

Au-pairs are rarely male:

Norms on the gender perception of role names across English, French and German

This study aims to evaluate people's gender representation of role names across three different languages, English, French and German. In order to provide norms for role names to be further used in research on gender stereotyping, 362 participants filled in a questionnaire about their estimates of the proportion of men and women in different roles. Role names are any names that incorporate features used to describe a person or a group of people, such as hobbies (e.g. soccer fan) or occupations (e.g. dentists, actors or students). As role names are particularly prone to gender stereotypes (Baudino, 2001), they are often used to investigate, for example, the influence of stereotypes on reading comprehension (e.g., Duffy & Keir, 2004; Garnham, Oakhill & Reynolds, 2002; Kennison & Trofe, 2003; Oakhill, Garnham & Reynolds, 2005; Sturt, 2003). Gender perceptions of professional occupations were previously investigated by Shinar (1975), Beggs and Doolittle (1993) and Couch and Sigler (2001) for US samples and by Sastre, Fouquereau, Igier, Salvatore and Mullet (2000) for Spanish and French samples. Only recently Kennison and Trofe (2003) provided norms for 405 English role names.

The present study is the first, to our knowledge, to simultaneously obtain gender norm ratings across languages that differ in the gender markedness of nouns. This issue may need to be taken into consideration when measuring norms. In German as well as in French, the gender of a character in a text is - in general - explicitly given by the form of the determiner and/or by the morphologically feminine or masculine form of the noun. In contrast to English, in which readers of the sentence "The neighbour could not stand the noise" cannot be sure whether it is about a man or a woman, in French or German, the sentence "La voisine ne pouvait pas tolérer le bruit" or "Die Nachbarin konnte den Lärm nicht ertragen"

unequivocally signifies that a woman is referred to. Although there are gender-marked role names in English (such as *watchman* or *midwife*), most role names are not gender marked. This lack of markedness results in readers relying on context information such as stereotypes to infer the gender of the person that is referred to (Garnham, Oakhill & Reynolds, 2002).

In French and German, the grammatical form is not always sufficient to determine the protagonists' gender as there is an order in the use of masculine and feminine forms: To refer to several persons of both sexes and to persons of unknown sex or where sex is irrelevant, the masculine form is used. This usage is called the generic-masculine (GM). As masculine words that serve as generics also have a male-specific meaning, readers and listeners have to use context information (provided by the text or derived from their own knowledge) to correctly identify whether a word is used in a specific or generic way. There is an ongoing discussion about whether the use of the GM leads to a "gender-open" representation, as claimed by conventional linguistics, or to a heightened cognitive accessibility of male prototypes, thus impeding the accessibility of female prototypes, as claimed by proponents of feminist linguistics and as supported by most empirical research on the use of the GM in German (for a review see: Stahlberg, Braun, Irmen & Sczesny, 2007) and in French (e.g., Colé & Segui, 1994).

In other words, English readers might more strongly rely on context information to determine the sex of (heard of or read about) protagonists than French or German readers, as in the last two languages gender is grammatically encoded. However, because the masculine form of gender-marked languages can be used in a specific or generic way, there can be a gender ambiguity when that form is used. When measuring norms in French or in German, the question arises whether the form in which the role names are presented influences participants' ratings. If the masculine in French and German is interpreted in a generic way, using either the masculine and feminine forms or only the masculine form, should not alter

the ratings. But if the use of the masculine impedes the accessibility of female prototypes, stereotypicality ratings should be more male biased when only masculine terms are used than when both masculine and feminine terms are presented. Qualifying the English nouns with gender-specific adjectives, i.e. explicitly marking the ends of the scales “male office workers” vs. “female office workers” might lower the influence of the stereotypicality information, thus leading to less extreme ratings compared to the non-qualified nouns.

Other factors, such as *sex of participants* and *scale direction*, have been identified as influencing such ratings. More specifically, women tend to perceive role names as being more gender neutral than do men (Beggs & Doolittle, 1993; Kennison & Trofe, 2003). Furthermore, Kennison and Trofe (2003) report male participants to be especially prone to changes in scale direction: Ratings were more male biased when the scale was directed towards “mostly male” (1 = mostly female, 7 = mostly male) than when it was directed towards “mostly female” (1 = mostly male, 7 = mostly female), and this effect was larger for male participants. Research on rating scales has shown that response scale formats with a *counterintuitive* combination of content and number (e.g. 1 = strongly agree to 8 = strongly disagree) can lead to more diversified answers resulting in less extreme means (Rammstedt & Krebs, 2007). Thus, the findings of Kennison and Trofe (2003) could be interpreted as a hint that a scale that is directed towards “mostly male” is the more intuitive one. To further investigate this issue in our research we varied the structure of our questionnaires with reference to *response scale format* and *stimulus material*. Two studies were conducted. In both studies, we presented participants with a list of role names for which they had to indicate the proportion of men and women in these roles on a rating scale. Within each study the scale direction was manipulated and between the studies different forms of the role names were employed. In *Study 1*, for each role name, both the feminine and the masculine forms were presented in the French and German samples (e.g.: docteurs/docteuses; Ärzte/Ärztinnen).

In the English sample, the role names were qualified by a gender-specific modifier (e.g.: male doctors/female doctors). In *Study 2* we used the masculine role name form, as generics, only. In this way, we explored whether the scale direction effect reported by Kennison and Trofe (2003) was replicable (English sample in *Study 2*), could be generalized across languages (French and German sample in *Study 2*) and whether it would be amplified by the gender-marking of the role names (*Study 1*).

To summarize, the purpose of our study is as follows:

First, we aim to provide norms for the gender stereotypicality of role names in three languages (English, French, German). We are interested in the consensus about the stereotypicality of each role name within each language as well as across the three languages.



Second, we investigated the stability of stereotypicality ratings by manipulating scale direction and by using different forms of the role names. With this, we not only seek to provide further evidence for the stability or fragility of the norms but to also examine two hypotheses derived from previous research: (1) If the scale direction effect reported by Kennison and Trofe (2003) reflects a general bias then the influence of the response format should hold across the variation of the stimulus material independent of language changes. (2) If the use of the masculine intended as generic impedes the accessibility of female examples in the French and German samples the ratings in *Study 2* should be more male biased compared to *Study 1* and, presumably, compared to the English sample.

The present work extends existing research by comparing gender-marked (French, German) and non gender-marked (English) languages and by systematically varying the grammatical form of the role names. In doing so, we wish to demonstrate that providing both masculine and feminine forms in French and German may influence the gender representation associated with a role name. By additionally manipulating the response scale format we seek to mark out the importance of the previously found scale direction effect.

Study 1

Method

Questionnaire and Design. Based on norms already published for English (Kennison & Trofe, 2003) 126 role names were selected that had clear cognates in all three languages (see Appendix A). Four of the role names are or can be gender marked in English (actors/actresses; craftsmen/craftswomen; postmen/postwomen; waiters/waitresses), whereas 34 are not marked in French (e.g., auteurs - authors) and 13 are not marked in German (e.g., Jugendliche - adolescents). These proportions reflect the typical proportions of the languages in question.

The participants' task was to decide on the proportion of males and females in that role on an 11-point rating scale as depicted in Figure 1. The rating scale's orientation was manipulated such that in one version (*100% male-left*) the 100/0 on the left side was graphically labeled as 100 % male () and the 0/100 on the right was labeled as 100 % female () whereas in the second version (*100% male-right*) the labeling was reversed.

Participants were instructed to estimate to what extent the groups are *actually* made up of women or men and not to be influenced by any thoughts about how proportions should be.

The questionnaires were administered in three languages: *English*, *French* and *German*. The female and male versions of the role names were given in accordance with the scale's orientation at the right and left side of the scale respectively (cf. Figure 1). For gender-marked role names (mostly in the French and the German versions) the *male* (specific masculine) as well as the *female* role names (specific feminine) were presented. For gender-unmarked role names (mostly in the English version) the words "male" or "female" (*hommes*

– *femmes* in the French and *männliche* – *weibliche* in the German version) were added to the role names (e.g., male tennis players vs. female tennis players). The order in which the role names were presented to individual participants was varied: In total, there were 24 different random orders of the 126 role names (8 for each language).

Sample and Procedure. A total of 229 students participated in the study. Thirty-five participants had to be excluded because they were not native speakers of the target languages ($N = 25$), because they did not provide information on their mother tongue ($N = 6$) or because they were outliers in age based on the overall mean ($M + 4 SD$: $N = 4$). Of the remaining 194 participants 60 (31 male, 29 female) were recruited at the Universities of Fribourg, Geneva and Neuchâtel (French-speaking part of Switzerland), 76 (41 male, 35 female) at the University of Bern (German-speaking part of Switzerland) and 58 (7 male, 51 female) at the University of Sussex (United Kingdom). Their ages ranged from 18 to 35 years ($M = 22.4$, $SD = 3.2$).

Versions of the questionnaire with the two scale directions were quasi-randomly assigned to the male and female participants independently in each of the three languages. In all languages, participants were recruited via seminars and lectures.

Results

Data from questionnaires in the 100% male-left condition were recoded in such a way that higher values reflect a higher proportion of men.

Inter-Participant Analyses. For each participant the mean rating across the role names was calculated ($M = 57.6$, $SD = 3.7$, Scale midpoint: 50). These normally distributed scores, $KS\ z = 1.1$, $p = .19$, ranged from 46.1 to 68.2 and were used as the dependent variable in a three (Language: English vs. French vs. German) by two (Scale Direction: 100% male-left vs. 100% male-right) ANOVA. The analysis revealed a main effect of the Scale Direction, $F(1,188) = 24.4$, $p < .001$, partial $\eta^2 = .12$: the proportion of males was rated to be lower for

the 100% male-right scale ($M = 56.4$, $SD = 3.4$) compared to the 100% male-left scale ($M = 58.9$, $SD = 3.6$). Neither the main effect of Language, $F(2,188) = 1.7$, $p = .19$, partial $\eta^2 = .02$, nor the Scale Direction by Language interaction was significant, $F(2,188) = 2.3$, $p = .11$, partial $\eta^2 = .02$, but post hoc contrasts on the scale direction effect revealed that the scale direction's influence was stronger in English and German (Tukey HSD, both $p < .001$), but comparably weak in the French version (Tukey HSD, $p = .89$, see Figure 2, left panel).

Female participants rated the proportion of males to be only slightly lower ($M = 57.3$, $SD = 3.7$) than male participants ($M = 57.9$, $SD = 3.7$), $t(192) = .99$, $p = .33$. To further investigate the influence of respondent's sex, the ANOVA was repeated with sex as an additional between subjects factor. As there were only seven male participants in the English sample, further analysis was restricted to the French and German versions. There was a Scale Direction effect, $F(1,128) = 10.8$, $p < .001$, but neither the Scale Direction by Language interaction, $F(1,128) = 2.7$, $p = .11$, nor the triple interaction, $F(1,128) = 2.7$, $p = .11$, reached statistical significance (all other $ps > .20$).

Inter-Item Analyses. For each role name the mean rating and the standard deviation were calculated separately for each Language as well as for each scale direction. Across the languages the ratings of the 126 role names were highly reliable, $\rho_I(E-F) = .91$; $\rho_I(E-G) = .87$; $\rho_I(F-G) = .94$, all $N = 126$, all $p < .001$, suggesting a clear consensus across languages.¹ This reliability also held for Scale Direction, $\rho_I(Fr-Fl) = .99$, $N = 126$, $p < .001$.

The Figures 3 to 5 display the scatterplots of mean rating for each item against its standard deviation for each language separately. The figures show that the closest agreement, as reflected in low standard deviations, was found for a group of items judged as neutral (e.g., Swimmers, Spectators, Authors). Standard deviations also decreased towards the male end of the scale, indicating a group of role names that was perceived as strongly male-oriented (e.g., Football coaches, Hunters, Bricklayers). At the female end of the scale fewer role names are

found (e.g. Beauticians, Birth attendants, Secretaries), and participants' ratings are more heterogeneous, at least in English (Figure 3) and French (Figure 4). This suggests that it is easy to agree upon neutral items, i.e. ratings around 50%, but that people vary in their decisions about how far other items, especially female stereotyped items, deviate from neutral. Mean ratings and standard deviations of each role name for each language are given in Appendix B, split by sex of respondents in Appendix C.

Discussion

In the inter-item analysis, the ranking of the items by their gender stereotypicality is highly reliable across languages. Furthermore, clusters of role names perceived as (a) neutrally stereotyped and (b) male stereotyped can clearly be identified. However, there are fewer female stereotyped role names and there is less consensus, except in German, on their degree of stereotypicality.

In the inter-participant analysis, no main effect of language or sex of respondent emerged. However, ratings were influenced by the scale direction. Furthermore, while previous research (Kennison & Trofe, 2003) reported the “mostly women”-presented-on-the-left scale to be associated with ratings tending more strongly towards male, in our study the 100% female-left scale was associated with ratings tending less strongly towards male. The Kennison and Trofe (2003) study used English role names without male or female specification. Therefore, an explanation for the divergent scale effects could be that in our study the scale direction effect is really a primacy effect: In our study scale direction is confounded with the position of the female and the male version of the role names. As we typically read from left to right, the 100% male-left version of the questionnaire leads participants to read the masculine version of the role name first (resulting in higher male ratings), whereas the 100% male-right version leads participants to read the feminine version

of the role names first resulting in higher female ratings (i.e. lower male ratings). If this interpretation is correct, this effect should not occur in Study 2, in which only the male role names appeared, as in the Kennison and Trofe (2003) study.

Study 2

Method

Questionnaire and Design. The questionnaires were identical to those used in Study 1, except that the female role names as well as additional male gender markings were removed. Thus, only the *male* role names (serving as generics) were presented.

Sample and Procedure. A total of 184 students participated in the study. Seven participants had to be excluded because they did not provide information on their mother tongue ($N = 6$) or because they were outliers in age based on the overall mean ($M + 4SD$: $N = 1$). Of the remaining 177 participants, 51 (3 male, 48 female) were recruited at the University of Sussex (United Kingdom), 67 (33 male, 34 female) were recruited at the Universities of Fribourg, Geneva and Neuchâtel (French-speaking part of Switzerland) and 59 (27 male, 32 female) at the University of Bern (German-speaking part of Switzerland). Their ages ranged from 17 to 37 years ($M = 22$, $SD = 3.5$).

Versions of the questionnaire with the two scale directions were quasi-randomly assigned to the male and female participants independently in each of the three languages. Participants were recruited via seminars and lectures.

Results

Data from questionnaires in the 100% male-left condition were recoded in such a way that higher values reflect a higher proportion of men.

Inter-Participant Analyses. For each participant the mean rating across the role names was calculated ($M = 58.8$, $SD = 3.6$, Scale midpoint: 50). These normally distributed scores

(KS $z = 1.0$, $p = .23$) ranged from 51.5 to 68.3 and were the dependent variable in a three (Language: English vs. French vs. German) by two (Scale Direction: 100% male-right vs. 100% male-left) ANOVA. The analysis revealed a main effect of Scale Direction, $F(1,171) = 5.5$, $p = .02$, partial $\eta^2 = .03$: Contrary to the findings of Study 1 but in line with those of Kennison and Trofe (2003) the proportion of males was rated to be higher in the 100% male-right condition ($M = 59.4$, $SD = 3.8$) compared to the 100% male-left condition ($M = 58.2$, $SD = 3.2$). Neither the main effect of Language, $F(2,171) = 2.6$, $p = .07$, partial $\eta^2 = .03$, nor the Scale Direction by Language interaction was significant, $F(1,171) = 1.8$, $p = .18$, partial $\eta^2 = .02$. But as can be seen from Figure 2 (right panel) scale direction had a stronger influence on the ratings in German than in French or English.

Again female participants participants rated the proportion of males to be slightly lower ($M = 58.5$, $SD = 3.3$) than male participants ($M = 59.4$, $SD = 4.0$), $t(175) = -1.64$, $p = .10$. To further investigate the influence of respondent's sex, the ANOVA was repeated with sex as an additional between subjects factor. As there were only three male participants in the English sample, further analysis was restricted to the French and German versions. There was a Scale Direction effect, $F(1,118) = 4.2$, $p = .04$, that was qualified by a three-way interaction ($F(1,118) = 4.0$, $p < .05$; Scale Direction by Language interaction: $F(1,118) = 2.9$, $p = .09$; all other $F < 1$). Female participants' ratings in the German version were most strongly influenced by the scale direction ($M_{F-l} = 61.1$, $SD = 3.6$; $M_{F-r} = 57.4$, $SD = 2.1$), followed by male participants in the German direction ($M_{F-l} = 59.7$, $SD = 3.6$; $M_{F-r} = 58.6$, $SD = 3.3$) and male participants in the French version direction ($M_{F-l} = 60.6$, $SD = 5.1$; $M_{F-r} = 59.1$, $SD = 3.5$), whereas female participants in the French version even showed a reversed pattern ($M_{F-l} = 58.2$, $SD = 3.0$; $M_{F-r} = 59.3$, $SD = 4.1$).

To investigate the effect of the grammatical form of the role names, the data from Study 1 and Study 2 were analyzed in a 3 (Language: English vs. French vs. German) by 2

(Role Name: specific - Study 1 vs. generic - Study 2) by 2 (Scale Direction: 100% male-right vs. 100% male-left) ANOVA. The analysis revealed significant main effects of the version of the role names ($F(1, 359) = 10.2, p = .002$; $M_{\text{specific}} = 57.6, M_{\text{generic}} = 58.8$) and language ($F(2, 359) = 4.2, p = .02, M_E = 57.5, M_F = 58.7, M_G = 58.5$). The main effect of Scale Direction was marginal: $F(1, 359) = 2.8, p = .09$. These effects were qualified by a Role Name by Scale Direction interaction ($F(1, 359) = 26.0, p < .001$) and a three-way interaction ($F(2, 359) = 3.4, p = .03$; all other F s < 1). For all three languages, changing the role names reversed the scale direction effect, which in addition was larger in the specific than the generic version (Role Name x Scale Direction Interaction). However, as shown in Figure 2, the influence of scale direction varied in magnitude across the three languages.² Furthermore, it is noteworthy that no language by version interaction emerged, although the shifts in the ratings due to changing the version of the role names are slightly more pronounced in French ($\Delta = 1.2\%$) and German ($\Delta = 1.4\%$) than in English ($\Delta = 1\%$).

Inter-Item Analyses. As in Study 1 for each role name the mean rating and the standard deviation were calculated separately for each language as well as for each scale direction. Across the languages the ratings of the 126 role names were highly reliable (ρ_I (E-F) = .91; ρ_I (E-G) = .95; ρ_I (F-G) = .95, all $N = 126, p < .001$). This reliability also held for scale direction (ρ_I (Fr-Fl) = .99, $N = 126, p < .001$). The intra-class correlation between the ratings in Study 1 and Study 2 was $\rho_I = .98$ ($N = 126, p < .001$). The mean difference between the ratings in Study 1 and Study 2 was 1.2 % ($SD = 3.3$), with a shift towards a higher proportion of males for role names in Study 2. This male shift was especially pronounced for au pairs (20.2%), flight attendants (8.8%), waiters (8.3%) and butchers (7.6%). Some participants wrote “exclusively male” for au pairs and flight attendants in Study 2, suggesting that they interpreted the masculine versions of those nouns in French (*jeunes hommes au pair*; *stewards*) and German (*Aupair-Jungen*; *Flugbegleiter*) as specific forms.

Mean ratings and standard deviations of each role name for each language are given in Appendix D, split by sex of respondents in Appendix E.

Discussion

Although presenting just the masculine forms of the role names did not change the ranking of the role names, it did induce a small but significant overall increase in the perceived proportion of males. Furthermore, there was a hint, from the fact that some role names, such as au-pairs and flight attendants, were interpreted (at least by some participants) as specifically masculine, that the notion of the generic use of the masculine is not applicable across the board. The interpretation of these role names as exclusively masculine occurred even though they were embedded in a series of role names that were obviously to be generically interpreted.

In the inter-participants analyses no main effect of language or sex of respondents emerged. Furthermore, removing the feminine forms led to a scale direction effect similar to that reported by Kennison and Trofe (2003). The proportion of males was rated higher with the male oriented scale (100% male on the right) than with the female oriented scale (100% female on the right).

General Discussion

The aim of this study was, first, to provide norms for the gender stereotypicality of role names in three languages (English, French, German), and, second, to explore design variables that might influence these ratings.

With reference to the first purpose of our study, it can be stated that the rankings of the role names was highly reliable across the three languages as well as across variations in questionnaire format. In addition, the absolute ratings varied only slightly, though some of the effects were significant. This consensus provides a good basis for using the norms to

select stimulus material for future experimental studies designed to investigate the influence of gender stereotypes within as well as across languages.

With reference to the second purpose of our study, we used different forms of the role names and manipulated the response scale direction to investigate the stability of the stereotypicality ratings. We found that when the role names were explicitly gender marked the proportions of males were rated lower than when generically intended forms were used. Furthermore, we were able to replicate the scale direction effect reported by Kennison and Trofe (2003) and to show that it can be generalized across languages (see Study 2).

But in addition the results from Study 1 also revealed the boundaries of that effect: When applying the specific feminine and masculine forms (Study 1) instead of the generic forms (Study 2) the scale direction effect was reversed: With “100% male” on the left, the ratings were similar when the specific forms (Study 1) and the generic forms (Study 2) were presented. But with “100% male” on the right, the proportions of males were rated lower when the specific forms were presented (Study 1) than when the generic form was presented (Study 2). Thus, presenting the specific feminine and masculine versions of the role names increased the judged percentage of females in that role but only if the feminine versions were presented on the left, i.e. if the feminine version was read first. This early reading of the feminine form might have led respondents to reflect more on the possibility that women could fill that role. The finding that the response scale direction effect is moderated by the version of the role names used suggests that different mechanisms are involved when responding to the generic vs. the specific forms. This issue deserves more attention and further research should include measures on monitoring the response-process, such as the think aloud technique (concurrent verbalisation) or the measurement of response times as an indication of answer elaboration (i.e. shorter times reflecting automatic processes).

In our studies a total of four versions of a questionnaire have been compared. There

are of course more possibilities on how to present the role-names in combination with the rating scales and there are even more design variables (such as the number of rating scale categories) that could be manipulated. Thus, future research might reveal our effects to be attenuated or further amplified by other questionnaire formats. But what is important here is the fact that different questionnaire structures lead to different responses. We therefore suggest that researchers conducting norming studies on role names should pay attention to the structure of their questionnaires, and that people using stereotype norms should consider whether the effects of questionnaire format on ratings affect their use of the norms.

In this research we hope to have provided excellent materials for further investigations on the interplay of grammar and stereotypicality on gender representation in language. Furthermore, we were able to document the malleability of those representations: Our results are a further indication that explicitly referring to the fact that roles could be filled by women heightens women's visibility but only if it is done in a prominent way, i.e. by placing the feminine versions first.

References

- Baudino, C. (2001). Politique de la langue et différence sexuelle : la politisation du genre des noms de métier. Paris: L'Harmattan.
- Beggs, J. M. & Doolittle, D. C. (1993). Perceptions now and then of occupational sex typing: A replication of Shinar's 1975 study. *Journal of Applied Psychology*, 23, 17, 1435-1453.
- Colé, P. & Segui, J. (1994). Grammatical incongruence and vocabulary types. *Memory & Cognition*, 22, 387-394.
- Couch, J. V., & Sigler, J. N. (2001). Gender perception of professional occupations. *Psychological Reports*, 88, 693-698.
- Duffy, S. A. & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, 32, 551-559.
- Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*, 30, 439-446.
- Kennison, S. M. & Trofe, L. (2003). Comprehending Pronouns : a role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32, 355-378.
- Oakhill, J., Garnham, A., & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory & Cognition*, 33, 972-983.
- Rammstedt, B., & Krebs, D. (2007). Does response scale format affect the answering of personality scales? Assessing the Big Five Dimensions of Personality with different response scales in a dependent sample. *European Journal of Psychological Assessment*, 23, 32-38.
- Sastre, M. T., Fouquereau, E., Igier, V., Salvatore, N., & Mullet, E. (2000). Perception of occupational gender typing: A replication on European samples of Shinar's (1975) and

- Beggs and Doolittle's (1993) studies. *Journal of Applied Psychology*, 30, 430-441.
- Shinar, E. H. (1975). Sexual stereotypes of occupations. *Journal of Vocational Behavior*, 7, 99-111.
- Stahlberg, D., Braun, F., Irmen L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), *Social communication* (pp. 163-187). New York: Psychology Press.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory & Language*, 48, 542-562.

Author Note

This research was supported by a grant from the Swiss National Foundation (100013-109705j) to Ute Gabriel and Pascal Gygax.

We thank Emanuelle Caruzzo and Corinne Moser for their help in conducting this study.

Footnotes

¹ Both the French and the German ratings were obtained from participants living in the same country (Switzerland). This might explain the slightly higher consensus between the ratings in those two languages compared to the English ratings.

² This result holds even if the analysis is restricted to those role names ($N = 86$) that have a specific feminine and masculine form in both, French and German.

Figure Captions

Figure 1. Rating scale (100% male--left) with specific male and female role names (Study 1)

Figure 2. Mean ratings of male proportion as a function of scale direction and language for gender specific role names (Study 1, $N = 194$) and generic role names (Study 2, $N = 177$)

Figure 3. Scatterplot of mean rating for each role name ($N = 126$) by standard deviation for each role name for the English sample

Figure 4. Scatterplot of mean rating for each role name ($N = 126$) by standard deviation for each role name for the French sample

Figure 5. Scatterplot of mean rating for each role name ($N = 126$) by standard deviation for each role name for the German sample

Figure #2

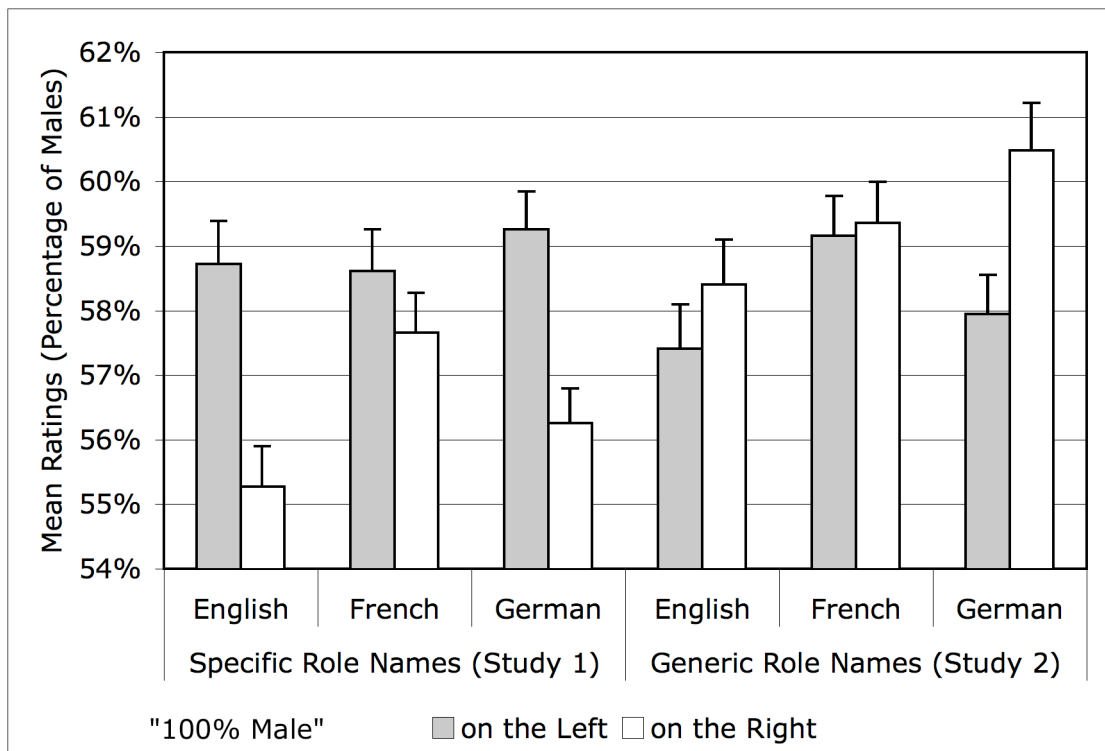


Figure #3

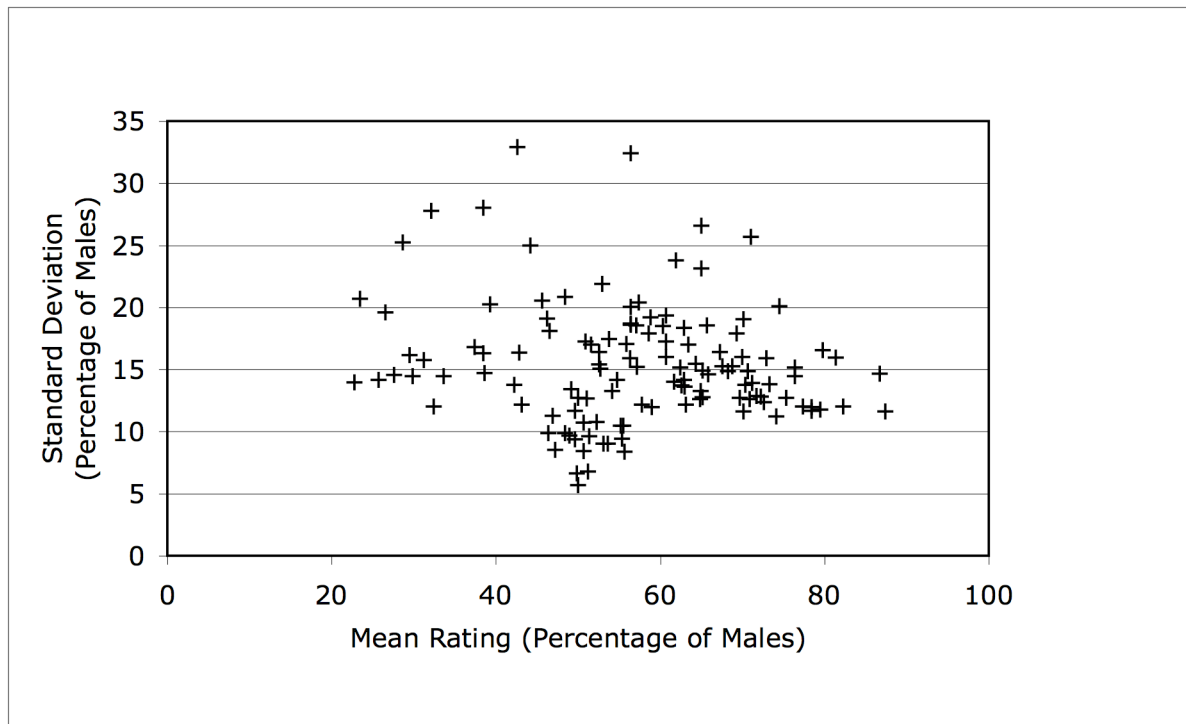


Figure #4

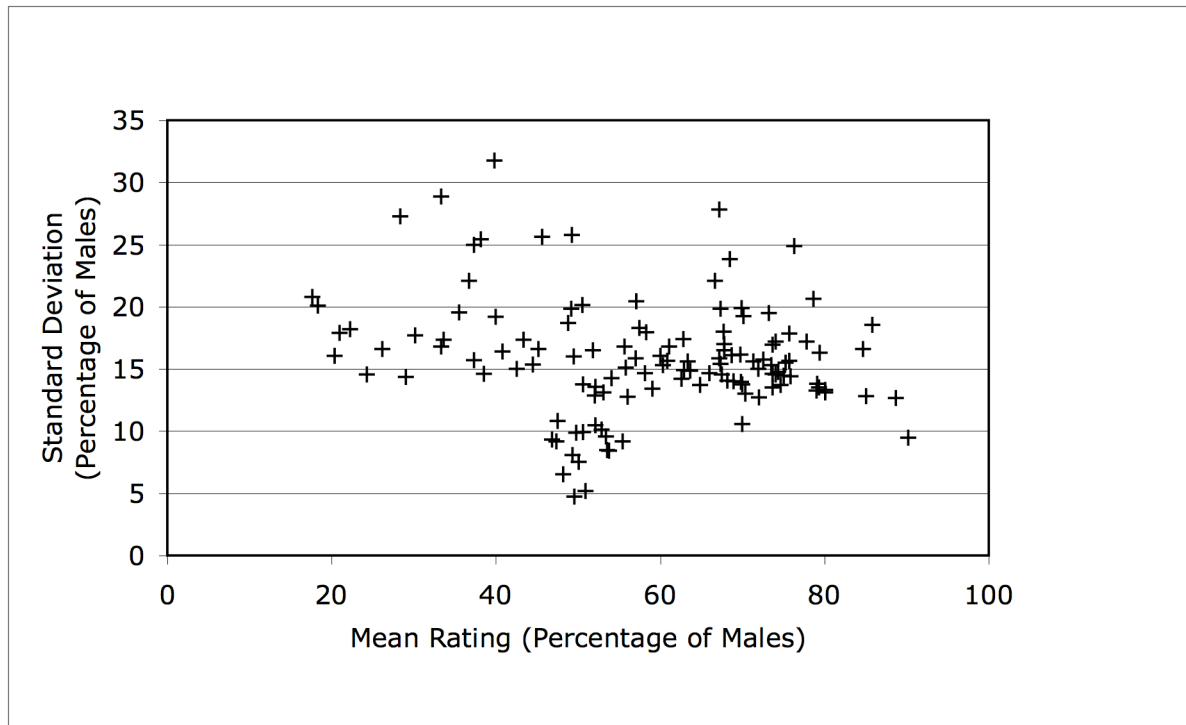


Figure #5

