



## Exploring gender stereotypes about interpersonal behavior and personality factors using digital matched-guise techniques

Inga Dennhag<sup>1</sup>, Anders Steinvall<sup>2</sup>, Camilla Hakelind<sup>3</sup>, Mats Deutschmann<sup>4</sup>

<sup>1</sup>Department of Clinical Science, Child and Adolescent Psychiatry, Umeå University, Sweden

<sup>2</sup>Department of Language Studies, Umeå University, Sweden

<sup>3</sup>Department of Psychology, Umeå University, Sweden

<sup>4</sup>School of Humanities, Education and Social Sciences, Örebro University, Sweden

How to cite: Dennhag, I., Steinvall, A., Hakelind, C., & Deutschmann, M. (2019). Exploring gender stereotypes about interpersonal behavior and personality factors using digital matched-guise techniques. *Social Behavior and Personality: An international journal*, 47(8), e8150

We explored gender stereotypes among Swedish university students ( $N = 101$ ) who were studying a course in psychology, using a matched-guise experimental design. The gender identity of a speaker in a dialogue, manifested by voice, was digitally manipulated to sound male or female. Responses to the recordings indicated that a speaker with a male voice was rated as significantly less conscientious, agreeable, extraverted, and open to experience than was the same speaker with a female voice. Regarding social behavior, there was a tendency for the speaker with a male voice to be rated as more hostile than was the same speaker with a female voice. The study findings suggest that stereotype effects, rather than real behavioral differences, may have an impact on perceived gender differences.

### Keywords

stereotypes; gender differences; interpersonal understanding; personality factors; social behavior; voice quality

*Stereotyping* has been envisaged as a selective filter acting on various social categories, such as gender—which is the category relevant to this article—with the purpose being to simplify how individuals perceive and process information about other individuals (Talbot, 2003). In short, stereotyping means that people are more likely to take in information that fits their model expectations, and ignore details that do not. Gender is of particular relevance to stereotyping because it cuts through age, race, and class (Ridgeway & Correll, 2004). In most Western societies, men have stereotypically been associated with control, individualism, and achievement, whereas women have been associated with being more emotional, more caring, and less devoted to their career (Brescoll, 2016; Talbot, 2003). Stereotypical expectations about gender also affect social situations by the way individuals predict, interpret, and react to the behavior of others. It has been shown that this is of particular relevance in educational contexts (e.g., Saar & Nordberg, 2016). For example, female students at all levels of education are generally considered to be conscientious and focused on academic achievement (Duckworth & Seligman, 2006; Spinath, Eckert, & Steinmayr, 2014). Expectations such as these may thereby limit behavioral scope and help to maintain gender inequality.

Typically, two models (Feingold, 1994) have been put forward to explain documented gender differences in traits: a biological explanatory model, whereby differences are regarded as corresponding to innate temperamental differences, and a sociocultural explanatory model, in which gender differences are seen as the product of social and cultural factors. In the sociocultural model emphasis is placed on the impact of people's expectations on the construction and perception of gendered traits and behaviors (Feingold, 1994). However, in her metastudy, Gerber (2009) showed that various personality traits that actually have their origin in status factors, often become associated with gender attributes. These include stereotypically male

personality trait categories, such as instrumental-assertive and instrumental-dominating, which are characterized by self-assertiveness and agentic and goal-oriented behaviors, and stereotypically female personality trait categories, such as expressive and submissive, which are characterized by accommodating and interpersonally oriented behavior. Complaining, which reflects an indirect way of exercising influence, is also seen as a stereotypically female trait category according to Gerber.

In accordance with Feingold (1994) and Gerber's (2009) observations, we argued that some perceived gender differences may be attributed to selective perception effects induced by stereotypic gender expectations. Consequently, a certain behavior motivated by the contextual situation, for example, being dominating, may be noticed and attributed to gender traits if it fits people's expectations (i.e., if the person being dominating is a man), or alternatively, may be ignored if it does not fit expectations (i.e., if the person being dominating is a woman). In the current experiment, we wanted to explore whether the same behavior is interpreted differently by university students depending on the perceived gender of the person.

The study is part of the Raising Awareness through Virtual Experiencing (RAVE) project, funded by the Swedish Research Council, which is aimed at developing pedagogical methods for raising students' awareness of how stereotyping may inadvertently color people's impressions of others. An essential component of our method is that our participants were exposed to this phenomenon firsthand, and for this purpose we used an updated matched-guise methodology (see, e.g., Lambert, Hodgson, Gardner, & Fillenbaum, 1960). More specifically, we investigated students' perceptions of a person's behavior during communication and their personality traits based on their interpretations of two variants of the same text recording, where the only difference between the versions was the perceived identity of the speaker as signaled by digitally manipulated voice quality. In several studies the results have shown that vocal cues and the quality of a voice trigger stereotype inferences about speakers' personality, and this phenomenon has been studied in relation to gender and masculine-/feminine-sounding voices (Ko, Judd, & Stapel, 2009; McAleer, Todorov, & Belin, 2014). Ko et al. (2009) showed that auditory cues (male or female voices) act as an overall between-category basis for gender stereotyping. In addition, they illustrated how voice quality itself has some effect on within-category judgements, even when the gender identity of the person is known.

In the current study we explored new ground in a number of ways. We focused on interpretations of interaction as represented by the behavior of a person in a constructed dialogic case. Previously, studies in which matched-guise methods were used have been based on short monologic utterances or readings (see, e.g., Ko et al., 2009). Further, as the gendered variants of our cases were digitally manipulated versions of the same recording, we were able to eliminate unwanted background variables, such as variation in pacing, pausing, and voice level, which have been highlighted as important in influencing interpretation but which are problematic to control for when different recordings, or indeed different actors, are used to produce the input stimulus (Tsalikis, DeShields, & LaTour, 1991).

### **Personality Factors, Social Behavior, and Stereotypical Preconceptions**

In our experiment we explored how students' interpretation of personality factors and social behavior differ depending on the perceived gender of an actor. The most commonly used personality theory in recent years is the five-factor model of personality, or Big Five (McCrae & Costa, 1987). In this model personality is described by five factors: extraversion, agreeableness, conscientiousness, emotional stability/neuroticism, and openness. Several scholars have documented that women report having, and are reported to have, somewhat higher levels of extraversion, agreeableness, conscientiousness, and neuroticism than men have (Costa, Terracciano, & McCrae, 2001; Feingold, 1994; McCrae et al., 2005; Vianello, Schnabel, Sriram, & Nosek, 2013). Of particular relevance for educational contexts is conscientiousness, the personality factor that has been associated with qualities such as competence, order, dutifulness, achievement striving, self-discipline, and deliberation (Costa et al., 2001; Spinath et al., 2014), and, in extension, with academic achievement (Duckworth & Seligman, 2006). It has been shown that the social context in which a person acts influences how differences in personality are evaluated (Heller, Watson, Komar, Min, & Perunovic,

2007; Robinson, 2009). For example, when Robinson (2009) included context in his study of personality traits he found that differences for conscientiousness were significant only in the context of colleagues, but not in either the parents or friends contexts. Accordingly, on the basis of studies such as those by Duckworth and Seligman (2006) and Spinath et al. (2014), we speculated that conscientiousness could be stereotypically gendered in school and university contexts.

It is important to note at this stage that all the Big Five studies discussed above are based on self-evaluations. This may be a potential source of error because of so-called *expectancy factors*, that is, social and cultural expectations of behaviors (cf. Feingold, 1994). Social expectations of behavior may influence self-reported gender characteristics in trait studies to mirror societal values of what are deemed to be desirable male and female traits. With our design in the current study of using one recording with two different voice manipulations, we aimed to capture how expectancy factors may color an individual's interpretation of another person. All observed differences are, in other words, "in the eyes of the beholder."

In addition to personality variables, we also investigated aspects related to social behaviors. Here, our focus was on dominance and submissiveness, as well as on positive and negative affective behaviors toward others, that is, behaviors that build on interpersonal theory (Kiesler, 1996), and that are included in Benjamin's (2000) theoretical framework for structural analysis of social behavior. It has been shown that social behavior and Big Five personality variables are correlated. For example, it has been demonstrated (Pincus, Gurtman, & Ruiz, 1998; Trapnell & Wiggins, 1990) that positive affective behavior correlates positively with low neuroticism (high emotional stability) and negative affective behavior correlates positively with high neuroticism (low emotional stability). Furthermore, autonomy correlates negatively with conscientiousness (Pincus et al., 1998), whereas control correlates positively with conscientiousness (Trapnell & Wiggins, 1990).

Scholars have identified gender stereotypes as being associated with social behaviors (see, e.g., Gerber, 2009), and the issue of affective behaviors and gender stereotyping has also been explored in experimental research (e.g., Durik et al., 2006; Plant, Hyde, Keltner, & Devine, 2000; Shields, 2002). Plant et al. (2000), for example, found that similar drawings of sexually ambiguous faces with different expressions were more likely to be identified as female when they were happy or sad, and male when they were angry.

## Study Aims and Hypotheses

Our overall aim was to explore whether psychology students interpret the same behavior differently depending on the perceived gender of the actor. Per the patterns identified in previous research on gender stereotypes, personality traits, and social behavior (Costa et al., 2001; Feingold, 1994; Gerber, 2009; McCrae et al., 2005; Plant et al., 2000; Robinson, 2009; Vianello et al., 2013), we formulated two working hypotheses with subsections:

**Hypothesis 1:** The perceived gender of an interlocutor, as signaled by voice quality, will affect students' ratings of the Big Five personality factors.

**Hypothesis 1a:** Compared to the female version actor, the male version actor's personality will be rated as less extraverted.

**Hypothesis 1b:** Compared to the female version actor, the male version actor's personality will be rated as less agreeable.

**Hypothesis 1c:** Compared to the female version actor, the male version actor's personality will be rated as less conscientious.

**Hypothesis 1d:** Compared to the female version actor, the male version actor's personality will be rated as less neurotic (more emotionally stable).

Note that we made no predictions regarding ratings of openness, as this factor specifically does not figure in the literature on gender stereotyping, as far as we know.

**Hypothesis 2:** The perceived gender of an interlocutor, as signaled by voice quality, will affect students'

ratings of social behavior.

**Hypothesis 2a:** Compared to the female version actor, the male version actor's personality will be rated as more assertive/controlling and less submissive.

**Hypothesis 2b:** Compared to the female version actor, the male version actor's personality will be rated as having more negative affect (e.g., acting resentfully, with dislike or ignoring).

**Hypothesis 2c:** Compared to the female version actor, the male version actor's personality will be rated as having less positive affect (e.g., less understanding, less supportive).

## Method

### Overall Framework

A key component of project RAVE is response cases, which are built on a digitally updated matched-guise methodology. The objective is to produce two gender-altered versions of a case dialogue based on a single recording. Thus, in one version, the actor's voice is manipulated so that Person A sounds like a woman, and the other recording is manipulated so that Person A sounds like a man. Respondents are split into two groups and each group member listens to one of the two versions of the recorded text. In an immediate postexposure survey, the test participants are then asked to respond to items related to personality factors and the social behavior of Person A.

### Contextualization and Presurvey

In a matched-guise set-up it is of critical importance that the participants are initially unaware of the real purpose of the experimental design. Thus, in this experiment we contextualized the purpose as a case task in personality psychology, where the students believed they were taking part in an exercise aimed at highlighting different methods for measuring personality factors and social behavior. A presurvey included demographic information about the respondents, comprising their age, gender, and self-evaluation of national identity.

### Participants

The sample consisted of 101 people (36 men, 64 women, one intersex individual). All participants described their ethnic identity as Swedish, and all were attending an undergraduate university course in personality psychology. At the time of the experiment the class was already divided into two groups: Group 1 (Case version 1;  $n = 55$ ; 23 men, 31 women, one intersex individual) and Group 2 (Case version 2;  $n = 46$ ; 13 men, 33 women). The mean age in Group 1 was 23.97 years ( $SD = 5.23$ ) and in Group 2 it was 23.86 years ( $SD = 4.83$ ). There were no significant differences between the groups regarding gender and age. The responses of two students were excluded from the final analysis because it was obvious that they had misunderstood which person in the case dialogues they were supposed to rate; their answers deviated markedly and systematically from the rest of the group.

### The Case Scenario

The case scenario was constructed as a spoken mixed-gender dialogue in four short scenes illustrating Students A and B working on a joint thesis project. In the first scene, the students discuss a critique they have received on their joint project work in a midprocess seminar, and how they can proceed with their work. The two subsequent scenes represent planning and negotiations of who does what, and in the final scene the students reflect on the final feedback they have received on their project. In the dialogue, Person A comes across as slightly more dominant, self-assured, and demanding than Person B does, but is also encouraging and positive. To avoid confusion and split attention, we asked students to evaluate only Person A.

The dialogue was recorded in a studio using two actors specifically chosen for their voice quality. The

deciding criterion for choice of voice quality was that the voices responded well to digital manipulation (voice morphing) and the resultant “male” and “female” versions had to be perceived as just that. In pursuit of this, we recorded five male and five female actors who each read a short passage (the Rainbow Passage) and we then voice morphed the recordings to produce male and female versions. We used Avid Pro Tools HD 12.0.0 to record the actors and then edited each voice with the same software. Pitch shifting was processed manually with X-Form (rendered only) using Elastic Audio properties in Pro Tools. The resultant 20 voice samples were then sent out to 25 peers whom we asked if, in their view, (a) the recordings sounded natural and (b) each sounded convincing as a male/female voice. On the basis of their responses, we chose the two voices (actors) that were evaluated most positively.

Once the dialogue between the two actors whose voices were chosen had been recorded and edited, we digitally manipulated the voices so that one case version gave the impression that Person A was female and Person B was male, and reversed the perceived gender identities of the speakers in the second version so that Person A sounded more masculine and Person B more feminine. It is important to note here that the actor playing Person A was female, and both versions of her recording were manipulated, as were both versions of the recording of Person B, which was read by a male actor. By building both case versions on the same recording we guaranteed that crucial aspects of the interpretation of the dialogue and its participants, such as intonation and force, remained identical. Finally, the sound files were incorporated into two digital packages consisting of one version in which Person A was “male” and Person B was “female” and vice versa for the second version. The package also included neutral silhouette images of a male student and female student engaged in activities relevant to the case scenario.

## Response Surveys

We also included a response survey in which the students rated Person A’s personality factors and social behavior, directly after they had listened to the manipulated recording of the dialogue.

**Ten-Item Personality Inventory.** The TIPI (Gosling, Rentfrow, & Swann, 2003) is an abbreviated form of the Big Five Inventory. It is a widely used ten-item personality inventory with good test–retest reliability and adequate levels of convergent validity (Gosling et al., 2003). It contains two items for each of the five personality traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness. The items are rated on a 7-point Likert scale ranging from 1 (*disagree*) to 7 (*agree*). Gosling et al. (2003) reported low to acceptable Cronbach’s alphas ranging from .40 (agreeableness) to .73 (emotional stability). In this study Cronbach’s alphas were .54 for extraversion, .72 for agreeableness, .73 for conscientiousness, .70 for emotional stability, and .50 for openness to experience. The low to acceptable reliability estimates could be attributable to the small number of items in each scale (Gosling et al., 2003).

**Social behavior.** We developed eight statements regarding perceived social behavior that were designed to measure different actions directed toward another interlocutor. These statements were inspired by the theoretical framework of Benjamin (2000), and consisted of two central axes: *control* (control–submissiveness) and *affiliation* (love–hate).

Submissiveness:

- Person A shows a submissive behavior and lets Person B decide.

Control:

- Person A asserts an opinion, decides over and controls Person B.

Positive affiliation (love):

- Person A shows openness, acceptance, and understanding toward Person B.
- Person A shows warmth and expresses liking toward Person B.
- Person A shows trust toward, helps, and supports Person B.

Negative affiliation (hate):

- Person A is resentful of, accuses, and criticizes Person B.
- Person A shows dislike toward, attacks, and threatens Person B.
- Person A withdraws from and ignores Person B.

The items are rated on an 11-point scale from 0 (*completely disagree*) to 10 (*completely agree*). In this study Cronbach's alphas for the positive and negative affect subscales showed high internal consistency at .83 and .78, respectively. As there was only one item to rate in the subscales of control and submissiveness, Cronbach's alpha could not be estimated; however, Pincus et al. (1998) demonstrated that short-form surveys with one item per subscale show acceptable to good reliability.

### Classroom Procedure

The class ( $N = 101$ ) was split into two subgroups, whereby Group 1 ( $n = 55$ ) listened to Case 1 (Person A = male, Person B = female), and Group 2 ( $n = 46$ ) listened to Case 2 (Person A = female, Person B = male). Each participant was given a headset and could listen to the recording only once. There were no indications either in an ensuing debriefing discussion or in the responses to a postsurvey that the students were aware of the voices being digitally manipulated, or that there was any confusion as to the gender of Persons A and B in the two case versions.

### Data Analysis

To compare TIPI and social behavior ratings of the students who listened to Person A with a male or a female voice, two multivariate analyses of variance were performed. Rated case version was used as the between-subjects factor, and scores on social behavior (four variables) and the TIPI (five variables) were used as dependent variables in the respective models. Age and gender were controlled for in both models. Effect sizes were expressed as eta-squared values (i.e.,  $\eta^2$ ). Effect sizes of .01, .06, and .14 indicate small, medium, and large effects, respectively (Cohen, 1988).

**Model fit and test of assumptions.** Following guidance from Tabachnick and Fidell (2013), the data were found to meet the assumptions of multivariate analyses in relation to normal distribution, linearity, univariate outliers, multivariate outliers, homogeneity of variance–covariance matrices, multicollinearity, and singularity.

## Results

### Bivariate Correlations Between Dependent Variables

Bivariate correlations were calculated to make sure that data were correlated as expected for all study variables, as well as for gender and age (see Table 1). All significant correlations between these two macrocategories were in the expected direction. For example, there was a significant positive correlation between positive affective behavior and emotional stability (i.e., low levels of neuroticism), which is in line with the findings reported in earlier studies (Pincus et al., 1998; Trapnell & Wiggins, 1990). There was no multicollinearity (Tabachnick & Fidell, 2013).

Table 1. *Pearson Correlations Between Social Behavior, Personality Factors, Gender, and Age*

	1	2	3	4	5	6	7	8	9	10
1. Extraversion										
2. Agreeableness	.18									
3. Conscientiousness	.09	.41**								
4. Emotional stability	.04	.32**	.93**							
5. Openness to experience	.34**	.55**	.25**	.08						
6. Submissive behavior	.17	.52**	.09	-.04	.29**					
7. Controlling behavior	-.11	-.63**	-.21*	-.15	-.37	-.64**				
8. Positive affective behavior	.33**	.68**	.38**	.26*	.58**	.50**	-.60**			
9. Negative affective behavior	-.23*	-.59**	-.39**	-.36**	-.41**	-.25**	.39**	-.43**		
10. Gender	.26*	.02	.13	.04	.12	-.01	-.02	-.04	-.18	
11. Age	-.22*	-.01	-.10	.10	-.04	-.08	-.08	-.02	-.22*	-.36**

Note. \*  $p < .05$ , \*\*  $p < .01$ .

### Differences Between Voice Conditions on the Ten-Item Personality Inventory and Social Behaviors

**Influence of rated case on the Ten-Item Personality Inventory's subscales.** A one-way between-groups multivariate analysis of variance was performed to investigate case differences in the TIPI personality factors. Five dependent variables were used: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. Sample descriptives are shown in Table 2.

Table 2. *Mean Item Scores on the Ten-Item Personality Inventory and Social Behaviors*

Subscales	Case 1 Person A with a male voice		Case 2 Person A with a female voice	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Extraversion	4.46	1.00	4.99	1.12
Agreeableness	2.49	1.18	3.18	1.32
Conscientiousness	4.68	1.55	5.83	1.17
Emotional stability	4.40	1.42	4.90	1.38
Openness to experience	3.32	0.97	3.85	1.22
Person A shows a submissive behavior and lets Person B decide	1.36	1.85	1.33	1.83
Person A asserts opinions; decides over and controls Person B	7.36	2.22	7.30	2.19
Person A shows positive affective behavior towards Person B	3.21	1.76	3.96	1.92

Note. Case 1,  $n = 55$ ; Case 2,  $n = 46$ .

When we controlled for gender and age in the analysis, there was a statistically significant overall difference between how the students rated the two case versions in relation to the TIPI factors:  $F(5, 93) = 4.41, p < .001$ ; Wilks' Lambda = .78;  $\eta^2 = .19$ . When the results for the dependent variables were considered separately, four results were significant: extraversion,  $F(1, 97) = 5.21, p = .025$ ;  $\eta^2 = .05$ ; agreeableness,  $F(1,$

97) = 7.66,  $p = .007$ ;  $\eta^2 = .07$ ; conscientiousness,  $F(1, 97) = 15.71$ ,  $p < .001$ ;  $\eta^2 = .14$ ; and openness to experience,  $F(1, 97) = 5.06$ ,  $p = .027$ ;  $\eta^2 = .05$ . The result for emotional stability was nonsignificant,  $F(1, 95) = 2.78$ ,  $p = .099$ ;  $\eta^2 = .03$ . Mean scores for the TIPI's subscales are shown in Figure 1.

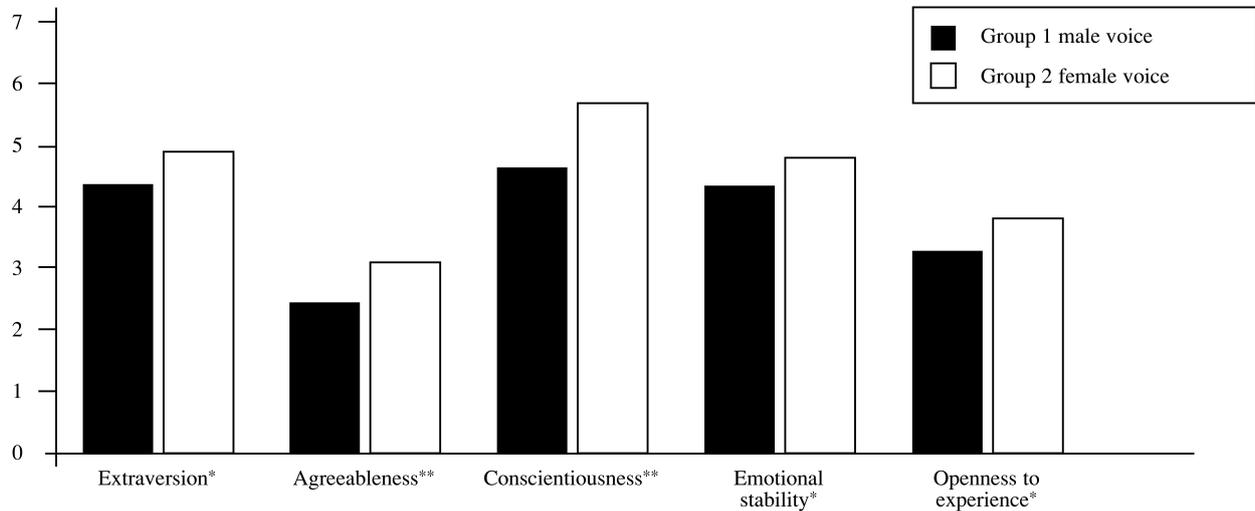
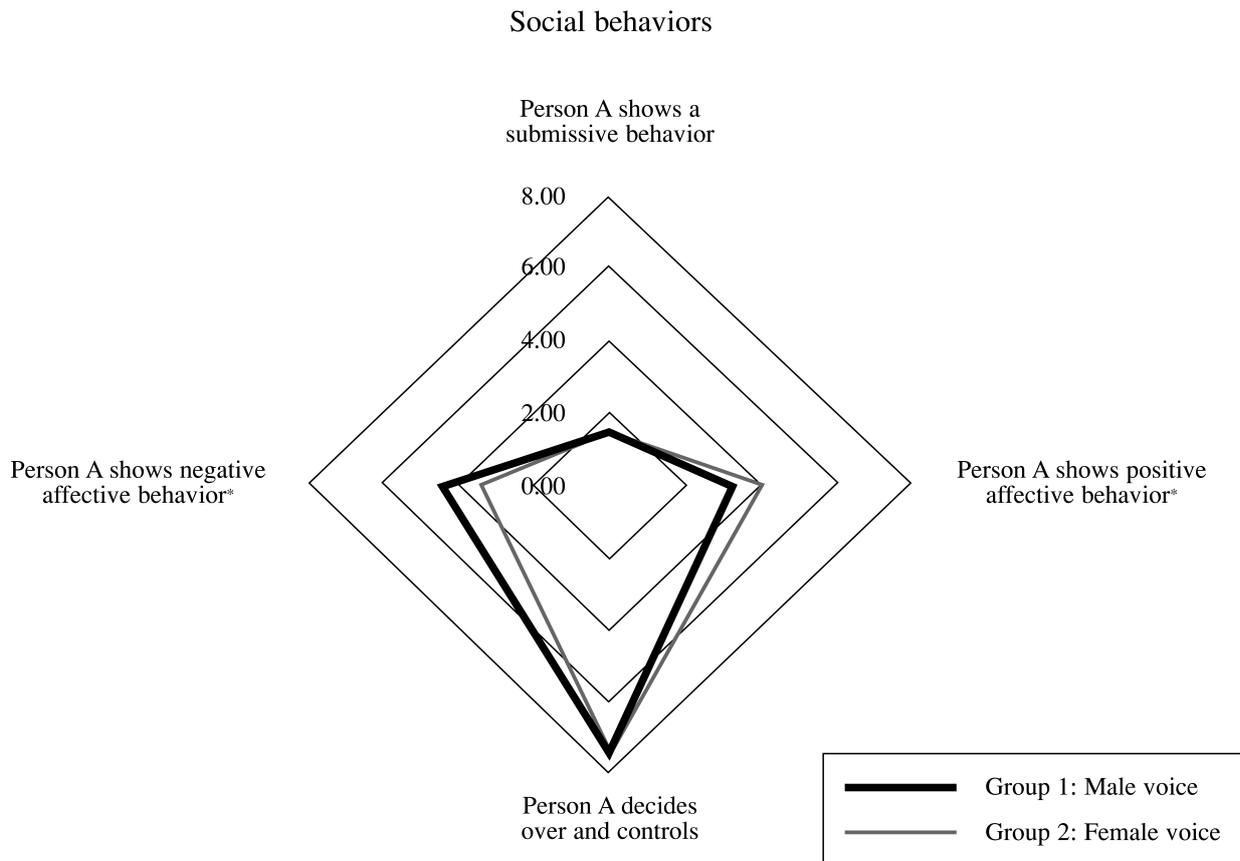


Figure 1. Mean scores for the subscales of the Ten-Item Personality Inventory for Group 1 (Person A rated with a male voice) and Group 2 (Person A rated with a female voice).

\*  $p < .05$ , \*\*  $p < .01$ .

**Influence of rated case on social behaviors.** We analyzed four dependent variables for social behavior: submissive behavior, controlling behavior, positive affective behavior, and negative affective behavior. Sample descriptives are shown in Table 2. When we controlled for gender and age in the analysis, there was a statistically significant overall difference between how the students rated the case (Person A with a male or female voice),  $F(4, 94) = 2.68$ ,  $p = .036$ ; Wilks' Lambda = .90;  $\eta^2 = .10$ . When the results for the dependent variables were considered separately, there were two significant results: When Person A spoke with a male voice, he was rated as showing less positive affective behaviors toward Person B,  $F(1, 97) = 4.08$ ,  $p = .046$ ;  $\eta^2 = .04$ , and more negative affective behaviors toward Person B,  $F(1, 97) = 5.60$ ,  $p = .020$ ;  $\eta^2 = .05$ . Eta square values were considered to be small. Submissive behavior and controlling behavior were rated as not significantly different for Person A with a male or female voice,  $F(1, 97) = .001$ ,  $p = .975$ ;  $\eta^2 = .00$ , and  $F(1, 97) = .002$ ,  $p = .968$ ;  $\eta^2 = .00$ . Mean scores for social behaviors are shown in Figure 2.



*Figure 2.* Mean scores for social behavior subscales for Group 1 (Person A rated with a male voice) and Group 2 (Person A rated with a female voice).

\*  $p < .05$ .

## Discussion

In this study we explored how two groups of undergraduate students rated the personality and social behavior of the same participant in a dialogue in two different gender guises (male and female). Our research was designed to address how the rating of the Big Five personality factors and social behavior may be affected by the perceived gender of the person as manifested through voice quality. The results support both our overall hypothesis that the perceived gender of an interlocutor, as signaled by voice quality, will affect the listener's ratings of personality factors, and our overall hypothesis that perception of gender will also affect listener ratings of social behavior. More specifically with reference to the Big Five factors, the results support Hypotheses 1a, 1b, and 1c, in that those who rated Person A with a male voice perceived the speaker as less extraverted, less agreeable, less conscientious, and less open to experience than did those who rated Person A with a female voice. The largest effect size was observed for conscientiousness and it was at a level considered to be large, Cohen's effect size ( $\eta^2$ ) = .14. The other effect sizes were small to medium. There were no differences in the interpretations of the two case versions with regard to emotional stability (neuroticism, Hypothesis 1d).

In the part of the study related to social behavior, our results supported Hypotheses 2b and 2c: The speaker

with a masculine voice was interpreted as showing less positive affective behaviors and more negative affective behaviors toward the other speaker than was a speaker with a feminine voice. Our results did not support Hypothesis 2a, in which we had predicted that the female speaker would be interpreted as more submissive/less dominating than the male character. Overall, our results mirror many of the findings from previous studies that point to the existence of gender stereotypes, according to which women are perceived as being more friendly and more conscientious than are their male counterparts, especially in professional contexts, and men are perceived as being more hostile and less caring (see, e.g., Gerber, 2009). According to our results, our respondents interpreted the two case versions differently, and focused on personality factors and social behaviors that confirmed their stereotypic gender beliefs.

In terms of Big Five traits, our results were very much in line with those reported in previous studies (Costa et al., 2001; Feingold, 1994; McCrae et al., 2005; Vianello et al., 2013) in that the speaker was perceived by the rater as more extraverted, more agreeable, and more conscientious when in the female guise. However, in contrast to results in previous studies, we found no difference according to the speaker's gender with respect to neuroticism, which has been identified as a female trait (McCrae et al., 2005; Vianello et al., 2013). The reason for this could be that the speaker being rated had few, if any, neurotic characteristics. An important difference between the present study and previous studies in which personality factors and gender have been addressed, is that self-evaluation tests have very often been employed in previous research. Feingold (1994) and Vianello et al. (2013) hypothesized that expectancy effects could be significant factors explaining gender differences in studies using self-evaluation tests. In such studies it has, however, been difficult to separate real differences, that is, social role model and/or biological differences, from perceived (expectancy model) differences. In this respect we further address below the methodological benefit of the approach we used in our study.

As regards the effect of context, the scenario we chose was firmly situated in a university setting familiar to the respondents, a feature that could have had an effect on the respondents' perception of traits in relation to gender (Duckworth & Seligman, 2006; Spinath et al., 2014). We found it interesting that the most significant gender difference was for conscientiousness. It has been previously found that conscientiousness correlates with qualities such as competence, dutifulness, and achievement striving (Costa et al., 2001; Spinath et al., 2014), and is specifically associated with girls' behavior in school settings (Duckworth & Seligman, 2006; Spinath et al., 2014). In accordance with the suggestions put forward in previous studies (Heller et al., 2007; Robinson, 2009), we speculate that context may have an influence on gender differences in the evaluation of personality factors, but this needs to be explored further.

With respect to social behaviors and positive and negative affect, our results are consistent with the findings of previous studies about gender stereotypes, where anger and pride have been associated with masculine behavior, whereas love, happiness, and most other emotions have been associated with how people perceive womanly behavior (Durik et al., 2006; Plant et al., 2000; Shields, 2002). Furthermore, our results are in line with those reported by Plant et al. (2000), who used faces in drawings and photographs as visual cues to investigate how gender stereotypes affect the perception of different emotional facial expressions.

We were surprised that there were no differences in the students' interpretations of social behavior related to dominance and submissiveness according to the case. This result goes against those of Gerber (2009), who showed that dominating traits are often associated with men and submissive traits are often associated with women. There could be several explanations for our result. One is that, in comparison with the general population, among the students that comprised our sample the roles of women and men are generally perceived as more equal with regard to dominance. Another explanation is that the specific context of the experiment comprising student-to-student interactions in a university setting influenced the respondents. In such a setting, we would argue, the expectation is that both men and women should be able to express dominance.

Although many of our results in this study are in line with what has been previously demonstrated, we argue that this study is unique in that our results show how gender stereotyping, activated through the cue of a male or female voice, can lead to stereotype-like interpretations of the same behavior in a dialogic situation. There is, consequently, a very strong case for claiming that the differences we found between the two groups in their perception of Person A's behavior and traits were solely in the ears, or rather, in the minds, of the listeners. Thus, the expectancy effects hypothesized by Feingold (1994) and Vianello et al. (2013) would seem to be confirmed in this study. Our results thereby suggest that stereotypic gender expectations affect judgements of personality and social behavior, and may, therefore, have a distortive impact on results in studies exploring gender differences in personality and social behavior. However, the degree of this impact remains to be explored. The method we used in this study further points to the opportunity that digital technology provides in terms of controllable material in matched-guise designs.

### **Limitations and Future Research Directions**

There are a number of limitations in the present study. First, the study group consisted of a homogeneous sample of Swedish psychology students, which raises issues of generalizability. Second, we acknowledge that the effects of voice quality itself may have been an important background variable influencing our results. As previously demonstrated (Ko et al., 2009), voice quality is not only an important cue leading to between-category stereotyping, but also affects within-category judgements. There is, thus, scope for further studies of similar design to ours, where voice quality variation and its effect on within-category stereotyping are explored. A third limitation is the lack of previous validation of the eight items we used to assess social behaviors. However, the measure we developed demonstrated good reliability for both positive and negative affectivity subscales and showed convergent validity with the TIPI through the correlations between similar scales. Nevertheless, the use of a single item for the submission and control subscales may be a weakness in that it could create polarization of behaviors. We note, however, that short-form questionnaires with single items for social behaviors have been used in several other studies with sustained reliability (e.g., Pincus et al., 1998). Finally, with the exception of conscientiousness and agreeableness, after controlling for gender and age, the majority of our critical probability values ranged from .02 to .05, so that their reliability can, thus, be questioned (Simmons, Nelson, & Simonsohn, 2011). Larger samples are needed to increase the reliability of our results.

### **Wider Implications**

Our study findings strongly imply that stereotypic gender expectations regarding social behavior and personality color people's interpretation of an event. This, in turn, may imply that differences found in previous studies of this nature are not entirely a reflection of real trait differences, but may also be affected by selective individual interpretations, created and maintained by stereotypic filters. Our results thereby carry important implications for studies on gender stereotyping in which investigation of personality traits and social behavior are based on impressions. As pointed out by Hyde (2014), it may be that in studies based on self-evaluations and evaluations of others, gender differences rather than similarities are exaggerated. There is, thus, a danger that research serves not only to confirm gender stereotypes, but also to run the risk of legitimizing them.

### **Acknowledgements**

Additional information regarding the equations and models in the study can be obtained directly from the first author.

### **References**

Benjamin, L. S. (2000). *Structural analysis of social behavior SASB/Intrex user's manual for short, medium, and long questionnaires. Part 1: Instruction for administering questionnaires, interpreting reports, and giving rater feedback*. Salt Lake City, UT: University of Utah.

Brescoll, V. L. (2016). Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders. *The Leadership Quarterly*, 27, 415–428. <https://doi.org/10.1016/j.leaqua.2016.02.005>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81, 322–331. <https://doi.org/10.1037/0022-3514.81.2.322>

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98, 198–208. <https://doi.org/10.1037/0022-0663.98.1.198>

Durik, A. M., Hyde, J. S., Marks, A. C., Roy, A. L., Anaya, D. A., & Schultz, G. (2006). Ethnicity and gender stereotypes of emotion. *Sex Roles*, 54, 429–445. <https://doi.org/10.1007/s11199-006-9020-4>

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116, 429–456. <https://doi.org/10.1037/0033-2909.116.3.429>

Gerber, G. L. (2009). Status and the gender stereotyped personality traits: Toward an integration. *Sex Roles*, 61, 297–316. <https://doi.org/10.1007/s11199-008-9529-9>

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)

Heller, D., Watson, D., Komar, J., Min, J.-A., & Perunovic, W. Q. E. (2007). Contextualized personality: Traditional and new assessment procedures. *Journal of Personality*, 75, 1229–1254. <https://doi.org/10.1111/j.1467-6494.2007.00474.x>

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 373–398. <https://doi.org/10.1146/annurev-psych-010213-115057>

Kiesler, D. J. (1996). *Contemporary interpersonal theory and research: Personality, psychopathology, and psychotherapy*. Oxford, UK: Wiley.

Ko, S. J., Judd, C. M., & Stapel, D. A. (2009). Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin*, 35, 198–211. <https://doi.org/10.1177/0146167208326477>

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, 60, 44–51. <https://doi.org/10.1037/h0044430>

McAlear, P., Todorov, A., & Belin, P. (2014). How do you say ‘Hello’? Personality impressions from brief novel voices. *PLoS ONE*, 9, e90779. <https://doi.org/10.1371/journal.pone.0090779>

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81–90. <https://doi.org/10.1037/0022-3514.52.1.81>

McCrae, R. R., Terracciano, A., & 78 members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer’s perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, 88, 547–561. <https://doi.org/10.1037/0022-3514.88.3.547>

Pincus, A. L., Gurtman, M. B., & Ruiz, M. A. (1998). Structural analysis of social behavior (SASB): Circumplex analyses and structural relations with the interpersonal circle and the five-factor model of

personality. *Journal of Personality and Social Psychology*, 74, 1629–1645.  
<https://doi.org/10.1037/0022-3514.74.6.1629>

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, 24, 81–92. <https://doi.org/10.1111/j.1471-6402.2000.tb01024.x>

Ridgeway, C. L., & Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & Society*, 18, 510–531.  
<https://doi.org/10.1177/0891243204265269>

Robinson, O. C. (2009). On the social malleability of traits: Variability and consistency in Big 5 trait expression across three interpersonal contexts. *Journal of Individual Differences*, 30, 201–208.  
<https://doi.org/10.1027/1614-0001.30.4.201>

Saar, T., & Nordberg, M. (2016). To be created as a “real” boy – Masculinity creation in school [In Swedish]. *Kapet* (elektronisk), 12, 5–20. Retrieved from <https://bit.ly/2S2D1EZ>

Shields, S. A. (2002). *Speaking from the heart: Gender and the social meaning of emotion*. Cambridge, UK: Cambridge University Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.  
<https://doi.org/10.1177/0956797611417632>

Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students’ intelligence, personality and motivation? *Educational Research*, 56, 230–243.  
<https://doi.org/10.1080/00131881.2014.898917>

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson Education.

Talbot, M. (2003). Gender stereotypes: Reproduction and challenge. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of language and gender* (pp. 468–486). Malden, MA: Blackwell.

Trapnell, P. D., & Wiggins, J. S. (1990). Extension of the Interpersonal Adjective Scales to include the Big Five dimensions of personality. *Journal of Personality and Social Psychology*, 59, 781–790.  
<https://doi.org/10.1037/0022-3514.59.4.781>

Tsalikis, J., DeShields, O. W., Jr., & LaTour, M. S. (1991). The role of accent on the credibility and effectiveness of the salesperson. *The Journal of Personal Selling & Sales Management*, 11, 31–41.  
<https://doi.org/10.1080/08853134.1991.10753857>

Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. (2013). Gender differences in implicit and explicit personality traits. *Personality and Individual Differences*, 55, 994–999.  
<https://doi.org/10.1016/j.paid.2013.08.008>