
medium	60	18	52	15	51	12	50	17	87	11	84	11
large	7	5	6	4	7	1	11	0	10	0	9	1

Note: F → Female participants rated the social category higher on given dimension; M → Male participants rated the social category higher on given dimension. Effect size interpretation is defined following the empirical guidelines provided by Lovakov and Agadullina (2021).

Table S3: Differences in evaluation ratings of pairs (m/f) of social categories

Level of analysis	*n*	*Freq*
Feminine social categories are rated more positively	141	.73
Significant Δ at 0.05 level	42	.22
thereof feminine social categories are rated more positively	38	.90
Significant Δ after Bonferroni correction	5	.03
thereof feminine social categories are rated more positively	5	1.0

Note: n = 192 social categories. N = Number of cases; Freq. = relative frequency.

Table S4: Differences in potency ratings of pairs (m/f) of social categories

Level of analysis	*n*	*Freq*
Masculine social categories are rated as more powerful	103	.54
Significant Δ at 0.05 level	34	.18
thereof masculine social categories are rated as more powerful	29	.85
Significant Δ after Bonferroni correction	4	.02
thereof masculine social categories are rated as more powerful	3	.75

Note: n = 192 social categories. N = Number of cases; Freq. = relative frequency.

Table S5: Differences in activity ratings of pairs (m/f) of social categories

Level of analysis	*n*	*Freq*
Feminine social categories are rated as more powerful	129	.67
Significant Δ at 0.05 level	30	.16
thereof feminine social categories are rated as more active	21	.70
Significant Δ after Bonferroni correction	0	0
thereof feminine social categories are rated as more active	0	0

Note: n = 192 social categories. N = Number of cases; Freq. = relative frequency.