Gendered Expectations for “Agreeableness” in Response to Requests and Opinions

Felicia Roberts & Alda Norris

To cite this article: Felicia Roberts & Alda Norris (2016) Gendered Expectations for “Agreeableness” in Response to Requests and Opinions, Communication Research Reports, 33:1, 16-23, DOI: 10.1080/08824096.2015.1117437

To link to this article: http://dx.doi.org/10.1080/08824096.2015.1117437

Published online: 12 Jan 2016.

Submit your article to this journal

View related articles

View Crossmark data
Gendered Expectations for “Agreeableness” in Response to Requests and Opinions
Felicia Roberts & Alda Norris

To address the need for more direct assessment of gendered expectations for communication behaviors, we devised a novel experimental approach that tests whether expectations for “agreeableness” are more salient when evaluating male or female speakers in their affirmative responses to a friend’s request or expression of opinion. Briefly, as lag times are introduced, it appears that females are expected to be more agreeable (or male delay is more tolerated), particularly when alignment with opinion is at stake or when a female is responding to another female. Findings are discussed in the context of Expectation Violations Theory.

Keywords: Conversation; Expectation Violations; Gender; Silence

The first few decades of research on gender and language were rightly criticized for characterizations of women as monolithically “agreeable” and “cooperative.” Scholars have since developed more theoretically rich understandings of women’s actual communicative behavior (Eckert & McConnell-Ginet, 2013). However, the question remains whether there is, nonetheless, an expectation for women to display affirming and compliant communication behaviors as reflected in early studies. Recent indirect evidence indicates that, despite advances in gender equity, women and men conform to normative expectations for communication behaviors in the workplace (Guadagno & Cialdini, 2007). This suggests that underlying assumptions about what is acceptable gendered behavior may be enduring. The goal of the current study is to more directly test this by asking whether expectations for “agreeableness” are more salient when

Felicia Roberts (PhD, University of Wisconsin–Madison, 1998) is a Professor of Communication at Purdue University. Alda Norris (PhD, Purdue University, 2011) is an Evaluation Specialist in Cooperative Extension at the University of Alaska-Fairbanks. Correspondence: Felicia Roberts, Brian Lamb School of Communication, Purdue University, 2170 Beering Hall, West Lafayette, IN 47907. E-mail: froberts@purdue.edu
evaluating male or female speakers. Our findings are discussed in the context of Expectation Violations Theory (Burgoon, 1978).

An experimental approach was created in which study participants listened to stimulated telephone conversations among an ostensible group of male and female friends. The conversations were devised so that the interactants were rated on the same communicative action: affirmative responses to a friend’s request or expression of opinion.

To examine whether the “agreeableness” ratings were based simply on gender or, more subtly, on expectations that agreement should be quick and forthcoming from either males or females, brief silences were inserted in the conversations to simulate hesitancy or reluctance to respond (Davidson, 1984). This technique allowed us to compare as directly as possible raters’ judgments by varying the sex of the speakers and the rapidity with which they responded. Our prediction is that in conditions where reluctance or lack of agreement is signaled by extended silence (i.e., lengths of 600 ms or longer; see Kendrick & Torreira, 2015; Roberts & Francis, 2013), female speakers will be judged as less agreeable than male speakers, even when responding affirmatively. Although assessments and requests are different speech acts, one requiring agreement/disagreement and the other requiring an indication of willingness (or not), both can be categorized as actions that seek alignment. Based on this, the following hypothesis was derived:

\[ H1: \text{In conditions where longer lag times are present, ratings for females responding to assessments and requests will be lower than ratings for males, regardless of the speech act type.} \]

Method

Experimental Design

In a mixed factorial design, three levels of interturn silence length (0 ms, 600 ms, 1200 ms) and two levels of speech act (assessment and request) were within-group factors. Gender pairing in the dialogues, which captured all combinations of same- and cross-sex possibilities, was a between-groups factor (female-female = FF; male-male = MM; male-female = MF; female-male = FM).

The dependent variable was the study participants’ ratings of “agreeableness” (in the assessment condition) or “willingness” (in the request condition.) These judgments were captured as ratings on an ordered scale from 1 (not willing or not in agreement) to 6 (very willing or very much in agreement). Only one measure for each construct was used. In what follows, we first provide an overview of the task for study participants and then explain in detail the rationale for, and construction of, the study stimuli.

Overview of Subject Recruitment and Study Task

Study participants (\(N = 142\)) were enrolled in independent convenience samples, recruited from classrooms where instructors had given permission for the last 10 minutes of the class to be used for voluntary participation in “a study about
communication.” Stimuli were counterbalanced across the various samples. For the task, participants first listened to recorded instructions and then listened to nine audio recordings of simulated telephone calls and answered a question after each dialogue. The constructed dialogues (further detailed in the following), concerned mundane themes (flyers for a school function, new furniture, and going to the gym), and each one ended with the caller either formulating a request in terms of that topic (e.g., getting a ride to pick up the flyers) or offering an opinion on the topic at hand (e.g., reporting that the flyers look good). The call recipients answered in the affirmative for both speech acts (“sure” for the request and “yeah” to display agreement with the opinion). By answering in the affirmative, compliance/agreement was lexically specific and clear, but the silent gaps introduced in lag-time conditions, especially over 600 ms, would likely color those affirmative responses as potentially reluctant.

Upon hearing the recipient’s affirmative response to the request or assessment, study participants rated, on a 6-point scale, their perception of the speaker’s enthusiasm for assenting to the request or agreeing with the assessment. Ratings higher on the scale indicated a perception of greater willingness on the call recipient’s part to comply with the request or agree with the assessment.

Construction of Stimuli

The stimuli for this study were simulated telephone calls based on an actual telephone call among friends of an age similar to the study population (Roberts & Robinson, 2004). The same two females and the same two males performed all target and masking (“distractor”) dialogues. For the targets, the same caller/call-recipient roles were maintained to control, within gender, for any variation in response that might be elicited by voice quality of the caller/responder. Because pitch is not always a reliable cue to speaker sex/gender, conventional gendered names (Kevin and Dan, Rachel and Amber) were used to cue listeners to the sex of the speakers. To control for possible confounding from the acoustic qualities within each actor’s slightly different productions of the “sure” and “yeah” response tokens in their various dialogues, we identified a median response token for each actor and pasted that into all of their relevant conversations. In other words, the quality of the affirmative response never varied within each actor.

Insertion of Interturn Silence

Once the median response token for each actor was edited into the appropriate conversations, silences were inserted between the focal speech act and the response token. Three levels of “gap” were used to test whether delay of an affirmative response would affect ratings. A baseline “no-gap” condition was also presented, so if there were any fundamental bias toward the voices in the dialogues, the baseline condition would capture that. This could serve as a kind of “control” or comparison measure for the delay conditions. Two delay conditions were constructed, one that was about half a second (600 ms) and one that was just over 1 second (based on Jefferson, 1988). If
there was an effect of gender even at the roughly half-second mark, the effect of gender was robust.

Presentation to Study Participants

Each counterbalanced audio presentation consisted of nine conversations: three ended with the assessment/response pairing (“I think it/they look(s) pretty good” followed by “Yeah”), which was separated by one of the three interturn silence lengths (0 ms, 600 ms, 1200 ms); three conversations ended with the request/response pairing (“Can you give me a ride over there” followed by “Sure”) separated by the interturn silence lengths; three nontarget conversations ended with an assessment or request (also paired with the affirmative response tokens “sure” and “yeah”). The groups were similar in terms of age, with most participants falling between 18 and 22 years. The ratio of male to female participants, overall, was 32% male, 68% female.

Results

Effects from age or sex of the study participants were tested by including the factors as covariates in an omnibus ANOVA. Because there were no interaction effects of age or sex with core factors of interest (gap length, speech act, or gender pairing in the dialogues), nor were there between-subjects effects for sex or age of study participant, $F(1,136) = 0.248, F(1,36) = 0.463$ respectively; sex and age were dropped from the model and Bonferroni corrected post hoc tests for multiple comparisons run.

The ANOVA obtained significant main effects for gap, $F(2,141) = 3.64, p < .001; \eta_p^2 = .72$, and for speech act, $F(1,141) = 6.09, p = .02; \eta_p^2 = .04$, as well as a significant two-way interaction between gap and gender pair, $F = 2.22, 6,141, p = .04; \eta_p^2 = .05$. A three-way interaction among gap, speech act, and gender pair was also statistically significant, $F(6,113) = 7.02, p < .001; \eta_p^2 = .13$. (See Table 1 for means and standard deviations in each condition). Note that effect sizes, calculated using partial Eta squared, were small, except for the main effect of gap length. While interpretation of effect sizes should be made cautiously, the effect of gap length accounts for most of the overall variance, indicating the importance of this factor on study participants’ ratings of an addressee’s willingness to comply with requests or agree with assessments. This was taken into account as the results were decomposed.

Female Responder Receives Lower Ratings Than Male as Lag Time Increases

Since ratings of willingness decrease with increased silence, regardless of language background (Roberts, Margutti, & Takano, 2011), rater response patterns within each speech act and gap length condition were examined (two one-way ANOVAs, one for each speech act, and with gender pair as the between-subjects factor.) For assessments, there were no statistically significant differences between the mean scores for the different gender pairings in the assessment condition at the 0 ms gap, $F(3,141) = 1.48,$
This suggests that in a normally responsive temporal framework, the male and female voices did not trigger different assumptions about agreeability when responding affirmatively ("yeah") to their friends’ offered opinions on mundane topics (e.g., “I saw the flyers and they look pretty good!”)

For requests, however, there was a statistically significant difference between gender pairs, $F(3,141) = 10.94, p < .001$, in the no-gap condition. Post hoc comparisons of ratings of willingness revealed that the ratings for the female responder in comparison with ratings for the male responder differed significantly. In other words, in the compliance condition (agreeing to help the friend by giving him/her a ride), the female responder was judged less willing even though there was no delay in her response. The largest of these differences was in the cross-sex condition comparing F-M (i.e., female asking male) and M-F ($M = 1.05; SE = .233, p < .001$). In that comparison, the female responding to the male received much lower scores of willingness ($M = 3.82$) than the male responding to the female ($M = 4.88$). In the 600 ms gap condition (roughly a half second), lag time significantly affects ratings in both the assessment, $F(3,141) = 4.48, p < .05$, and the request conditions, $F(3,141) = 5.75, p < .050$, a finding that was confirmed in post hoc comparisons of all gender pairs, with female response consistently receiving lower scores than male responses in the assessment condition. In the longest gap condition, 1200 ms, the pattern of ratings in the two speech acts was the inverse from the findings in the no-gap condition: The ratings for agreement with the assessment were significantly different between gender pairings, $F(3,141) = 4.39, p < .05$, which was confirmed in follow-up tests of gender pair comparisons. However, in the request condition, while the ANOVA indicated an overall statistically significant difference between gender conditions, $F(3,141) = 3.03, p < .05$, the post hoc tests indicated that only one comparison was

### Table 1 Mean Scores for Gender Pairs Across Speech Acts and Inturn Silence Lengths

<table>
<thead>
<tr>
<th>Gap length</th>
<th>Caller-Responder</th>
<th>N</th>
<th>Requests Mean (SD)</th>
<th>Assessments Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ms</td>
<td>M-M</td>
<td>36</td>
<td>4.81 (0.89)</td>
<td>4.25 (0.84)</td>
</tr>
<tr>
<td></td>
<td>F-M</td>
<td>33</td>
<td>4.88 (0.89)</td>
<td>4.24 (0.94)</td>
</tr>
<tr>
<td></td>
<td>F-F</td>
<td>39</td>
<td>4.03 (1.18)</td>
<td>3.87 (1.00)</td>
</tr>
<tr>
<td></td>
<td>M-F</td>
<td>37</td>
<td>3.82 (0.87)</td>
<td>4.00 (0.91)</td>
</tr>
<tr>
<td>600 ms</td>
<td>M-M</td>
<td>36</td>
<td>3.67 (0.99)</td>
<td>3.61 (1.10)</td>
</tr>
<tr>
<td></td>
<td>F-M</td>
<td>33</td>
<td>3.15 (0.97)</td>
<td>3.21 (1.05)</td>
</tr>
<tr>
<td></td>
<td>F-F</td>
<td>39</td>
<td>2.79 (0.95)</td>
<td>2.97 (0.81)</td>
</tr>
<tr>
<td></td>
<td>M-F</td>
<td>37</td>
<td>3.32 (0.78)</td>
<td>2.81 (1.00)</td>
</tr>
<tr>
<td>1200 ms</td>
<td>M-M</td>
<td>36</td>
<td>2.44 (0.88)</td>
<td>2.50 (1.08)</td>
</tr>
<tr>
<td></td>
<td>F-M</td>
<td>33</td>
<td>2.36 (0.90)</td>
<td>2.61 (0.79)</td>
</tr>
<tr>
<td></td>
<td>F-F</td>
<td>39</td>
<td>1.97 (0.81)</td>
<td>2.18 (0.79)</td>
</tr>
<tr>
<td></td>
<td>M-F</td>
<td>37</td>
<td>2.57 (1.04)</td>
<td>1.95 (0.73)</td>
</tr>
</tbody>
</table>
Discussion

The purpose of this study was to gather judgments of female and male speakers doing the same communication behavior: agreeing with their ostensible friend’s mundane requests and opinions. By holding the behavior constant (affirmative responses), but manipulating the sex of the speakers and the lag time before an affirmative response, some access was gained to listener perceptions of “agreeableness.” Overall, in terms of H1, that raters would score females as less agreeable than males as lag times increased, we see support for that; however, the pattern of results for the different speech acts is more complex than predicted. In the assessment condition, when agreement with another’s opinion is at stake, the actors are not rated differently when the response is quick and affirmative (i.e., no gap); however, as lag times are introduced, it appears that females are expected to be more agreeable (or male delay is more tolerated). In the request condition, when compliance is at stake, the female is consistently rated lower, even without lag time, and particularly in her response to a male’s request for help. Interestingly, as the delay to comply increases, the ratings drop further when the female is responding to a request for help from another female. In other words, “cooperativeness” of a female may be judged more severely when she displays reluctance to help another female. Apparently, the compliant dimension of the feminine social role is differentially applied, depending on interlocutor.

Expectancy Violations Theory (Burgoon, 1978) suggests that when behavioral expectations are violated, social actors may orient to them in one of two ways: in terms of social norms that they have internalized through experience (predictive expectations) or through what they have come to think is right or correct for a certain group (prescriptive expectations; Staines & Libby, 1986). Thus, if there is either a predictive or prescriptive expectation that women’s stereotypical social role is to be more agreeable and compliant, then expectations are violated when their affirmation of another’s opinion or compliance with another’s request is delayed. This normative orientation to agreeableness translates into lower ratings for females who delay in their responses.

The differential ratings for the female and male conditions could simply be an artifact of the vocal qualities of the different actors. However, several contradictory findings allay this concern. First, the ratings of the female respondent are not always lower than ratings of the male respondents (see Table 1); second, if vocal attractiveness alone were driving the judgments, then one might expect statistically significant differences in all no-lag-time conditions. In fact, it is primarily in the lag conditions where the gender effect is salient, supporting a conclusion that it is the delay in concert with the perception of a gendered voice that is affecting ratings. Finally, because the female tends to be rated as more agreeable in conversation with the male than she is with the female (Table 1), we
do not believe that it was something intangible in the actors’ voices causing the different ratings, but more likely the different expectations for behavior in the different gender pairings.

This approach opens several avenues for research. First, the findings can be tested with other voices to see if they are peculiar to this group of actors; second, the content of the opinions and requests can be manipulated to investigate whether the findings fade or intensify in settings where the interpersonal stakes are higher, such as helping an ill friend or family member, agreeing or disagreeing with politically charged opinions, and so on. And finally, more complex measures of these constructs could be used as researchers begin to decompose the interaction of displays of reluctance (lag time before response) and gender roles. This study intentionally used mundane requests and opinions, controlled the vocal quality as much as possible using sound editing, and simplified the task to one measure each of “willingness” and “agreement.” Future research should empirically test these initial findings about gendered expectations for agreeableness and cooperativeness and to more finely explore under what conditions those expectations might be attenuated or amplified.

Notes


[2] By using two gap conditions, we could approach the limit of what listeners would accept as a reasonable or normal delay (1200 ms) as well as test what might be considered a more subtle or less easily noticed delay (600 ms).

[3] Although this imbalance in sex of study participants might have caused some concern, there was no statistically significant effect from the sex of study subjects.

References


