

What the Voice Reveals

RIJKSUNIVERSITEIT GRONINGEN

What the Voice Reveals

Proefschrift

ter verkrijging van het doctoraat in de
Gedrags- en Maatschappijwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
maandag 11 juni 2007
om 14.45 uur

door

Sei Jin Ko
geboren op 15 mei 1971
te Seoul, Korea

Promotores:

Prof. dr. D. A. Stapel

Prof. dr. C. M. Judd

Copromotores:

Dr. E. Gordijn

Dr. S. Otten

Beoordelingscommissie:

Prof. dr. N. Ellemers

Prof. dr. D. Wigboldus

Prof. dr. N. W. van Yperen

Acknowledgments

For me, this section has been the most difficult to write. The cynical Sei Jin told me to skip it all together since one is always bound to forget somebody deserving thanks. The sentimental Sei Jin was horrified. We struggled, we negotiated, and we finally came to a settlement...

I will not specify names as my meager recall would shame me and all of you. However, I would like to thank everyone who has crossed my life, past and present, friends, family, and otherwise—you have taught and inspired me a great deal. Most of all, I am extremely in awe of you for putting up with me. I offer you my profoundest thanks.

Contents

Chapter 1	General Introduction	1
Chapter 2	What The Voice Reveals: Within And Between Category Stereotyping on the Basis of Voice	13
Chapter 3	A Voice in the Workplace: How Vocal Cues Impact Judgments Related to Hiring Decisions	35
Chapter 4	Sneaking in Through the Back Door: How Category-Based Stereotype Suppression Leads to Rebound in Feature-based Effects	55
Chapter 5	General Discussion, Conclusions and Final Remarks	61
References		70
Summary		82
Nederlandse Samenvatting		84
Footnotes		87
Tables		90
Figures		96
Appendix A		102

Chapter 1

General Introduction

True to our social nature, it would hardly be an exaggeration to state that we humans devote most of our waking hours to verbal interactions. In fact, one need only reflect on the many social interactions that occur each day to realize the extent to which people rely on their voices. Indeed, in some situations the voice is our only medium of communication (e.g., telephone). For example, imagine two people, who met through an internet-dating service, having their first phone conversation or potential employees on phone interviews—what kind of perceptions are elicited by their voices? And, consequently, how might these perceptions influence the dating relationship and whether the interviews are successful? Given this, there is surprisingly little research on how our voice may impact the way we are perceived in these social interactions. Perhaps this is because social psychologists assume that auditory cues are not as salient as visual ones and hence auditory cues provide less information than that provided by visual cues. However, the innumerable tales of the profundity with which people's lives have been impacted by the way their voice sounds—from schoolteachers to CEOs—suggests that the public does not share this assumption. The research presented in this dissertation was an attempt to make the anecdotal empirical.

It never ceases to amaze me just how much we put faith into first impressions to make sense of world. As you become more familiar with a person, you obviously become privy to more and more information about the person that was not available in the first meeting. Nonetheless, your initial impressions of the person may very well have set the scene for the type of relationship that you forge with the person. For instance, if your first impression leads you believe that the person is slightly dim and lazy but warm and funny then you may act towards the person in such a way as to leave little choice but to elicit behaviors that confirm your beliefs (Dougherty, Turban, Callender, 1994; Skrypnek & Snyder, 1982; for a review

see Jussim & Fleming, 1996). Hence, examining the triggers that lead to these first impressions may provide intriguing insight to the way we construct and interpret our world. To this end, in my dissertation, I focus on the stereotypic inferences made on the basis of meeting a person for the first time. More precisely, I have examined the stereotypic inferences elicited by speakers' vocal cues.

When you think about vocal cues that have the potential to affect gender-stereotyping, the first thing that you probably think of are the differences between men and women's voices. Vocal cues certainly vary between gender category but they also vary within a gender category. Specifically, voices within each gender vary in the degree to which they sound masculine/feminine, which we define as "vocal femininity". The overarching theme of the presented research is the examination of how this vocal femininity affects gender-stereotypic inferences of individuals of the same gender, over and above any stereotyping based on differences between gender category. In the rest of this introduction, I provide an overview of the chapters that follow.

Overview of the Chapters

Chapter 2

What The Voice Reveals:

Within And Between Category Stereotyping on the Basis of Voice

There exists a small body of voice research suggesting that people infer many different qualities about speakers on the basis of their voice. For example, the voice provides cues to the speaker's emotions (Yogo, Ando, Hashi, Tsutsui, & Yamada, 2000), personality traits (Aronovitch, 1976; Scherer, 1979), attractiveness (Berry, 1992; Collins, 2000; Zuckerman & Driver, 1989), maturity (Berry, 1992; Hummert, Mazloff, & Henry, 1999; Linville, 1996; Mulac & Giles, 1996), and likely occupation (Yamada, Hakoda, Yuda, & Kusuhara, 2000). However, surprisingly little research in social psychology has been

conducted in the auditory compared to the visual domain. A further challenge to social psychology's bias towards visual cues is offered by the recent surge of interest in the role of vocal cues in the expression of emotion (for a review see Scherer, Johnstone, Klasmeyer, 2003). Scherer (2003), for instance, asserts that there are many possible variations of vocal expressions of an emotion (e.g., quiet, brooding anger vs. raging anger) whereas visual expressions of the same emotion are not as versatile (see also Banse & Scherer, 1996). Thus, rather than yielding relatively meager information, vocal cues may sometimes be even richer than visual cues.

In sum, the above evidence clearly indicates that people's inferences about others are not confined to visual cues. Concurrently, it also hints at the potentially important role the voice may play in social interaction but there is currently no systematic research of its role in such a context. Chapter 2 presents a first step in examining how voice affects one essential aspect of social interaction—stereotyping. We started by investigating the acoustic characteristics that give rise to perceptions of vocal femininity and category. Then we examined how these perceptions of vocal femininity impacts stereotypic inferences over and above category. Finally, we provide a model of the relation among acoustics, vocal femininity, and stereotypic inferences.

Chapter 3

A Voice in the Workplace:

How Vocal Cues Impact Judgments Related to Hiring Decisions

In Chapter 3, I present research examining the consequences of vocal stereotyping in the specific social context of job interviews where these consequences can have far-reaching implications. Gender discrimination, especially as it applies in the workplace, has been well documented in the literature (e.g., Davison & Burke, 2000; Deal & Stevenson, 1998; Eagly & Karau, 2002; Eagly, Makhijani, & Klonsky, 1992; Glick, Zion, & Nelson, 1988; Heilman,

2001; Rudman & Glick, 1999; Schein, 1975). However, less is known about the factors that may trigger discrimination in the first place. Therefore, it is important to conduct systematic investigations of the factors that are present in specific contexts where gender discrimination is likely to occur. Accordingly, in Chapter 3, we focus on a point in the hiring process that appears particularly prone to gender discrimination, namely the interview stage. Focusing on the interview stage also seems important given that recruiters place a great deal of emphasis on interviews since most, if not all, hiring decisions rely upon some form of interview. We were specifically interested in the evaluative processes involved in the interview stage where recruiters are provided with many sources of information about the job applicants. The following quote nicely illustrates the importance of interviews:

“You decide that you want a job. You send in your brilliant (and mostly accurate) résumé, ask for an outrageous salary (with full benefits, perks, and a generous signing bonus), and get a letter one week later saying that you're hired. Congratulations! You are officially living in fantasyland. Today, a résumé only functions as an entrance test to snag an interview. It's during the interview where a company decides if it wants to hire you.”
(<http://www.soyouwanna.com/>)

The level of anxiety that most job candidates associate with interviews would clearly suggest that they too place considerable emphasis upon interviews. The question then is what kind of information is given in personal interviews that is not available on written résumés and why is such information regarded as so important. One obvious kind of information that is revealed in interviews are physical cues about a person, (e.g., their appearance, tone of voice, manner of dress). As a start to answering this question, we examined how job applicants' vocal and background information cues, in the form of résumés, affect judgments of applicants in an interview context.

The work presented in Chapter 3 extends our work from Chapter 2 in important ways. Specifically, even though the work in Chapter 2 was the first demonstration of within-gender stereotyping, based on vocal femininity, one could still argue that because perceivers were

only presented with vocal cues, it was only logical that they used these cues in their inferences. Thus, a more compelling case for the role of vocal femininity in stereotyping would be the demonstration of its impact alongside other competing information. Accordingly, the research in Chapter 3 extends our past work on voice by examining whether vocal femininity cues continue to exert an influence on judgments even when other rich competing cues are concurrently available.

Chapter 4

Sneaking in Through the Back Door:

How Category-Based Stereotype Suppression Leads to Rebound in Feature-based Effects

The seemingly paradoxical persistence of discrimination in the face of strong societal norms that dictate against category-based discrimination prompted us to investigate the relationship between category- and within-category feature-based stereotyping, which is presented in Chapter 4. Over the past fifty years there has been a dramatic reduction in the overt expression of racist and sexist attitudes and stereotypes in Western societies (Campbell, 1971; Greeley & Sheatsley, 1971; Hyman & Sheatsley, 1956, 1964; Schuman, Steeh, & Bobo, 1985; Taylor, Sheatsley, & Greeley, 1978). Yet, more subtle approaches to the measurement of stereotypes and prejudice, focusing on implicit or automatic evaluations and beliefs that are not under conscious control, suggest that racial and gender biases are still alive and well (Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995; Wittenbrink, Judd, & Park, 1997). This suggests that strong social norms have developed that encourage many people to curb the open-expression of prejudice and stereotypes based on social categories such as race and gender (Klonis, Plant, & Devine, 2005).

Certainly this effort to suppress the expression of category-based stereotypes and prejudices is not without consequences. According to the social psychological research on suppression, active efforts to suppress a thought actually increase the thought's activation and

the probability that the suppressed thought may leak out in other forms or on other occasions (Wegner & Erber, 1992; Wenzlaff, Wegener, & Klein, 1991). With continued practice and monitoring, of course, it may be the case that overt category-based stereotypes and prejudice never leak out, at least in their original form (cf. Macrae, Bodenhausen, Milne, & Jetten, 1994). Rather, the active suppression of such category-based beliefs may mean that they leak out in more subtle and seemingly harder-to-control forms.

We propose that one of these subtle ways is through within-category feature or cue effects. The hypothesis that motivated the presented work is that efforts to suppress category-based stereotyping are well practiced and therefore unlikely to show rebound effects. However, suppressing category-based stereotyping may lead to rebound in another form, with greater feature-based stereotyping following active attempts to suppress category-based stereotyping. In our view this might be the case because feature-based stereotyping is more subtle and harder to control than category-based stereotyping.

Indirect support for our proposal comes from recent work on racial stereotyping showing that perceivers had little trouble suppressing category-based stereotyping but were unable to suppress stereotypes arising from variations in the degree to which targets, within race, had African American (Afrocentric) features (Blair, Judd, & Fallman, 2004). Specifically, in this research, prior to a judgment task, participants were either instructed to “try avoid using the person’s race as a basis for judgment” (Blair et al., 2004, p. 770; category suppression condition) or instructed to “try to avoid using the person’s Afrocentric features as a basis for judgment” (Blair et al., 2004, p. 770; features suppression condition). In a last condition, participants were given no such instructions (control condition). The judgment task results showed that in the category suppression condition participants were less likely to use targets’ racial category as a basis for their judgments than in the control condition. No such effect was observed in the features suppression condition; participants’

use of within-category features were unaffected by the instruction manipulation. To explain the lack of control over feature-based stereotyping, the authors argued that perceivers are largely unaware of feature-based influences. Furthermore, suppressing these features is complex because feature-based judgments are literally not black or white as are category-based judgments—features are riddled with myriad shades of gray (e.g., think of variations in nose size). This complexity is further heightened by the fact that not one but many different cues (e.g., nose size, kinkiness of hair, fullness of lips) constitute Afrocentric features, hence suppressing feature-based stereotypes would require monitoring the influence of the variation in all these cues. The work in Chapter 4 was the first demonstration of the relationship between category- and feature-based stereotyping that may have important societal implications.

Due to the fact that the empirical chapters are based on articles that have been published or under review, there will be overlaps in the theoretical introductions and discussions.

As time passes and the inevitable happens—the detailed memory of this research totally evaporates—I have a hope and a wish. My hope is that I will have opened the eyes of the reader to the richness of our voice and its ability to impact our lives. My wish is that the simple memory of this will never decay. Without further ado, let me lead you through a journey of what the voice reveals.

dictates that men, on average, would have more masculine voices than women, and hence a speaker's gender can be easily determined on the basis of voice (Coleman, 1976; P. Smith, 1979). There seems to be both biological and cultural bases for this gender difference. For instance, differences in body size and vocal tract morphology dictate that men's voices on average will be more resonant (Fant, 1966; Fitch & Giedd, 1999) and lower pitched than that of women's (Graddol & Swann, 1983; Rendall, Kollias, Ney, & Lloyd, 2005). Whereas factors such as men's (as compared to women's) restricted pitch variability and tendency to use the lowest part of their pitch register, has been suggested to be culturally based (Aronovitch, 1976; Graddol & Swann, 1983). Whatever the bases, it is clear that many aspects of men and women's vocal characteristics differ. So when the credit card clerk Chris' telephone greeting sounds masculine, one would categorize Chris as a man but if Chris's voice sounds feminine then one is likely to instead categorize Chris as a woman. Even if we know of no such empirical work to date, it stands to reason that gender-consistent stereotypic inferences would ensue on the basis of gender-categorizing Chris—a male Chris might be seen as assertive, analytical, and cold, whereas a female Chris might be seen as sensitive, kind, and gossipy. From our vantage point, a more interesting possibility is that this same gender-signaling set of cues may also be used to stereotype people *within* each gender category. That is, voices within each gender naturally vary in the degree to which they contain feminine (vs. masculine) cues (e.g., high pitch)²—just as a male Chris' vocal femininity differs from another male clerk's, a female Chris' vocal femininity would also differ from another female clerk's. Consequentially, listeners' stereotypic inferences may not only be confined to those ensuing on the basis of between-gender categorization but also on the basis of this variance in vocal femininity within gender. Putting this in the context of Chris, our credit-card clerk—perhaps a more feminine-sounding Chris would lead us to associate more female stereotypic qualities with him/her (e.g., sensitive, kind, gossipy), and a

more masculine-sounding Chris would lead us to associate more male-stereotypic qualities with him/her (e.g., assertive, analytical, cold), *regardless* of whether Chris is a man or a woman.

Work by Blair and colleagues, focusing on the association between racial stereotypes and facial features (e.g., Blair, Judd, Sadler, & Jenkins, 2002; Blair, Judd, & Chapleau, 2004; Blair, Judd, & Fallman, 2004) provides support for this distinction between category-based and (within-category) feature-based stereotyping. Even though these findings are confined to the visual rather than the auditory domain, they, nonetheless support our general theoretical framework.

Overview of Current Research

We designed a series of studies to examine the role of vocal cues in gender stereotyping. In the first study, we collected vocal samples from male and female speakers to verify that voices can be reliably scaled (within each gender) on femininity. We also conducted a preliminary examination of the specific acoustic characteristics (e.g., pitch, resonance) associated with femininity. Two additional studies were then conducted to determine whether vocal femininity would lead to gender stereotypic inferences, over and above the influence of gender category. In Study 2.2, we examined this question by holding the gender category of the voices constant: For some participants, all stimulus voices were clearly male, for others, they were all clearly female. In Study 2.3 participants judged voices of both genders. Here our goal was to determine whether both gender category and (within category) feature-based stereotyping could be observed simultaneously.

Study 2.1

The goal of this study was to determine if, within each category, male and female voices could be reliably scaled on the degree to which they sounded feminine. A secondary goal of this study was to examine some of the acoustic characteristics that might be

associated with speakers' gender category and vocal femininity. To do this, we recruited two sets of participants for the study, one set whose voices were recorded and another who judged those recorded voices on femininity. We then analyzed several acoustic characteristics of the voices and examined their relation with gender category and vocal femininity.

Method

Recording Participants and Procedure

One hundred and twelve undergraduate students (54 male and 58 female) at the University of Colorado were recorded in exchange for partial fulfillment of course credit or monetary compensation. All were native English speakers with an average age of 19 years ($SD = 2.33$ years).

We recorded the participants in a quiet room using a head-worn microphone and the software Sound Studio installed on IMAC computers. During the recording session the participants read several passages, one of which was the "Rainbow Passage" (Fairbanks, 1940 p.127), which is the only passage that was used as stimuli in the research reported in this article. This passage is commonly used in voice research for a number of reasons, the most relevant for our purpose being its unvalenced content and neutrality with regard to gender stereotypes (e.g., Addington, 1968; Batstone & Tuomi, 1981; Ray, 1986). From these recordings, 47 female voices and 47 male voices, saying the first two sentences of the "Rainbow Passage", were used as the final stimuli. We had to discard 11 of the female and 7 of the male recordings because they contained too many reading errors.

Judgment Participants and Procedure

Fifty-four undergraduate students (24 male and 30 female) at the University of Colorado provided judgments of the recorded voices, in exchange for partial fulfillment of course credit or monetary compensation. These participants were assigned to judge either the male or the female voices. Each participant listened to the 47 voices in the assigned set,

played in one of three random orders, and rated each voice on femininity, using a 1 (not at all) to 7 (very) scale.³

Analysis of Acoustic Characteristics

Our choice of acoustic characteristics was guided by results from established acoustical research (e.g., Byrd, 1994; Coleman, 1976; Graddol & Swann, 1983; Kramer, 1977; Peterson & Barney, 1952; Ray, 1986; Tusing & Dillard, 2000; Whiteside, 2001). Specifically, we chose cues that have been associated with gender. Although the list of acoustic cues that we selected is far from exhaustive, our chosen cues of pitch (Coleman, 1976; Graddol & Swann, 1983), variability in pitch (Graddol & Swann, 1983; Aronovitch, 1976), formant frequency (Coleman, 1976; Fitch & Giedd, 1999; Peterson & Barney, 1952; Whiteside, 2001), and speech rate/duration (Byrd, 1994; Aronovitch, 1976) were some of the cues shown to differ most consistently between gender. Our primary goal was to examine whether these cues, that are distinct between genders, are also related to perceptions of vocal femininity *within* each gender.

In our measurement of these cues, we combined the controlled methods commonly used in acoustical research (see Peterson & Barney, 1952) with a more real-world method. Namely, the “Rainbow Passage”, upon which we base our acoustic measurements, is controlled in that the same passage is spoken by all voices but concurrently more “real-world” in that it is in complete sentences which seems more akin to what listeners hear everyday than single syllables that are frequently used in acoustic analyses (for an exception see Graddol & Swann, 1983).

Pitch. The pitch of a voice is measured by its fundamental frequency (F0) in Hertz (Hz), an index of the number of vibrations per second of the vocal folds to produce phonation (Fitch, 1994; Tusing & Dillard, 2000). Higher pitched voices are expected to be judged as more feminine.

Resonance. The resonance of the voice is akin to the timbre of an instrument and, as such, is orthogonal to the pitch (Fitch, 2002): A violin and a cello may be playing the same middle C but the two instruments will have quite different resonant qualities which provide a direct auditory cue to the instruments' different sizes. Only relatively recently did Fitch (1997) introduce a reliable way to measure the resonance of the voice through what he termed the "formant dispersion" (D_f). D_f is derived by averaging the distance between adjacent pairs of formant frequencies (also measured in Hz). When formant frequencies are closer together, the D_f will be lower, which translates to a more resonant voice, one which may be perceived to be less feminine.

Pitch variability. Momentary variability in pitch during speech was assessed by calculating the standard deviations of pitch (*SD* of pitch) across the vocal sample for each voice. More variability in pitch is akin to a less monotone (more "sing-song") voice, which may be perceived as more feminine.

Speech duration. This was defined as the time taken to say the first two sentences from the "Rainbow Passage," including both the spoken and silent parts of each voice clip, in seconds.

The vocal analysis software *Wavesurfer* (Sjölander & Beskow 2002) outputted the F0 and the first four formant frequencies for each of the 47 male and 47 female voice clips. To assess the four vocal characteristics, for each voice we averaged the F0, computed and averaged the D_f from the four formant frequencies (as detailed in Fitch, 1997), and computed the *SD* of F0. The software Sound Studio outputted, for each voice, the time taken to read the first two sentences of the passage.

Results and Discussion

Judgments of Femininity

To examine the reliabilities of the mean femininity scores for each voice, the individual femininity ratings (for each voice by each participant) were analyzed with a two-way analyses of variance within the male and female voice sets, treating judges/participants as one factor and voice as another. From these analyses, variance components were estimated and intraclass correlations (due to voice, treating participants as a random factor; ICC 2.1 from Shrout & Fleiss, 1979) were calculated in order to estimate the reliabilities of the resulting vocal femininity values (following procedures outlined in Judd & McClelland, 1998). For the female voices, the mean femininity score was 4.44 ($SD = 1.60$) and the reliability of the voice means was .96. For the male voices, the mean femininity score was 3.58 ($SD = 1.62$) and their reliability was .95. Not surprisingly, the female voices were rated to be significantly more feminine than the male voices, $F(1, 92) = 12.18, p < .001$.

These data clearly demonstrate that voices can be reliably scaled on femininity. There was high consensus among the participants on the voices' femininity, even though they were not told which vocal cues they should attend to in making their judgments. As a consequence, the mean femininity score for each voice will be used throughout the research reported in this paper to define that voice's standing on perceived femininity.

Acoustic Characteristics, Gender Category and Judged Femininity

Having established the reliability of the femininity ratings, we now turn to the question of the extent to which the acoustic characteristics we measured are related to the speakers' gender category and the perceived femininity of the voices within each gender. Means (and standard deviations) for the four acoustic cues are given in the top half of Table 2.1, broken down by the gender of the voices. Presented in the bottom half of Table 2.1 are the bivariate and partial correlations (partialling out the other three acoustic cues) between

speakers' gender category (0=male; 1=female) and each acoustic cue. The pitch, resonance, and variability of pitch were strongly, positively correlated with category, which is consistent with past acoustical research showing gender differences in voice (e.g., Coleman, 1979; Graddol & Swann, 1983). However, contrary to research suggesting gender differences in speech duration (Byrd, 1994), we did not find such a relationship. A plausible cause may be the short length and uniformity of the text (i.e., the "Rainbow Passage") that did not allow for speakers to show much variability on this dimension. In the partial correlations, pitch and resonance remained significantly related to category.

Table 2.2 presents both simple correlations of each of the four acoustic characteristics with judged vocal femininity and partial correlation coefficients for each characteristic controlling for the others. For the female voices, the bivariate correlations suggest that those voices that were judged as more feminine had higher pitch, were less resonant, and were more variable in their pitch, with the latter two correlations only marginally significant. In the partial correlations, both pitch and resonance showed significant partial relations with femininity. In the case of the male voices, the bivariate correlations suggest that higher voices and those with more variable pitch were judged as higher in femininity. In the partial correlations, only pitch manifested a significant partial relation with femininity.

Comparisons between the results of Tables 2.2 and 2.3 showed that the acoustic cue of pitch, which yielded the largest correlation with gender category, was also the cue that was most highly correlated with perceptions of vocal femininity, both within male and female voices. Resonance and pitch variability showed correlations both with category and with female vocal femininity, albeit weaker with the latter, but only variability was related to male vocal femininity. The partial relations between acoustic cues and category were also apparent between acoustic cues and femininity, though within male voices, only pitch yielded a significant effect over and above the other cues.

Obviously, these results are preliminary and the measured acoustic cues are far from exhaustive in that other potentially important acoustic characteristics, such as vocal energy or volume, were not assessed. Nonetheless, these acoustic cues yielded some interesting patterns. Specifically, cues of pitch, resonance, and variability in pitch that distinguished between speakers' gender category were also the same cues that were important determinants of the degree to which voices within gender were perceived to be feminine.

Study 2.2

Having shown in Study 2.1 that there is variance in the femininity of both male and female voices, and that listeners can detect this variance with high consensus, the purpose of our second study was to demonstrate that listeners spontaneously make gender-stereotypic inferences on this basis. Thus, the participants in this study either made judgments about a set of voices that were clearly identifiable as female or about a set of voices that were clearly identifiable as male. In other words, the gender of the voices was kept constant for each participant and the question of interest was whether within-gender variation in vocal femininity (as judged in Study 2.1) would predict participants' judgments.

To assess more spontaneous inferences that participants might make on the basis of vocal cues, without explicit instructions to make impression ratings, we adapted the experimental procedure used by Blair et al. (2002). Specifically, the participants were lead to believe that the study was about accuracy in interpersonal judgments and, accordingly, they were asked to read several short self-descriptions and, following each description, to judge a series of voice clips on the probability that each speaker was the person who wrote the description. The self-descriptions varied in the degree to which they depicted someone who was stereotypically female or male and in the person's likeability (valence). Our purpose was to assess, for each of the self-descriptions, the extent to which participants' probability ratings of the male or female voices were associated with the voices' femininity. We

predicted that, within each gender, more feminine voices would be seen as more probable for the female-stereotypic self-descriptions and as less probable for the male-stereotypic self-descriptions. We included valence as a factor in the study since this is an inherent aspect of stereotypes and of social judgment, more generally. However, we did not make a clear prediction of how the valence of the self-descriptions might influence the judgments of more or less feminine sounding voices. Over the years, researchers have suggested that women are viewed more negatively than men (Broverman, Broverman, Clarkson, Rosenkrantz, & Vogel, 1970; Rosenkrantz, Vogel, Bee, Broverman, & Broverman, 1968), that women are viewed more positively than men (Eagly & Mladinic, 1989; Eagly, Mladnic, & Otto, 1991), and most recently, that feelings toward women are often ambivalent and depend on the context and salience of particular attributes (Fiske, Xu, Cuddy, & Glick, 1999; Glick & Fiske, 1996; Glick, Diebold, Bailey-Werner, & Zhu, 1997). In sum, the valence findings to date are mixed. Thus, we were most interested in judgment variance due to the confluence of voice femininity and the stereotypicality of the self-descriptions, leaving open the question of whether judgments might also be influenced by the valence of the descriptions.

Method

Participants

Seventy-four undergraduate students at the University of Colorado participated for partial fulfillment of course credit or monetary compensation. Equal numbers of these students were male and female.

Stimuli

Written self-descriptions. We developed eight different “self-descriptions” of three to four sentences each. Two of the descriptions were stereotypically-female and positively valenced (e.g., “As an elementary school teacher, I like to create an environment where students learn to cooperate and build self-confidence...”), two were stereotypically-female

and negative valenced (e.g., “Most of the time I stay home and watch soap operas ...”), two were stereotypically-male and positively valenced (e.g., “As an engineer, I thrive on solving challenging problems...”), and two were stereotypically-male and negatively valenced (e.g., “I have always been so immersed in my work that I never had the time to make any friends...”). Thus, there were 4 types of self-descriptions, varying in stereotypicality and valence.⁴

Voices. We selected a set of 20 female and a set of 20 male voices reading the “Rainbow Passage” from the voices scaled in Study 2.1. The mean femininity ratings for these voices (and standard deviations) were very similar to those obtained for the full set of voices in Study 2.1. To ensure that the gender of these voices was unambiguous, these female and male voices were intermixed and a new group of 30 participants was asked to categorize each voice as quickly as possible. The results showed that the 20 female voices were categorized as female 99% of the time and the 20 male voices were categorized as male 100% of the time.

Procedure

Upon arrival, participants were seated at individual computers with headphones and told that they would be involved in a probability judgment task with the aim of matching the speakers with the appropriate self-descriptions. They were further told that past research in visual perception showed people to be very accurate in their interpersonal judgments, and that through the present study, we were examining if this held true for judgments based on the voice. Participants then received one of the eight self-descriptions. They were asked to read the self-description carefully and to imagine what the person who wrote the self-description might sound like so that they could identify that person from among the set of 20 voices that would follow. After the participants had read the description, they either heard the set of 20 female voice clips or the set of 20 male voice clips.⁵ For each of the voice clips, they

were asked to estimate the probability, on a 0 to 100 scale, that the person speaking was the one who wrote the self-description. After completing the probability ratings for all 20 male or 20 female voice clips, the participants repeated the procedure with the other seven self-descriptions. Participants were asked to rate the set of 20 voice clips each time without letting previous ratings influence their judgments. The order of the voice clips and self-descriptions was randomized for each participant. After completing the ratings, participants were debriefed and dismissed.

Analysis

Each participant made probability ratings of the 20 male or female voices for each of the 8 self-descriptions that varied according to a 2 (Stereotypicality: female vs. male) by 2 (Valence: positive vs. negative) design. Of interest was the degree to which higher probabilities would be assigned to more feminine-sounding voices (within each gender) in the stereotypically female descriptions. Thus, our analytic approach involved a multilevel estimation procedure, estimating regression models within each participant and within each self-description (following Blair et al., 2002). In these regressions, each participant's probability ratings for the 20 voices for a particular self-description served as the criterion. These were regressed on one predictor variable (across the 20 voices): the femininity mean scale values of the voices from the pretest.

For each participant and for each self-description, these regressions yielded a slope for femininity that became the data for higher-level analyses. A positive slope means that probability ratings for a given self-description provided by a given participant increase as vocal femininity increases. A negative slope means they decrease. To determine if the effect of vocal femininity on probability judgments varied as a function of the valence and stereotypicality of the self-descriptions, we averaged the slopes for each participant across the two self-descriptions that had the same valence and the same stereotypicality values. These

average slopes (four for each participant) then became the data on which we subsequently performed analyses of variance with valence and stereotypicality of the self-descriptions as within-subjects factors.

Results

Table 2.3 presents the mean within-participant slopes for femininity, for each type of self-description and for both the female and male voices.⁶ All of these mean slopes are significantly different from zero, indicating that femininity of voice was used in making the probability ratings for each type of self-description: for the female-stereotypic self-descriptions, the higher the femininity of the voice, the higher the probability ratings; for the male-stereotypic self-descriptions, the higher the femininity, the lower the probability ratings.⁷

Female Voices

Unsurprisingly, given the fact that all slopes were significant and in the predicted directions, the slopes for the female voices showed a large difference as a function of the stereotypicality of the self-descriptions, $F(1, 38) = 193.75, p < .001$. The slopes also manifested a valence difference, $F(1, 38) = 9.13, p < .005$, such that more feminine voices were seen as more probable for the negatively-valenced self-descriptions than the positively-valenced self-descriptions. These differences were qualified by a stereotypicality by valence interaction, $F(1, 38) = 4.22, p < .05$. Simple effects tests revealed that the predicted stereotypicality difference was significant for both positive and negative self-descriptions, but it was somewhat larger in the latter case, $F(1, 38) = 108.65, p < .0001$, and $F(1, 38) = 135.55, p < .0001$, respectively.

Male Voices

In the case of the male voices, there was a large difference between the slopes for the male-stereotypic and female-stereotypic self descriptions, $F(1, 34) = 157.92, p < .0001$, again

with more feminine-sounding male voices rated as more probable for the female-stereotypic descriptions and more improbable for the male-stereotypic descriptions. There was neither a valence difference in these slopes, nor a valence by stereotypicality interaction.

Discussion

The data from this study strongly suggest that participants use differences in vocal femininity within voices of the same gender to make inferences about what speakers are like, with more feminine-sounding voices deemed more probable for female-stereotypic descriptions and less probable for male-stereotypic descriptions.

Although this basic effect strongly held for voices of both genders, there were some interesting (and unpredicted) gender differences. For the female voices, although more feminine voices were judged more probable for the female-stereotypic self-descriptions and less probable for the male-stereotypic ones, this difference was larger for the negatively-valenced descriptions than for the positively-valenced ones. No such interaction emerged for the male voices. Since this difference was not predicted, we hesitate to attach theoretical importance to it, pending its replication.

The bottom line from this study is that vocal femininity is used to make inferences about gender stereotypic attributes of speakers, even when gender does not vary and thus its salience is minimized. An obvious objection, however, to these results is that when decisions could not be made on the basis of gender category, the participants had no choice but to rely on differences in vocal femininity within a category to make their judgments. Indeed, in the categorization literature, there is evidence that when stimuli can be categorized, between-category differences are accentuated while within-category differences are minimized (Corneille, Klein, Lambert, & Judd, 2002; Ford & Stangor, 1992; Tajfel & Wilkes, 1963). In the context of the stereotyping literature, many dominant models of the link between categorization and stereotyping also suggest that category effects may be potent enough to

leave little room for within-gender variations to play a role in stereotyping (e.g., Bodenhausen & Macrae, 1998; Brewer, 1988; Brewer, 1996; Fiske & Neuberg, 1990). Consider, for instance, a quotation from an authoritative source (Fiske & Taylor, 1991): "Once a person is categorized as Black or White, male or female, young or old, the stereotypic content of the schema is likely to apply regardless of how much or how little the person looks like the typical category member" (p. 121). If we substitute "sounds like" for "looks like," this seems to leave little role for within gender differences when gender category also varies. So, if we want to show that vocal femininity is used even in a context when there are other useful cues that might be exclusively relied upon (such as the speaker's gender), then it is important to demonstrate the effects of (within-gender) vocal femininity on stereotypic inferences in a context in which judgments are asked of a mixed set of male and female voices. Our third study therefore examined whether stereotypic inferences would be made on the basis of within-gender vocal femininity even when the voices obviously differed in gender.

Study 2.3

Method

Participants

Thirty-eight undergraduate students at the University of Colorado participated for partial fulfillment of course credit. Of these students, 28 were female and 10 were male.

Materials and Procedure

The materials and procedure were the same as those used in Study 2.2 with two exceptions:

1) Each participant judged 30 voices in this study, 15 female and 15 male, for each self-description. These voices were selected from among those used in Study 2.2. As before,

each voice was easily identifiable as female or male, and within each gender, the voices varied in judged femininity.

2) Because the participants were asked to rate more voices in this study, only four self-descriptions were used rather than eight. Thus, there was only one self-description in each of the cells resulting from the crossing of gender stereotypicality with valence.

Analysis

Our questions that guided this study were slightly different from those of the previous study. First, we wanted to estimate the effects of gender categorization on stereotyping. For this purpose, we coded the gender of each voice using a contrast code, +1 if female, -1 if male. And then, in the within-participant regression models, we regressed the probability ratings of the 30 mixed gender voices on this contrast-coded predictor.

Second and more importantly, we wanted to see if even over and above gender category the judged femininity of the voice would affect stereotypic inferences. Accordingly, in the second set of within-participant regression models, the contrast-coded category variable and the mean scale values of femininity were simultaneously used as predictors.

Results

Presented in the top half of Table 2.4 are the mean within-participant slopes for gender category from the first model, where it alone was used to predict probability ratings. All category slopes were significantly different from zero and in the expected direction: Female voices were rated to be more probable for female-stereotypic self-descriptions and male voices more probable for male-stereotypic self-descriptions. Unsurprisingly these slopes showed a highly significant stereotypicality difference, $F(1, 37) = 198.92, p < .0001$. In addition, they manifested a difference due to valence, $F(1, 37) = 4.81, p < .05$, and a stereotypicality by valence interaction, $F(1, 37) = 37.30, p < .0001$. As shown in Table 2.4, gender category resulted in stereotypic inferences for both positive and negative self-

descriptions, but the influence was stronger in the latter, $F(1, 37) = 196.10, p < .0001$ than in the former, $F(1, 37) = 48.01, p < .0001$.

Mean partial slopes when both gender category and femininity were used to predict probability ratings for the self-descriptions are presented in the lower half of Table 2.4. The mean partial slopes for gender category remained very similar to what they were in the model that did not control for femininity. Specifically, they varied as a function of stereotypicality, $F(1, 37) = 116.67, p < .0001$, valence, $F(1, 37) = 4.94, p < .05$, and the interaction of stereotypicality and valence, $F(1, 37) = 57.70, p < .0001$.

In the case of femininity, over and above gender category, all slopes (with the exception of the slope for the negative male-stereotypic description) were significantly different from zero. These significant slopes indicate that vocal femininity continued to be used to make inferences within-gender, even when gender category was an available and influential cue for judgment. As expected, the femininity slopes varied significantly as a function of the descriptions' stereotypicality, $F(1, 37) = 75.48, p < .0001$, with more feminine sounding voices given higher probability ratings for the female-stereotypic self-descriptions, whereas the opposite was the case for the male-stereotypic ones, controlling for gender category. The significant interaction of stereotypicality and valence, $F(1, 37) = 67.08, p < .0001$, suggested that the stereotypic inferences were stronger in the case of the positively valenced self-descriptions than in the case of the negatively valenced ones, $F(1, 37) = 134.05, p < .0001$, and $F(1, 37) = 13.71, p < .001$.⁸

Discussion

In this study, where voices differed in gender, one might have supposed, following the prevailing models of stereotyping discussed earlier, that categorization by gender would be the only factor that affected stereotypic inferences. This was definitely not the case: While male and female voices were clearly judged differently, the effect of vocal femininity within

gender remained very potent. Within gender, those voices with higher femininity were given higher probability ratings in the female-stereotypic self-descriptions (and lower probabilities in the male-stereotypic ones), and this was true even while the gender of the voice exerted a large and significant effect on stereotypic inferences. Clearly gender stereotyping by voice is not based exclusively on the gender category of the target individual: Vocal femininity within gender continues to have a large impact on the types of trait inferences that are made.

We did not make predictions about differences in the role of either gender or vocal femininity as a function of the valence of the self-descriptions. Researchers have suggested that women may be viewed both positively and negatively, depending on the context and specific attributes (Fiske et al., 1999; Glick & Fiske, 1996; Glick et al., 1997). The inconsistent results with regard to evaluation across the studies may reflect such ambivalence. In Study 2.2, women's vocal femininity resulted in stronger stereotypic inferences for negatively than positively valenced self-descriptions; men's vocal femininity did not show such valence effects. In Study 2.3, stereotypic inferences based on gender category were stronger for negatively than positively valenced self-descriptions, but stereotypic inferences based on femininity, controlling for category, were stronger for positively than negatively valenced self-descriptions. Since the voices and self-descriptions varied across the studies, systematic research is needed to establish which effects are replicable and theoretically important. At present, it is difficult to draw any conclusions on the valence findings except to say that stereotyping based on vocal cues may be influenced by the valence of the attributes. More importantly, despite these variations, both studies showed that vocal femininity was a strong predictor of stereotypic inferences, and in both studies this was true for both positive and negative attributes.

Acoustic Characteristics, Perceived Femininity, and Stereotypic Inferences

Our findings as a whole led us to posit a theoretical model about the joint relationships among the acoustic characteristics (measured in Study 2.1), perceived femininity, and stereotypic inferences made on the basis of the voices. In Study 2.3 we showed that over and above gender category, stereotypic inferences were made on the basis of vocal femininity. Ultimately we would argue that it is the acoustic characteristics of the voices that give rise to perceived femininity, which in turn gives rise to the stereotypic inferences we have documented. Accordingly, it seemed to us that an appropriate mediational model was one in which, controlling for gender category, acoustic characteristics affect stereotypic inferences through perceived femininity. The model that we would suggest is diagrammed in Figure 2.1. In the top half of the figure, we represent the total effect of the acoustic characteristics on stereotypic inferences. Then in the bottom, we present the mediated model, with part of the total effect of the acoustic characteristics on stereotypic inferences mediated via perceived femininity.

To demonstrate support for the proposed mediational model, a series of conditions must be met (Baron & Kenny, 1986; Judd & Kenny, 1981). First we must show that the acoustic characteristics have an effect on stereotypic inferences, over and above gender category (effect A in the model at the top of Figure 2.1). Second, it is important to show that these characteristics also affect perceived femininity, again controlling for gender category (effect B in the mediational model at the bottom of the Figure). Third, controlling for both category and acoustic characteristics, it should be the case that perceived femininity affects stereotypic inferences (effect C in the mediational model). And finally, the direct effect of the acoustic characteristics on stereotypic inferences should be reduced once we control for perceived femininity (i.e., the residual effect D in the mediational model should be smaller than the total effect A). Using the data from Study 2.3, in which probability judgments were

made about 15 female and 15 male voices in the four kinds of self-descriptions, we examined these four conditions, thus testing the mediation model of Figure 2.1.

The assessment of mediation in the present case is somewhat complicated by the multilevel nature of the data, requiring some models to be estimated within-participants and some not. Accordingly, to simplify things as much as possible, we computed for each participant and for each of the 30 voices a stereotypic inference score: the degree to which a voice was given a higher probability by that participant in the two female-stereotypic self-descriptions than in the two male-stereotypic self-descriptions. Thus, rather than presenting four different mediational models, one for the probabilities given for each of the four self-descriptions in Study 2.3, we present only one model, with the stereotypic-inference dependent variable representing the extent to which a given voice is judged more probable in the female than the male self-descriptions.

To assess the total effects of acoustic characteristics on stereotypic inferences (effect A), we estimated two models for each participant, one regressing the stereotypic-inference score only on the gender category of each voice and the second including both gender category and the four acoustic characteristics as predictors. We then computed the difference in the *R*-squares between these two models (for each participant)⁹ and then tested (across participants) whether the mean increment to *R*-square, as a function of adding the acoustic characteristics, was significant. The resulting mean increment to *R*-square was .131 (reported in parentheses under effect A in Figure 2.1), a value that differed significantly from zero, $p < .0001$. Thus, the acoustic characteristics show an overall effect on stereotypic inferences over and above gender category.

Next, we wanted to demonstrate that for the voices used in Study 2.3, the acoustic characteristics predicted vocal femininity (as determined in the pretest), again over and above category membership (effect B in Figure 2.1). Across the 30 voices used in Study 2.3, we

estimated two models, the first only using category to predict vocal femininity and the second adding the four acoustic characteristics. We then computed the increment to *R*-square produced by adding the four acoustic characteristics. The resulting increment to *R*-square equaled .48 (reported in parentheses under effect B in Figure 2.1) and was again significantly different from zero, $p < .01$. Thus acoustic characteristics were largely responsible for judged femininity within gender category.

Then, we asked whether vocal femininity predicted stereotypic inferences over and above the acoustic characteristics (effect C in Figure 2.1), controlling for gender category. These analyses were again done within each participant. First we regressed the stereotypic-inference difference score on the voice's gender category and acoustic characteristics. Next we added femininity to this model. We then calculated for each participant the increment to *R*-square as we moved from the first model to the second. The mean increment equaled .064 (reported in parentheses under effect C in Figure 2.1), and was again significantly different from zero across participants, $p < .0001$). Thus over and above the acoustic characteristics (and controlling for gender category), higher femininity is associated with more stereotypic inferences.

Finally, we needed to demonstrate that the acoustic characteristics had a smaller impact on stereotypic inferences once we controlled for vocal femininity compared to their impact when femininity was not controlled (i.e., that effect D is smaller than effect A in Figure 2.1). We have already estimated the effects of the acoustic characteristics on stereotypic inferences not controlling for femininity (effect A of Figure 2.1). Now we need to estimate the contribution of the acoustic characteristics on stereotypic inferences once perceived femininity is controlled (effect D). Accordingly we computed for each participant the increment to *R*-square when we moved from a model that used gender category and perceived femininity as predictors of the stereotypic-inference difference score to a model

Chapter 2

What The Voice Reveals:

Within And Between Category Stereotyping on the Basis of Voice¹

Voice and Stereotyping

People have a fundamental need to make sense of the world quickly and efficiently (Fiske, 2004). As a consequence, one of the first things people do upon encountering a stranger is to use salient cues as a basis for making stereotypic inferences about that person (e.g., Bodenhausen & Macrae, 1998; Brewer, 1988; Fiske & Neuberg, 1990). Although social psychology has predominantly focused on salient visual cues associated with race, gender, and age, as a basis for categorization, the cues afforded by the voice may be equally salient in social settings, as these are some of the first cues that are encountered when meeting another person. Some of the most important of these may be those that signal the gender of the speaker because of inherent biological differences between men and women (Fitch & Giedd, 1999). Although we know of no study that examines the relationship between voice and stereotyping, it stands to reason that when people hear a voice, vocal cues signaling gender would be used to categorize the speaker's gender and then this category information would lead to gender-stereotypic inferences. Furthermore, because these same cues also vary within gender, a more interesting possibility is that these cues may also affect gender stereotypic inferences among individuals of the same gender. In other words, stereotypic inferences may be associated with speakers depending on the variance in vocal cues, over and above any stereotyping based on gender alone. The present research was conducted to examine this possibility.

Defining Vocal Cues

We will refer to the set of vocal cues that define gender as the femininity/masculinity of the voice. The physical distinction between men and women (Fitch & Giedd, 1999)

that included these two plus the acoustic characteristics. Across participants, the resulting mean increment to *R*-square was .104 (reported in parentheses below effect D in Figure 2.1), again significantly different from zero, $p < .0001$.

For each participant we could then compute the difference between the two increments to *R*-square due to the acoustic characteristics, comparing the model where femininity was not controlled (effect A in Figure 2.1) with the model where femininity was controlled (effect D in Figure 2.1). On average, the increment to *R*-square due to the acoustic characteristics was .027 less when femininity was controlled compared to when femininity was not controlled (i.e., effect A minus effect D: .131 - .104) and, again across participants, this mean difference was significantly different from zero, $p < .05$). Accordingly, we can conclude that the residual direct effects of the acoustic characteristics on stereotypic inferences once femininity is controlled (effect D) are smaller than the total effects of those characteristics when femininity is not controlled (effect A).

In sum, these analyses provide consistent support for the mediational model we have suggested. The acoustic characteristics as a set affect stereotypic inferences about the voices in part because they affect judgments of vocal femininity, which in turn leads to stereotypic inferences. That said, however, it is important to recognize that the mediation we have examined is only partial. There remain substantial effects of the acoustic characteristics on stereotypic inferences even controlling for judged femininity (effect D is highly significant).

Chapter 3

A Voice in the Workplace:

How Vocal Cues Impact Judgments Related to Hiring Decisions¹⁰

Researchers have posited that gender discrimination arises when a job/position is perceived to be incongruent with the gender of the applicant (e.g., Cejka & Eagly, 1999; Eagly & Karau, 2002; Eagly et al., 1992; Glick, et al., 1988; Heilman, 1983, 2001; Kalin & Hodgins, 1984; Martinko & Gardner, 1983). Specifically, people hold beliefs about characteristics that are necessary to be successful in certain jobs (Heilman, 2001) and they also hold beliefs that one or the other gender is more likely to possess those necessary characteristics. To illustrate, researchers have found high consensus not only among the lay-public but also among corporate employees in the type of characteristics that embody the ideal, successful manager: Male stereotypic characteristics were perceived to be the prerequisite for an ideal manager (Deal & Stevenson, 1998; Heilman, 1989; Powell & Butterfield, 1979, 1989; Schein, 1973, 1975; Schein & Davidson, 1993). When the characteristics believed to be necessary for a given job are incongruent with the characteristics associated with a given applicant on the basis of his or her gender category, then discrimination is proposed to occur (Eagly & Karau, 2002; Heilman, 1983).

What is clear from this literature is that the characteristics believed to be necessary to be successful in the workplace—especially in the more desirable positions—are subsumed under the dimension of *competence* and hence this dimension is the major criterion when judging job applicants' suitability for hire (e.g., Bertrand & Hallock, 2001; Blau & Kahn, 2000; Eagly & Karau, 2002; Nesbitt & Penn, 2000; Powell et al., 2002; Schein, 1975, 1978). However, this is not to suggest that applicants are not judged on any other dimension. In fact, competence is one of two dimensions that is believed to underlie most, if not all, judgments of others (Bakan, 1966; Cejka & Eagly, 1999; Eagly & Karau, 2002; Eagly & Steffen, 1984;

Fiske, Cuddy, Glick, & Xu, 2002; Heilman, 2001; Judd, Hawkins, Yzerbyt, & Kahimia, 2005; Rudman & Glick, 1999): The other dimension is warmth, which is predominantly associated with service orientation while competence is predominantly associated with achievement orientation. There exists a body of research devoted to the interactive consequences of these two dimensions (and gender) for employees. For instance, the “backlash” effect (Rudman, 1998; Rudman & Glick, 1999; 2001) suggests that female employees who are obviously highly competent can be strongly disliked when seen as cold (or low on warmth), whereas there does not seem to be such consequences for equivalent male employees (also see Biernat, Manis, & Nelson, 1991; Biernat & Manis, 1994). The main message gleaned from this research is that the fundamental affect of the warmth dimension is on valence (i.e., likeability) judgments rather than hireability judgments because warmth has more to do with how well the target may get along with colleagues and less to do with whether the target has potential to attain job success. Indeed, Rudman and Glick (1999) found that even when the manipulated job description clearly stressed that the ideal job applicant required characteristics high on warmth (“feminized job description”), competent applicants—both male and female—were still rated as more suitable for hire than warm applicants.

Our focus in the current research was on understanding how vocal cues affect perceptions of applicants’ characteristics that have the most impact on hiring decisions. Hence, we were predominantly interested in the dimension of competence. Specifically, the importance of competence as a criterion for judgments of job and applicant congruency coupled with the fact that recruiters rely heavily on interviews to make final applicant selection decisions, lead us to hypothesize that the physical cues of voice might specifically influence competence judgments in job interview contexts. Consequently, support for our hypothesis would provide the first evidence of a link between vocal femininity cues and

specific characteristics that form an important basis upon which discrimination in the workplace can occur.

The present studies extends the work presented in Chapter 2 by examining the interplay of vocal and other cues, for instance those taken from a résumé, within an interview context. That is, our question is whether the judgments of target applicants are different as a function of their vocal cues even when abundant other information is also available

Study 3.1: Overview

For the purpose of examining how judgments of applicants are affected by vocal and résumé cues, we modified a research paradigm used by Glick et al. (1988). In the real world, hiring processes generally involve considering more than one applicant to fill a position, thus the comparisons that recruiters make *between* applicants must play a vital role. Paradoxically, past research using individuating information (i.e., résumés) to examine this recruitment process have predominantly used a design where each participant only rated one applicant résumé (e.g., Cohen & Bunker, 1975; Gerdes & Garber, 1983; Glick et al. 1988; Rudman & Glick, 1999), hence it is unclear whether such a design allows participants to make critical comparisons among applicants that would be analogous to those made by real recruiters. Accordingly, we had each participant evaluate all applicants rather than just one applicant, as was the case in Glick et al.'s paradigm. Applicants' background information was presented through résumé-type information, which depicted either a stereotypically masculine or feminine applicant. For applicant voices, we recorded male and female speakers posing as applicants and had their voices scaled on vocal femininity/masculinity. On the basis of this scaling, male and female applicants whose voices were rated as either high or low on femininity (for their gender) were chosen for the main part of the study, where vocal femininity was crossed with voice gender category and résumé sex type such that each participant heard a different voice-résumé pairing. This allowed us to assess the nature of the

separate and joint effects of vocal cues, résumés, and applicant gender on the judgments made of the applicants. The dependent variables were judgments of each applicant's levels of competence and warmth.

Like past research on individuating/résumé information (e.g., Beckett & Park, 1995; Glick et al., 1988; Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982; also see Davison & Burke, 2000), we expected résumé cues to affect judgments of applicants. More importantly for our purposes, we anticipated that vocal femininity cues would also influence judgments (Ko et al., 2006), while gender category, itself, may not. Our rationale for these predictions comes from recent work suggesting that people in western societies may have become quite good at controlling the more blatant category-based stereotyping but not the more subtle within-category based stereotyping, particularly in situations where people may be wary of appearing biased (Blair, Judd, & Chapleau, 2004; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Ko, Muller, Judd, & Stapel, 2007). Given the abundance of anti-discrimination laws in the corporate sector, the interview is one situation where recruiters are likely to be cautious about making blatant stereotypic judgments of applicants on the basis category membership. However, we predict that the more subtle within-category vocal femininity cues would impact applicant judgments because perceivers may not always be aware of or able to control the influences from these cues (see Blair, Judd, & Fallman, 2004).

Given the prominent role that competency plays in evaluations of applicant suitability and the degree to which interviews are relied upon to make such evaluations, we predicted that vocal femininity cues would be used to judge applicants on the dimension of competence. Specifically, we anticipated that variation in vocal femininity within each gender should impact perceivers' judgments such that applicants with relatively more masculine voices would be perceived as more competent than applicants with relatively more

feminine voices *regardless* of the applicants' gender or résumé. Such a finding would provide the first evidence for our assertion that within-gender vocal femininity effects persist even in the presence of other rich informational cues (i.e., résumés and gender category derived from voice).

Study 3.1: Stimuli Materials

Résumés: Construction and Selection

We constructed 24 sex-typed résumés, which were designed to depict either masculine or feminine applicants. The résumés' sex-type was manipulated by varying the three pieces of information that made up each résumé. These three pieces of information were 1) a holiday job (e.g., masculine job: *salesperson at an auto supply store*, feminine job: *sales assistant at a flower shop*), 2) a job held during University term time (e.g., masculine job: *security patrol staff of a company*, feminine job: *an aerobics instructor*), and 3) a favorite hobby (e.g., masculine hobby: *playing basketball*, feminine hobby: *yoga*). The three pieces of information contained in each résumé were always stereotypically consistent with each other. In other words, if a résumé started with a masculine holiday job, then the term-time job and the favorite hobby were also masculine.

These résumés were then pretested for stereotypicality by 19 participants from the University of Groningen who rated them on a scale ranging from 1 (*very feminine*) to 7 (*very masculine*). Consequently, we chose six feminine sex-typed résumés ($M = 1.79$), and six masculine sex-typed résumés ($M = 6.25$).¹¹

Applicant Voices: Recording and Voice Scaling

In order to examine our central question of how the interplay of résumés and applicants' voice might influence judgments, we recorded male and female participants reading the résumés as if they were the actual applicants depicted in the résumés. Each of these participants also read the "Rainbow Passage" (Fairbanks, 1940 p.127; see also Ko et al.,

2006), which is a neutral passage commonly used in voice research and particularly useful in the present context because of its unvalenced content and neutrality with regard to gender stereotypes. This passage was used to scale the voices on femininity and the resulting scale values were used to select the voices used in the study.

Method

Recording Participants. Eighty-five students (45 female and 40 male) at the University of Groningen were recorded in exchange for monetary compensation. All were native Dutch speakers with an average age of 23 years ($SD = 2.4$).

Procedure

We recorded the participants in a soundproofed recording room. During the recording session, each participant was asked to read aloud the “Rainbow Passage”, and each of the 12 résumés. For each résumé, male participants always began by stating a different male name (e.g., “My name is Mark”) and female participants always began by stating a different female name (e.g., “My name is Ellen”). We had to discard 6 of the female and 2 of the male recordings because they contained too much background noise or too many reading errors. On the basis of the rainbow passage, the remaining 39 female and 38 male voice recordings were then scaled on vocal femininity.

Vocal Femininity Judgment Participants. Seventy-nine undergraduate students (72 female and 7 male) at the University of Groningen provided judgments of the vocal femininity of the recorded voices, in exchange for partial fulfillment of course credit. Participants were seated in front of individual computers with headphones and assigned to judge either the 38 male or 39 female voices speaking the Rainbow Passage. The order of the voices varied randomly across participants. The participants’ task was to rate each voice on how feminine it sounded, using a 1 (*not at all feminine*) to 7 (*very feminine*) scale.¹²

Voice Scaling Results

To examine the reliabilities of the mean femininity scores for each voice, the individual femininity ratings (for each voice by each participant) were analyzed with a two-way analysis of variance within the male and female voice sets, treating judges/participants as one factor and voice as another. From these analyses, variance components were estimated and intraclass correlations (due to voice, treating participants as a random factor; ICC 2.1 from Shrout & Fleiss, 1979) were calculated in order to estimate the reliabilities of the resulting vocal femininity values (following procedures outlined in Judd & McClelland, 1998). For the female voices, the mean femininity score was 4.56 ($SD = .55$) and the reliability of the individual voice means was .92. For the male voices the mean femininity score was 3.10 ($SD = .63$) and the reliability of the individual voice means was .97. As one would expect, the female voices were rated as significantly more feminine than the male voices, $F(1, 77) = 120.25, p < .001$.

On the basis of the vocal femininity judgments three female applicants high on vocal femininity ($M=5.56$), three female applicants low on vocal femininity ($M=3.33$), three male applicants high on vocal femininity ($M=4.9$), and three male applicants low on vocal femininity ($M=1.9$) were chosen to represent the final twelve applicant speakers for the main study.

Study 3.1: Ratings of Applicants

Our goal was to examine how job applicants would be judged on the dimensions of warmth and competence. Accordingly participants were presented with a series of hypothetical applicants, some in written form only (with their resumé and gender provided) and some orally (varying in resumé type, gender, and vocal femininity (within gender)). Each applicant was then judged by each participant on rating scales relevant to the dimensions of competence and warmth.

Method

Participants

Sixty-two undergraduate students (40 female and 22 male) at the University of Groningen participated for partial fulfillment of course credit or monetary compensation.

Research Design

Based on pretesting, as described above, we selected a total of 12 résumés—six stereotypically masculine and six stereotypically feminine. Participants in this study evaluated the applicants described by these résumés. Four of the résumés were judged in text form only, with participants making their judgments on the basis of reading the résumés. Eight of the résumés were associated with voices, such that each résumé was orally presented by a different speaker, the speaker ostensibly self-presenting him or herself. In the oral presentation, each participant was presented with two randomly selected voices out of the total three per voice type (e.g., male applicants low on femininity). However, across participants, all three voices per voice-type were presented. Which résumés were text-only and which were orally presented was counterbalanced across participants.

The six text-only résumés varied according to a 2 X 2 within-participant design, where the factors were the gender of the applicant to whom the résumé was attributed (male vs. female) and the sex-type of the résumé (masculine vs. feminine). The former factor was accomplished by varying the name at the top of the résumé. A further counterbalancing here was done such that each of the two stereotypically feminine résumés, for instance, was attributed to a male applicant for half the participants and to a female applicant for the other half of the participants.

For the 8 résumés that were orally presented, the design involved three within-participant factors—résumé sex-type (masculine vs. feminine), speaker gender (male vs. female) and vocal femininity (high vs. low within gender). Here we also did extensive

counterbalancing across participants so that each résumé was equally often presented by a male or a female and equally often by a high feminine voice (within gender) or a low feminine voice (again within gender).

Dependant Measures of Warmth and Competence

Each applicant, either presented in text only or orally, was rated by participants on eight gender stereotypic traits that captured the fundamental dimensions of warmth and competence. The positively-valenced masculine traits were *assertive* and *decisive*; the negative masculine traits were *cold* and *aggressive*; the feminine positive traits were *supportive* and *caring*; and the feminine negative traits were *insecure* and *passive*. Participants rated each applicant on all eight characteristics using a 1 (*not at all*) to 7 (*very*) scale. Following the definitions put forth in the above-mentioned literature on warmth and competence, the positive feminine and negative masculine stereotypic traits represent opposite ends of the warmth dimension (e.g., *caring* versus *cold*) and the negative feminine and positive masculine stereotypic traits represented opposite ends of the competence dimension (e.g., *decisive* versus *passive*). From these eight ratings of each stimulus résumé, we derived one overall competence composite rating by subtracting ratings on the negatively valenced feminine traits (*insecure* and *passive*) from the positively valenced masculine ones (*assertive* and *decisive*) and one overall warmth composite ratings by subtracting the ratings on the negatively valenced masculine traits (*cold* and *aggressive*) from the positively valenced feminine ones (*supportive* and *caring*).

Procedure

On arrival, participants were seated in front of individual computers with headphones. The computer automatically assigned each participant to the appropriate counterbalancing condition. All instructions were presented via the computer. The instructions informed the participants that the study was about the judgment process involved in job applicant

selection. Participants were further informed that our final goal was to make comparisons between how they (as lay public) judge applicants and how corporate professionals judge them. As a rationale for the text presentation of the résumés (as opposed to the oral presentation), participants were given the cover-story that a few of the audiotapes of applicant résumés turned out to be corrupt and, consequently, only the transcripts of the résumés would be presented for these applicants.

After the instructions, the computer randomly presented the 12 résumés one after the other, either spoken by applicant voices or in text format. After each résumé, participants rated the applicant on the eight traits. The order of these was randomized for each participant. While the ratings were being done for each applicant, either their spoken résumé looped or each text résumé stayed on the screen.

Results

Text Presentation

In order to examine how applicants' gender category (derived from name) and/or résumé information affected judgments, the composite ratings were subjected to a 2 (applicant gender category: female vs. male) X 2 (judgment dimension: warmth vs. competence) X 2 (résumé sex-type: feminine vs. masculine) within-participants ANOVA.¹³ Presented in Table 3.1 are the two dimensions' means and standard deviations, broken down by gender category and résumé type. There was a category main effect, $F(1, 61) = 4.97, p < .05$, due to higher average ratings given to female than male applicants, a judgment dimension main effect, $F(1, 61) = 4.84, p < .05$, due to higher average ratings on warmth than on competence, and a résumé sex-type main effect, $F(1, 61) = 35.53, p < .0001$, due to higher average ratings for feminine than masculine résumés. The only significant interaction was between judgment dimension and résumé, $F(1, 61) = 138.40, p < .0001$, showing that feminine résumés were rated higher on warmth while masculine one were rated higher on

competence. In sum, female applicants were rating more highly than males on both dimensions, while résumé sex type seemed to be matched with judgmental dimension, i.e., masculine résumé judged more positively on competence and female one judged more positively on warmth. The absence of a gender category by judgmental dimension interaction suggests that traditional gender stereotypes did not affect these applicant ratings.

Oral Presentation

The composite ratings for the oral presentation were subjected to a 2 (applicant gender: female versus male) X 2 (applicant vocal femininity: high feminine vs. low feminine) X 2 (judgment dimension: warmth vs. competence) X 2 (résumé sex-type: feminine vs. masculine) within-participants ANOVA.

Presented in Figure 3.1 are the average two dimension ratings for each résumé type. There were main effects of gender category, $F(1, 61) = 82.69, p < .0001$, due to higher ratings for female than male applicant, vocal femininity, $F(1, 61) = 33.09, p < .0001$, due to higher ratings for low than high feminine voices, judgment dimension, $F(1, 61) = 30.71, p < .0001$, due to higher ratings on warmth than competence, and résumé, $F(1, 61) = 38.92, p < .0001$, due to higher ratings for feminine than masculine résumés

Importantly, these main effects were moderated by higher order interactions. First, there was the expected vocal femininity by judgment dimension interaction, $F(1, 61) = 76.85, p < .0001$. For judgments of competence, applicants with low feminine voices (within gender) were rated as significantly more competent than those with high feminine voices, $F(1, 61) = 84.50, p < .0001$, whereas for judgments of warmth, applicants with high feminine voices were rated as significantly warmer than those with low feminine voices, $F(1, 61) = 8.43, p < .01$. In accordance with our predictions, however, the magnitude of the vocal femininity effect was much greater on competence than warmth judgments. Second, trait dimension interacted with résumé type, $F(1, 61) = 55.11, p < .0001$, such that on the warmth

dimension, applicants with feminine résumés were judged as significantly warmer than applicants with masculine résumés, $F(1, 61) = 91.72, p < .0001$, whereas these résumé types did not differentially affect judgments on the competence dimension, $F(1, 61) = 1.02, p > .30$. While the vocal femininity by judgment dimension interaction suggests stereotyping on the basis of within-gender vocal femininity, these data, like the text-presentation data, offer no evidence of stereotyping based on gender category, in that the gender category by judgmental dimension interaction was not significant.

Discussion

The goal of this study was to investigate how the voice might impact judgments in a job interview context where résumé information is concurrently available. The overall findings provide the first evidence suggesting a distinct informational basis from which people make competence and warmth judgments. Specifically, when vocal and résumé cues were simultaneously available, vocal cues were more strongly associated with judgments of competence than were résumé cues while résumé cues were more strongly associated with judgments of warmth than were vocal cues. Figure 3.2 provides a model of the distinct roles that vocal and résumé cues may play in job hiring contexts.

Gender category did not differentially impact the judgment dimensions, either in the text or oral presentations. The result in the oral presentation may be seen as somewhat surprising as past work shows that when salient category cues are present those cues are used for judgment (Beckett & Park, 1995; also see Blair et al., 2002; Ko et al., 2006). In the larger scheme of things, this may be an encouraging indication that discrimination is no longer such a problem in the workplace. However, the significant impact of applicants' vocal femininity cues on competence—the most critical criterion for judging applicant suitability—suggests otherwise. Specifically, masculine-voiced applicants were perceived to be more competent than feminine-voiced ones regardless of the applicants' gender category and resume

information. These results provide strong support for our assertion that in a work context, where discrimination is a sensitive issue, people may be motivated to avoid gender-stereotypic biases and the simplest way to do so may be to curb the use of category cues in judgments. Unfortunately, people are still vulnerable to biased judgments arising from within-category cues (i.e., vocal femininity). They may be less able to monitor the use of these cues, if they are even aware of their influence in the first place (Ko, Muller, Judd, & Stapel, 2006; Blair, Judd, & Fallman, 2004).

A secondary finding of interest was the difference in the effects of résumé type between the text and the oral presentations. That is, consistent with past research on the effects of individuating information (e.g., Glick et al., 1988; Locksley et al., 1980; Locksley et al., 1982), we found that written résumé information in the text presentation significantly affected both judgment dimensions in expected ways—for competence judgments, applicants with masculine résumés were rated as more competent than those with feminine résumés and *vice versa* for warmth judgments. In the oral presentation, however, where vocal cues were concurrently available, résumés no longer had any effect on competence judgments but solely affected warmth judgments. Given that competence may be more central than warmth in hiring decisions, this finding raises an interesting possibility that gender stereotypic information conveyed by résumés may be less impactful in the presence of vocal information than when they are the only information about an applicant that is available

Now that we have established that vocal cues and resumes lead to distinct judgments a necessary next step was to test more stringently the robustness of these novel results. Additionally, study 3.2 allowed us to clarify some of the issues raised in study 3.1. First, because we only included voices that were rated extremely on vocal femininity for their gender, it is unclear whether gender stereotyping on the basis of vocal femininity would emerge with voices that varied more continuously in their femininity within gender. Second,

it is possible that the absence of gender category stereotyping was due to the particular context (i.e., a job application) in which ratings were done. This is a context in which participants are sensitized to issues of discrimination. It therefore seemed likely that in other judgment contexts we might find evidence for both gender-based stereotyping and also stereotyping on the basis of within-gender vocal femininity. Third, we wanted to extend these results by providing individuating information that was directly relevant to the two judgmental dimensions to examine whether the vocal femininity effects would continue to be found even in the presence of individuating information that was directly diagnostic of the two judgmental dimensions.

Study 3.2: Overview

For this study we modified a version of Locksley et al.'s (1982) paradigm (also see, Beckett & Park, 1995) where participants had to predict how targets would have behaved in a particular scenario after learning about targets' behaviors in other diagnostically relevant scenarios (individuating information). All the behaviors were directly related to either the dimension of competence or warmth. Furthermore, the individuating information was presented by target voices that varied both gender category and (within-gender) vocal femininity.

In Study 3.1 gender discrimination could have been a salient and sensitive issue because hiring decisions have long been associated with such biases. Hence, participants may have been particularly motivated to avoid using stereotypes on the basis of category membership. In contrast, gender discrimination should be a less salient concern in the current experiment because the experimental task is one that is not that makes salient gender biases. Accordingly, we expected gender category to impact competence such that, on average, male speakers would be rated as more likely to behave competently than female speakers. However, given past work on feature effects within categories (e.g., Blair et al., 2002; Ko et

al., 2006) coupled with our findings from study 3.1, we predicted that vocal femininity would be a potent enough cue to affect competence predictions over and above the influence coming from gender category and individuating information.

Study 3.2: Stimuli Materials

Competence and Warmth Behavioral Scenarios

We derived five scenarios in which the behavior of stimulus persons would be seen to indicate their competence (competence condition) and another five scenarios in which the behavior of stimulus persons would be seen to indicate their warmth (warmth condition).

Recording Participants

Seventy-one students (35 female and 36 male) at the University of Groningen were recorded in exchange for monetary compensation. All were native Dutch speakers with an average age of 23 years ($SD = 2.90$).

We recorded the participants in a quiet room using a head-worn microphone. Each participant was asked to read aloud the “Rainbow Passage” and the 16 possible outcome combinations of competent and incompetent responses in the competence condition’s first four scenarios and the 16 possible outcome combinations of warmth and cold responses in the warmth condition’s first four scenarios.

Voice Judgments, Scaling, and Results

Sixty undergraduate students (44 female and 16 male) at the University of Groningen provided vocal femininity judgments of the recorded voices, in exchange for partial fulfillment of course credit or monetary compensation. The procedure and methods for the judgments and scaling of these voices were identical to those described in Study 3.1 above. For the female voices, the mean femininity score was 4.52 ($SD = .89$) and the reliability of the means was .94. For the male voices the mean femininity score was 3.18 ($SD = .93$) and the reliability of the means was .93. Again, the female voices were rated as significantly more

feminine than the male voices, $F(1, 69) = 38.63, p < .0001$. On the basis of the scaling, we chose a set of 16 female voices and a set of 16 male voices that maximized the variance on femininity for the main part of experiment 3.2. The mean femininity ratings for these voices (and standard deviations) were very similar to those obtained for the full set of voices.

Study 3.2: Behavioral Predictions

Methods

Participants

Sixty-six undergraduate students at the University of Groningen participated for partial fulfillment of course credit or monetary compensation. Half of them (25 female and 8 male) were randomly assigned to the competence condition and the other half (22 female and 11 male) were randomly assigned to the warmth condition.

Procedure

On arrival, participants were seated in front of individual computers with headphones. All instructions were presented via the computer. The instructions informed the participants that the study was designed to investigate whether given knowledge of a person's prior behavior, people are able to correctly predict the person's subsequent behavior. Accordingly, participants would be listening to 32 speakers who participated in a series of five experiments (i.e., scenarios) intended to measure either, depending on the condition, competent behavior or warmth behavior. Specifically, the speakers would be relating how they responded in the first four scenarios and the participant's task was to predict how the speakers responded in the fifth. The participants were told that their predictions would be compared with the speakers' actual behavior in the fifth scenario. Following these instructions, participants read the description of each of the five scenarios relevant to their condition and for each were concretely told the behavior that was considered a competent (warm) response and the behavior that was considered an incompetent (not warm/cold) response (see Appendix A).

Participants then listened to all 32 male and female speakers and for each speaker, they had to estimate the probability that the speaker behaved competently (warmly) in the fifth scenario. In each condition, the first four scenarios provided 16 possible outcome combinations (i.e., combinations of warm and cold behaviors or competent and incompetent behaviors) and prior to the study, per condition, the set of 16 speakers in each gender category was randomly paired with the 16 outcome combinations. The order of the speakers was randomized for each participant. After completing the ratings, participants were debriefed and dismissed.

Analysis and Results

We first estimated the effect of individuating information alone on the competence and warmth judgments. For this purpose, we conducted separate regression analysis per participant where the likelihood of competent (warm) behavior in the fifth scenario for the 32 speakers was regressed onto the proportion of competent (warm) behavior(s) in the first four scenarios. Next, in order to examine the critical question of whether gender category of the speaker and vocal femininity affected competence and warmth judgments, over and above the individuating information given about each target person in the four scenarios, we performed separate regressions per participant where the likelihood of competent (warm) behavior in the fifth scenario was regressed onto contrast-coded gender category, vocal femininity, and the proportion of competent (warm) behavior(s) in the first four scenarios.¹⁴ We then tested whether the resulting slopes from the first model and the three partial slopes from the second model differed by condition.¹⁵

Presented in the top half of Table 2 are the mean within-participant slopes from the first model, where individuating information was the sole predictor of the likelihood ratings. The slopes were significantly different from zero in the expected direction. That is, for the competence ratings, the more competent behaviors that target persons had manifested in the previous four scenarios, the more likely they were predicted to behave in the competent

manner in the fifth (mean R -square = .51). Likewise for the warmth ratings, the more warm behaviors that the target persons had manifested in the first four scenarios, the more likely they were predicted to behave warmly in the fifth (mean R -square = .60). Furthermore, the competence slopes did not differ from the warmth slopes, $F(1, 64) = 1.74, p < .20$, showing that the strong influence of individuating information did not differ by condition.

Presented in bottom half of Table 2 are the means of the partial slopes, broken down by condition. All partial slopes in the competence condition were significantly different from zero and in the expected direction. That is, the more competent behaviors that targets manifested in the previous four scenarios, the more likely they were predicted to behave competently in the fifth; male targets were predicted to behave more competently, controlling for their individuating information (i.e., behavior in the previous four scenarios; and the less feminine the voice, the more competent the prediction, controlling both for individuating information and gender category. On the other hand, in the warmth condition, only the partial slopes for individuating information was significantly different from zero. Testing the between condition differences in these slopes revealed, as expected, that the effect of gender category on judgments in the fifth scenario differed significantly by condition, $F(1, 64) = 7.27, p < .01$, such that in the competence condition, male speakers were rated as more likely to behave competently than female speakers in the fifth scenario, controlling for individuating information and vocal femininity, whereas category had no effect on the likelihood ratings in the warmth judgments. Additionally and importantly, the partial slopes for vocal femininity differed significantly between the two conditions, $F(1, 64) = 4.70, p < .05$, such that in the competence condition, less vocal femininity was associated with higher competence predictions, controlling for gender category and individuating information whereas vocal femininity had no affect on warmth behavioral predictions.

Discussion

This study allowed a more rigorous and controlled test of the key findings and predictions from Study 3.1. Specifically, we used a well-established experimental paradigm first developed by Locksley et al., (1982, Experiment 2), where the diagnostic individuating information has consistently been shown to exert a strong influence on judgments. For our purpose, this provided an even stronger test vocal cue effects than were afforded by résumé information. For the vocal stimuli, we selected voices that represented all ranges of vocal femininity, hence, countering the possibility that the vocal femininity effect in the first study resulted from femininity functioning as a categorical cue.

Even under these more stringent conditions, our primary predictions were supported. That is, regardless of the substantial influence of individuating information, both gender category and within-category vocal femininity of the speakers affected behavioral predictions in the competence condition, whereas in the warmth condition, no such effects were observed. Specifically, speakers on the basis of their voice category were predicted to differ on how competently they behaved in the fifth scenario. Even more remarkable, however, was the confirmation of our prediction about vocal femininity—even over and above the potent effects of individuating information and category, vocal femininity still affected competence but not warmth predictions. The pattern of these effects replicated the one found in Study 3.1 by showing that speakers whose voices sound more masculine /less feminine were more likely to be rated as behaving competently, *regardless* of the speakers' gender and diagnostic information.

In the first study we found that the individuating information contained in the résumés had a greater impact on warmth judgments than on competence judgments. That effect was not replicated here, since the individuating information conveyed by behaviors in the first four scenarios had a large impact on both competence and warmth predictions in the fifth

scenario. It is possible that these effects are context dependent such that in specific contexts of job interviews where judgments on both dimensions are predominantly related to the workplace (i.e., competence is related to job success and warmth is related to being a good colleague) résumés may be used more to make warmth than competence judgments because vocal cues may be perceived to be more diagnostic of job success. Hence, vocal cues may exert a powerful enough influence on competence judgments to leave little room for résumés to have an effect on this judgment dimension. In other contexts, such as those depicted in our experimental scenarios, where judgments on the two dimensions are no longer work dependent, vocal cues may not exert as powerful an influence on competence judgments leaving room for individuating information to affect this judgment dimension while concurrently affecting the warmth dimension. Further work needs to be conducted in order to examine this idea.

Chapter 4

Sneaking in Through the Back Door:

How Category-Based Stereotype Suppression Leads to Rebound in Feature-based Effects¹⁶

As discussed in the introduction to this dissertation, there are clear indications that people today are quite capable of curbing relatively blatant category-based stereotypes but not the relatively subtle feature-based ones. However, to date no research has examined the question of whether an actual relationship between category- and feature-based stereotyping exists such that the constant societal pressure to suppress category-based stereotyping may ironically lead to a rebound in feature-based stereotyping. We argue that an examination of this topic is particularly timely and important, given the dual presence of strong norms in our society not to discriminate and the nevertheless strong and persistent experience that discrimination is alive and well. Perhaps the norms that dictate against category-based discrimination actually lead to stronger feature-based discrimination and subjective feelings that minority targets are the objects of discrimination persist in spite of those norms.

The present research investigated the relationship between category- and feature-based stereotyping within the auditory domain. In particular, we examined whether suppressing the use of gender-category stereotypes in one task affected stereotyping due to variations in within-gender vocal femininity in another task.

Study 4.1

Method

Stimuli For Probability

Self-descriptions. The four descriptions developed by Ko, Judd, and Blair (2006, Study 3) were used in this study. These varied on the degree to which the depicted person had gender stereotypic attributes and the degree to which the person was likeable (e.g., the

positively valenced male stereotypic description depicted an engineer who thrives on solving challenging problems at work).¹⁷

Recorded Voices. 45 female and 40 male voices saying the “Rainbow Passage” (Fairbanks, 1940, p. 127) were scaled on vocal femininity.¹⁸ Based on the results, 15 male and 15 female voices were chosen.

Participants

Thirty-eight male and 138 female University of Groningen students participated in exchange for course credit or monetary compensation.

Procedure

Participants were told that they would complete a number of short unrelated tasks. The first task was explained as an investigation of patterns in how perceivers write about others. Accordingly, participants were told that they would be given 10 minutes to write about a day in the life of a male and female target depicted in two photographs. Furthermore, those in the suppression condition were told that they should avoid gender stereotypes and be careful that these not affect what they wrote. Those in the control condition were not given these extra instructions. Participants were then put in separate cubicles with an envelope containing the first target photograph and a sheet of writing paper. Those in the suppression condition had written reinforcement to suppress gender stereotypes at the top of their writing paper. The gender of the first target randomly varied between participants. After 5 minutes, participants were given a second envelope, with a photograph of the second target, a new sheet of writing paper, and another 5 minutes to write. Following this writing task, participants engaged in a probability judgment task, similar to that used by Ko, Judd, & Blair (2006). This task was presented as an interpersonal accuracy task, where they had to match the self-description of a target with the correct speaker. Participants were randomly assigned to read one of the 4 self-descriptions before listening to the 30 voices, presented in random

order. For each, they estimated the probability, on a 0% to 100% scale, that the person speaking was the one who wrote the self-description. Since any one of the voices could be the correct speaker, participants were allowed to assign the same probability value to more than one speaker. After the probability task, participants completed a final unrelated questionnaire.

Analysis

We were interested in whether the variations in the relation between the probability ratings with gender category and within-gender vocal femininity were moderated by the suppression manipulation. This involved a multilevel modeling procedure, estimating a separate regression model for each participant (see Blair et al., 2002; Ko, Judd, & Blair, 2006). In these regressions, participants' probability ratings for the 30 voices were regressed onto two predictors: (a) contrast-coded gender category (+1 if female, -1 if male) and (b) degree of vocal femininity. The resulting two partial regression slopes per participant—one assessing the partial relationship between category and probability ratings and another assessing the partial relationship between vocal femininity and probability ratings—became the two measures which we subsequently analyzed. These partial slopes were examined in the context of a 2 (self-description stereotypicality: female vs. male) X 2 (self-description valence: positive vs. negative) X 2 (Condition: Suppression vs. Control) between-participant factors.

Results

Category Use

For the measure of category use, positive category slopes mean higher probabilities were assigned to female voices and negative category slopes mean higher probabilities were assigned to male voices, controlling for vocal femininity. Presented in Figure 1 are the mean category slopes, broken down by condition and self-description stereotypicality.

Unsurprisingly, there was a stereotypicality main effect, $F(1,168) = 180, p < .001$, indicating

that the mean category slopes between female stereotypical self-descriptions and male stereotypical ones were significantly different. More interestingly, this effect was moderated by the suppression manipulation, $F(1,168) = 6.97, p < .01$. In other words, participants who were told to avoid using gender stereotypes in the first writing task (suppression condition) used gender category less in their subsequent stereotypic inferences than participants in the control condition.^{19 20}

Vocal Femininity Use

For the measure of vocal femininity use, positive femininity slopes mean higher probabilities with increases in vocal femininity and negative slopes mean lower probabilities with increases in vocal femininity, controlling for category. Presented in Figure 2 are the mean femininity slopes, broken down by condition and self-description stereotypicality. Again, this measure yielded a stereotypicality main effect, $F(1,168) = 166, p < .001$, due to the femininity slopes for female and male stereotypic self-descriptions being significantly different. More central to our reasoning was the predicted stereotypicality by condition interaction, $F(1,168) = 5.34, p < .023$. Participants who were told to suppress gender stereotypes in the first writing task subsequently made more use of vocal femininity in their stereotypic inferences than participants in the control condition.²¹

Complementary analyses

As can be seen in Figures 1 and 2, results observed on the category and femininity slopes at the mean level seem to suggest opposite effects of our suppression manipulation: Suppression decreased category use but increased femininity use. As a way to investigate whether there was a relationship between the use of category and femininity, we computed partial correlations (controlling for stereotypicality and valence) between these two measures. These analyses revealed no correlation in the control condition, ($r = .10, p < .38$). However,

in the suppression condition, the less participants used category gender, the more they used vocal femininity features ($r = -.33, p < .01$).

Discussion

The goal of the current research was to investigate how suppressing category-based stereotypes might affect stereotyping due to variations in within-category cues. First, consistent with our reasoning that people may have become quite practiced at suppressing category-based stereotypes and that categorical cues are relatively simple to monitor, we found that participants who were told to suppress category-based stereotypes in the writing task continued to do so in the probability task. Second, we found strong support for our hypothesis that such suppression heightened the use of within-category cues as a basis for stereotyping. This led us to examine more closely the relation between category- and feature-based stereotyping. In support of our reasoning, the correlations revealed that when participants were told to suppress the use of stereotypes (i.e., in the suppression condition) a relation existed such that the less these participants used category cues the more they used within-category cues in their judgments.

The research reported earlier by Blair and colleagues importantly outlined the ways in which stereotyping could be based not only on social categories but also on features associated with category membership but that also vary within those categories (Blair et al., 2002; Ko, Judd, & Blair, 2006). They also demonstrated that these feature-based stereotyping effects were relatively subtle in the sense that participants did not seem to be able to suppress them when instructed to do so (Blair et al., 2004). The present set of results importantly extend these conclusions by examining the relationship between the two sorts of stereotyping, category-based and feature-based, and how efforts to suppress the former, while successful, may actually increase feature-based stereotyping. We would suggest that these results are particularly timely, given the widely-shared and largely effective norms that encourage

people to avoid the expression of category-based stereotypes. Ironically, our work suggests that the adoption of such norm-prescribed behavior may end up increasing the extent to which one stereotypes others based on those features that are associated with category membership but that vary within categories.

A number of novel aspects of this research set it apart from past work on suppression and rebound. First, this is the first demonstration of cues involved in suppression (i.e., between-category cues) being different from the cues involved in rebound (i.e., within-category features rather than the same between-category cues involved in suppression). Second, we showed rebound effects in a totally different physical modality from the one in which suppression was introduced. That is, we manipulated the suppression instructions through the visual modality using photographs and a writing task, whereas we presented cues and observed rebound in the auditory modality (i.e., as a function of listening to voices). Some may think that Macrae et al. (Experiment 2, 1994) were the first to demonstrate suppression and rebound in two different modalities because their suppression task involved writing about a target skinhead depicted in a photograph whereas rebound was measured behaviorally. However, we would argue that this is not entirely accurate since both the stimuli used in their suppression task (i.e., photographs) and rebound task (i.e., jacket and bag supposedly belonging to the target skinhead) were visual ones. Hence, in essence, the cues used to trigger both suppression and rebound were of the same visual modality.

Chapter 5

General Discussion and Conclusions

The goal of the present dissertation was to examine the everyday snap judgments that people make in a variety of situations and contexts on the basis of speakers' voice. We were particularly interested in whether the vocal cues that vary within each gender category have the potential to influence such judgments. We examined this through a program of research that was presented in chapters 2 to 4. In this final chapter, I summarize and provide some discussion of the main findings detailed in those chapters.

Overview of the Main Findings

Chapter 2

What The Voice Reveals:

Within And Between Category Stereotyping on the Basis of Voice

As a necessary first step in my research program, we examined whether listeners' actually do use gender stereotypes on the basis of voice. Furthermore, we went beyond vocal differences between gender categories and also looked at how differences in vocal cues/features within each gender category impacts stereotyping. Specifically, voices within each gender vary in the degree to which they sound feminine/masculine and our results showed that people's perception of this vocal femininity is remarkably consensual. Most central to our purpose was the finding that vocal femininity affected stereotypic inferences over and above those due to between-category differences: Voices that sounded more feminine/less masculine were more associated with traits stereotypic of females and those that sounded less feminine/more masculine were more associated with traits stereotypic of males *regardless* of the speakers' gender. We also investigated some of the voices' acoustical cues related to these inferences and found that the cues—such as pitch, formant frequency, and variability pitch—which distinguish men and women's voices (i.e., between category

differences) are also the ones that featured importantly in the perception of within-gender vocal femininity. Finally, mediation analyses showed that the relationship between acoustics and stereotyping was in part due to femininity.

Chapter 3

A Voice in the Workplace:

How Vocal Cues Impact Judgments Related to Hiring Decisions

In Chapter 3, we examined some of the implications of vocal stereotyping in a corporate context where sex discrimination to this day leads to unequal job opportunities. We were specifically interested in the evaluative processes involved in the interview stage where recruiters are provided with many sources of information about the job applicants.

Understanding vocal effects during this stage seems particularly important given that recruiters place a great deal of emphasis on interviews and since most, if not all, hiring decisions rely upon some form of interview. Accordingly, in Study 3.1 we recorded several male and female speakers posing as job applicants reading different sex-typed résumés.

Using vocal femininity as a criterion, a subset of these voices was chosen and presented to participants. Each participant heard a different voice-résumé pairing, allowing us to assess the nature of the separate and joint effects of vocal cues, résumés, and applicant gender on the participants' judgments of applicants. Interestingly, vocal femininity cues almost exclusively affected judgments of applicants' competence such that those with masculine voices were perceived to be significantly more competent than those with feminine voices, irrespective of the applicants' gender, whereas résumés solely affected judgments of applicants' warmth such that those with stereotypically feminine résumés were perceived to be warmer than those with stereotypically masculine résumés. The potent effect of vocal femininity on competence was replicated in a second study under more stringent and controlled conditions.

Chapter 4

Sneaking in Through the Back Door:

How Category-Based Stereotype Suppression Leads to Rebound in Feature-based Effects

In this Chapter, we examined the relation between category- and feature-based stereotyping. Consistent with our reasoning that people may have become quite practiced at suppressing category-based stereotypes and that categorical cues are relatively simple to monitor, participants who were told to suppress category-based stereotypes in the writing task continued to do so in the probability task. Even more important was the predicted finding that such suppression simultaneously heightened the use of within-category vocal femininity cues as a basis for stereotyping. Partial correlations examining the relation between category and vocal femininity in the probability task (partialling out the effects of the self-descriptions' stereotypicality and valence) no relation in the control condition but a strong negative relation in the suppression condition showing that in this latter condition, the less participants made use of category the more they used vocal femininity in their stereotypic inferences.

General Discussion

It makes sense that cues about a person's gender category, be they visual or auditory or socially defined, can affect gender stereotyping through categorization: Those individuals who are judged to be of one gender are assumed to have attributes different from those who are judged to be of the other gender. The research in Chapter 2 provides the first evidence of the effects of gender-signaling vocal cue on perceivers' stereotypes. However, the main interest that drove our research was not whether perceivers stereotype on the basis of gender but whether the gender-signaling vocal cues that perceivers normally use to make gender categorization judgments may lead to gender stereotypic inferences even within category. We proposed that variations in gender-signaling cues, which we defined as vocal femininity, may

also influence gender stereotyping, over and above the more gross category distinction. To examine our reasoning, we first showed that perceivers have little trouble making highly consensual judgments of variation among voices of the same gender in vocal femininity (Study 2.1). Additionally, we examined some of the acoustic characteristics of voices that are associated with gender and that may also account for variations in judged femininity within gender. In our samples male voices were lower pitched, had more resonance, and were less variable in pitch than female voices. Within genders, pitch was consistently related to the judged femininity of voices and, within the female voices, resonance was as well. What is most interesting in our acoustic analyses was the finding that the same cues that distinguish voice category are also the cues that are related to judgments of vocal femininity *within* each voice category. In Study 2.2, we showed that participants spontaneously associate more feminine sounding voices with female stereotypic attributes even when the gender category of the stimulus voices does not vary. Importantly, these stereotypic effects, based on vocal femininity, were found for both samples of female and male voices. In Study 2.3, we found that within-gender stereotypic inferences on the basis of vocal femininity continue to be made even when gender varies and is clearly a salient cue. Participants in this study could easily have relied solely on the gender of voices in making their inferences, but they did not. Even though strong stereotypic inferences were in fact made on the basis of the gender category of a voice, within-gender stereotypic inferences also ensued from the degree of vocal femininity of a voice. Finally, we showed that the acoustic characteristics that led to perceptions of vocal femininity exerted their effects on within-gender stereotyping in part through those perceptions, supporting our mediation model.

Our work clearly demonstrates that the perceived femininity of voices varies reliably within both genders and is used to make gender-stereotypic inferences. This is true both when speakers do not vary in their gender and when they do. What is theoretically provocative

about the results in Chapter 2 is that they suggest that stereotyping is not exclusively category-based. Rather, stereotyping ensues both from category membership and, independently, from those perceptual cues that are indicative of category membership but also vary within those categories (see also Blair et al., 2002). Concretely, this means that independent of the gender stereotypes arising from a target's gender, the variations in targets' vocal femininity may also lead perceivers to make the same stereotypic inferences. In essence, we contend that those perceptual cues, which are used to establish the gender category of a person (i.e., vocal femininity), and lead to stereotypic inferences through categorization, have acquired the ability to elicit those same stereotypic inferences independent of categorization.

On the surface, our findings may not appear all that new in light of other research documenting that a variety of personality inferences are made on the basis of vocal and paralinguistic cues (e.g., Scherer, 1979). For instance, people who speak more quickly are seen as more competent (B. Smith, Brown, Strong, & Rencher, 1975) and have more persuasive influence (Miller, Maruyama, Beaver, & Valone, 1976). And such personality attributes are likely to be related to judgments of the femininity/masculinity of the person. However, at a deeper level, it seems theoretically important to differentiate between such personality inferences from vocal cues and stereotypic inferences made on the basis of the perceived femininity of the *voice* (rather than of the *person*). It is not that our participants are judging some individuals as more feminine than others. Rather they are attentive to cues that indicate the femininity of a voice and they then stereotype on the basis of those cues. Though having said this, we would not exclude the possibility that judgment of vocal femininity could have lead to judgments of trait femininity and that these trait femininity judgments could have played a role in our effects. Future research should examine this issue.

Some may think that our idea of within-gender variation in vocal femininity is more of a “reinvention” of theories that have been around for years. After all, from the seminal work of Rosch (see Mervis & Rosch, 1981 for a review) we have learned that people are quite capable of perceiving gradient structures of physical cues within categories. However, where it concerns stereotypic associations on the basis of these cues, dominant models of stereotyping have long presumed that these subtle gradient structures play no part in the stereotypic associations: Consider again the quote by Fiske and Taylor (1991) "Once a person is categorized as Black or White, male or female, young or old, the stereotypic content of the schema is likely to apply regardless of how much or how little the person looks like the typical category member" (p. 121). This point of view suggests that as long as exemplars within a given category are still unequivocally perceived to be part of that category, then only their category and *not* their goodness of fit to their category will be used as a basis for stereotypic inferences. Hence, questioning whether people process variations in cues within a gender was not the purpose of our research, because this is clearly well established in the literature. What we do question and try to address are the theories of how these within-category variations might lead to stereotypic inferences. In our research, voices were clearly categorized as those of male or female. Nonetheless, contrary to what leading theories of stereotyping suggests, the gradient structures of vocal femininity within gender did play a role in stereotypic inferences.

Most, if not all, judgments of others can be classified under the fundamental dimensions of competence and warmth (e.g., Fiske et al., 2002). However, established theories posit that judgments on the dimension of competence have the greatest potential to influence discrimination in the workplace because competent characteristics are believed to be prerequisites for job success and therefore competence is the dominant criterion upon which applicants are judged when making hiring/job suitability decisions. The question that

we set out to examine in Chapter 2 was whether applicants' vocal femininity cues specifically affects judgments on this dimension and consequently has the potential to impact discrimination. Our rationale comes from observing the strong emphasis placed on interviews to make final hiring decisions and the fact that throughout the job recruitment process, it is only during interviews that recruiters are exposed to applicants' vocal cues. The results from both studies confirmed our predictions about vocal femininity effects. That is, both in judgments of applicants (Study 3.1) and behavioral predictions (Study 3.2), participants specifically relied on vocal cues as a basis to judge the competence of others. One may ask why we targeted vocal cues among all the other physical cues afforded by interviews. The answer lies in the popularity with which telephone interviews are conducted today. In fact, telephone interviews may be the most common form of interview and under these circumstances, the only physical cue available is the applicants' voice. Yet, recruiters seem to have little trouble basing their hiring decisions on such interviews, suggesting that vocal cues may play a particularly influential role in job competence judgments. Taken together the result in Chapter 3 provide an exciting initial demonstration that something as subtle and yet prevalent as vocal cues continues to impact judgments in the face of other potentially competing information. In the end, it may not be *who* you are but *how you sound* that ultimately affects whether you get the job or not.

Affirmative action and equal opportunity laws are some examples of how greatly Western society has advanced in terms of attempting to neutralize category-based influences that may unfairly bias judgments. However, biases still exist. Our findings in Chapter 4 suggest that we can no longer rely exclusively on categorical differences to elucidate why this is the case. Instead, a more complete understanding of the biases that occur must necessarily include examinations of how the more complex yet subtle within-category cues exert their influences on judgment and behavior. Our research imparts perhaps an even more

sobering message; feature-based biases do more than merely exist—they can be exacerbated by norms (and legislation) that appropriately dampen category-based stereotyping and discrimination.

Final Remarks

Although these results, we think, are very compelling in suggesting how voices may impact gender stereotyping, we would not want to argue that vocal cues are pre-eminently important in social interaction. Nevertheless, even in situations where many other cues are available, such as résumés and rich individuating behavioral information (Study 3.1), vocal qualities still played a role. That is, the sorts of effects that we have documented in Chapter 2 definitely made a difference even when considerably richer and individuating information were available (Chapter 3). Consider, for instance, the famous public figure, Margaret Thatcher. It is well documented that she underwent extensive vocal training in order to “...lower the pitch of her voice and make her sound more authoritative” (Paton, March, 24, 2003); Thatcher and her advisors obviously believed that visual cues were not enough to portray the desired persona. In other words, they assumed that vocal qualities do make a difference even when much richer visual and verbal information is available.

Obviously, the voice is only one of several physical characteristics that perceivers can use to make judgments. However, the voice stands apart from all other physical characteristics because it is probably the most dynamic of them all: we can manipulate our voice at any chosen moment. We have all experienced hearing the change in the tone of a voice as a speaker goes from one role to the next. The official voice on the phone with one’s boss is not the same as the voice one uses in chiding one’s child or the voice that appears in intimate situations with one’s lover. Indeed, a recent article in the *New York Times* (Jaret, 2005) detailed a major increase in the number of people who seek so-called “voice makeovers” to change the impressions that they are making. These people are no longer just actors, singers,

and famous people, as was the case in the past: Now they range from marketing consultants from San Jose to second grade teachers in Manhattan. Clearly, their everyday experiences have taught them just how important the voice may be in the impressions that they are making.

References

- Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perceptions. *Speech Monographs*, *35*, 492-503.
- Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *Journal of Social Psychology*, *99*, 207-220.
- Bakan, D. (1966). *The duality of human existence*. Chicago, IL: Rand McNally.
- Banse, R., & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality Social Psychology*, *70*, 614–636.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.
- Batstone, S., & Tuomi, S. K. (1981). Perceptual characteristics of female voices. *Language and Speech*, *24*, 111-123.
- Beckett, N. E., & Park, B. (1995). Use of category versus individuating information: Making base-rates salient. *Personality and Social Psychology Bulletin*, *21*, 21-31.
- Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *Industrial and Labor Relations Review*, *55*, 3-21.
- Berry, D. (1992). Vocal types and stereotypes: Joint effects of vocal attractiveness and vocal maturity on person perception. *Journal of Nonverbal Behavior*, *16*, 41-54.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*, 5-20.
- Biernat, M., & Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*, 485-499.
- Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*, 674-679.

- Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, 87, 763–778.
- Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, 83, 5-25.
- Blau, F. D., & Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic Perspectives*, 14, 75-99.
- Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 11 pp. 1-52). Hillsdale, NJ: Erlbaum.
- Brewer M. B. (1996). When stereotypes lead to stereotyping: The use of stereotypes in person perception. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds), *Stereotypes and stereotyping* (pp. 254 – 275). New York: Guilford Press.
- Broverman, I.K., Broverman, D. M., Clarkson, F. E., Rosenkrantz, P. S., & Vogel, S. R. (1970). Sex-role stereotypes and clinical judgments of mental health. *Journal of Consulting and Clinical Psychology*, 34, 1-7.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39-54.
- Campbell, A. (1971). *White attitudes toward black people*. Ann Arbor, MI: Institute for Social Research.
- Cejka, M.A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, 25, 413-423.
- Cohen, S. L., & Bunker, K. A. (1975). Subtle effects of sex role stereotypes on recruiters' hiring decisions. *Journal of Applied Psychology*, 60, 566-572.

- Coleman, R. O. (1976). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research, 19*, 168-180.
- Collins, S. A. (2000). Men's voices and women's choices. *Animal Behavior, 60*, 773-780.
- Corneille, O., Klein, O., Lambert, S., and Judd, C. M. (2002). On the role of familiarity with units of measurement in categorical accentuation: Tajfel and Wilkes (1963) revisited and replicated. *Psychological Science, 13*, 380-383.
- Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior, 56*, 225-248.
- Deal, J. J., & Stevenson, M. A. (1998). Perceptions of Female and Male Managers in the 1990s: Plus ça change . . . *Sex Roles, 38*, 287-300.
- Deaux K., & Lewis, L. L. (1984). Structure of gender stereotypes—interrelationships among components and gender label. *Journal of Personality and Social Psychology, 46*, 991-1004.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573-598.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111*, 3-22.
- Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin, 15*, 543-558.
- Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? An analysis of attitudes, beliefs, and emotions. *Psychology of Women Quarterly, 15*, 203-216.

- Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*, 735-754.
- Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383-386.
- Fairbanks, G. (1940). *Voice and articulation drillbook* (p.127). New York: Harper & Brothers.
- Fant, G. (1966). A note on vocal tract size factors and nonuniform F-pattern scalings. *Speech Transactions Laboratory Quarterly Progress and Status Report*, *4*, 22-30.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013-1027.
- Fiske, S. T. (2004). *Social beings: A core motives approach to social psychology*. New York: Wiley.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878-902.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1-74). New York: Guildford Press.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

- Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues, 55*, 473-489.
- Fitch, W. T. (1994). *Vocal tract length perception and the evolution of language*. Unpublished Doctoral Dissertation, Brown University.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America, 102*, 1213-1222.
- Fitch, W. T. (2002). Comparative Vocal Production and the Evolution of Speech: Reinterpreting the Descent of the Larynx. In A. Wray (Ed.), *In transition to language*. Oxford: Oxford University Press.
- Fitch, W. T. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America, 106*, 1511-1522.
- Ford, T. E., & Stangor, C. (1992). The role of diagnosticity in stereotype formation: Perceiving group means and variances. *Journal of Personality and Social Psychology, 63*, 356-367.
- Gerdes E. P., & Garber, D. M. (1983). Sex bias in hiring: Effects of job demands and applicant competence. *Sex Roles, 9*, 307-319.
- Glick, P., Diebold, J., Bailey-Werner, B., & Zhu, L. (1997). The two faces of Adam: Ambivalent sexism and polarized attitudes toward women. *Personality and Social Psychology Bulletin, 23*, 1323-1334.
- Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491-512.

- Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, 55, 178-186.
- Graddol, D. & Swann, J. (1983). Speaking fundamental frequency: Some physical and social correlates. *Language and Speech*, 26, 351-366.
- Greeley, A. M., & Sheatsley, P. B. (1971). Attitudes toward racial integration. *Scientific American*, Vol. 225, 13-19.
- Hogg, M. A. (1985). Masculine and feminine speech in dyads and groups: A study of speech style and gender salience. *Journal of Language and Social Psychology*, 4, 99-112.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior*, 5, 269-298.
- Heilman, M. E. (1989). Has anything changed? Current characterization of men, women, and managers. *Journal of Applied Psychology*, 74, 935-942.
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57, 657-674.
- Hummert, M. L., Mazloff, D., & Henry, C. (1999). Vocal characteristics of older adults and stereotyping. *Journal of Nonverbal Behavior*, 23, 111-132.
- Hyman, H. H., & Sheatsley, P. B. (1956). Attitudes toward desegregation. *Scientific American*, 195, 35-39.
- Hyman, H. H., & Sheatsley, P. B. (1964). Attitudes toward desegregation. *Scientific American*, 211, 16-23.
- Jaret, P. (2005 July 21). My voice has got to go. *New York Times*. Retrieved (4/29/2006) from <http://www.nytimes.com/2005/07/21/fashion/thursdaystyles/21voice.html?ei=5090&en=8db17c665c6700a8&ex=1279598400&partner=rssuserland&emc=rss&pagewanted=all>

- Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kahimia, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*, 899-913.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602-619.
- Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 180-232). New York: McGraw-Hill.
- Jussim, L., & Fleming, C. (1996). Self-fulfilling prophecies and the maintenance of social stereotypes: The role of dyadic interactions and social forces. In C. N. Macrae, C. Stangor, and M. Hewstone (Eds.), *Stereotypes and stereotyping* (p. 161-192). New York: Guilford Press.
- Kalin, R & Hodgins, D. C. (1984). Sex bias in judgments of occupational suitability. *Canadian Journal of Behavioral Science*, *16*, 311-325.
- Klonis, S. C., Plant, E. A., & Devine, P.G. (2005). Internal and external motivation to respond without sexism. *Personality and Social Psychology Bulletin*, *31*, 1237-1249.
- Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the voice reveals: Within- and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin*, *32*, 806-819.
- Ko, S. J., Judd, C. M., & Stapel, D. A. (2007). A Voice in the Workplace: How Vocal Cues Impact Judgments Related to Hiring Decisions. *Under review*.
- Ko, S. J., Muller, D., Judd, C. M., & Stapel, D. A. (2007). Sneaking in Through the Back Door: How Category-Based Stereotype Suppression Leads to Rebound in Feature-based Effects. *Under review*.

- Kramer, C. (1977). Perceptions of Female and Male Speech. *Language and Speech*, 20, 151-161.
- Linville, S. E. (1996). The sound senescence. *Journal of Voice*, 10, 190-200.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, 39, 821-831.
- Locksley A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, 18, 23-42.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, 67, 808-817.
- Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, 28, 250-259.
- Martinko, M., & Gardner, W. (1983). A methodological review of sex-related access discrimination problems. *Sex Roles*, 9, 825-839.
- Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of Speech and Persuasion. *Journal of Personality and Social Psychology*, 34, 615-624.
- Mulac, A., & Giles, H. (1996). "You're only as old as you sound"; Perceived vocal age and social meaning. *Health Communication*, 8, 199-215.
- Montepare, J. M., & Zebrowitz, L. (1998). Person perception comes of age: The salience and significance of age in social judgments. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 30, pp. 93-161). New York: Academic Press.
- Nesbitt, M. N., & Penn, N. E. (2000). Gender stereotyping after thirty years: A replication of Rozenkrantz, et al. (1968). *Psychological Reports*, 87, 493-511.

- Paton, N. (2003, March 24). "Breaking the sound barrier". Guardian Unlimited. Retrieved (3/29/2004) from <http://jobs.guardian.co.uk/officehours/story/0,9897,920386,00.html>
- Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
- Powell, G. N., & Butterfield, D. A. (1979). The 'good manager': Masculine or androgynous? *Academy of Management Journal*, 22, 395-403.
- Powell, G. N., & Butterfield, D. A. (1989). The 'good manager': Does androgyny fare better in the 1980's? *Group and Organizational Studies*, 14, 216-233.
- Powell, G. N., Butterfield, D. A., and Parent, J. D. (2002). Gender and managerial stereotypes: Have times changed? *Journal of Management*, 28, 177-193.
- Ray, G. B. (1986). Vocally cued personality prototypes: An implicit personality theory approach. *Communication Monographs*, 53, 266-276.
- Rendall, D., Kollias, S., Ney C., & Lloyd P. (2005). Pitch (F₀) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal Of The Acoustical Society Of America*, 117, 944-955.
- Rosch, E. & Mervis, C. B. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89-115.
- Rosenkrantz, P., Vogel, S., Bee, H., Broverman, I., & Broverman, D. M. (1968). Sex-role stereotypes and self-concepts in college students. *Journal of Clinical and Consulting Psychology*, 32, 287-285.
- Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counter-stereotypical impression management. *Journal of Personality and Social Psychology*, 74, 629-645.

- Rudman, L. A., & Glick, P. (1999). Feminized management and backlash towards agentic women: The hidden cost to women of a kinder, gentler image of middle-managers. *Journal of Personality and Social Psychology*, *77*, 1004-1010.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*, 743-762.
- Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, *57*, 95-100.
- Schein, V. E. (1975). Relations between sex role stereotypes and requisite management characteristics among female managers. *Journal of Applied Psychology*, *60*, 340-344.
- Schein, V. E. (1978). Sex role stereotyping, ability and performance: Prior research and new directions. *Personnel Psychology*, *31*, 259-268.
- Schein, V. E., & Davidson, M. J. (1993). Think manager, think male. *Management Development Review*, *6*, 24-28.
- Scherer, K. R. (1979). Personality markers in speech. In K.R. Scherer and H. Giles (Eds.), *Social Markers in Speech*, Cambridge: Cambridge University Press.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227-256.
- Scherer, K.R., Johnstone, T., Klasmeyer, G., (2003). Vocal expression of emotion. In: R.J. Davidson, K.R. Scherer, and H. Goldsmith (Eds.), *Handbook of affective sciences*, Oxford University Press, New York and Oxford.
- Schuman, H., Steeh, C., & Bobo, L. (1985). *Racial attitudes in America: Trends and interpretations*. Cambridge, MA: Harvard University Press.
- Szesny, S., & Kühnen, U. (2004). Meta-cognition about biological sex and gender-stereotypic physical appearance: Consequences for the assessment of leadership competence. *Personality and Social Psychology Bulletin*, *30*, 13-21.

- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.
- Sjölander K, & Beskow J. (2002). Wavesurfer (version 1.4.6) [Computer software].
Stockholm, Sweden: Centre for Speech Technology (CTT).
- Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality. *Language and Speech*, *18*, 145-152.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*, Harvard University Press, Cambridge, M.A.
- Smith, P. M. (1979). Sex markers in speech. In K.R. Scherer and H. Giles (Eds.), *Social Markers in Speech*, Cambridge: Cambridge University Press.
- Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology*, *54*, 101–114.
- Taylor, D. G., Sheatsley, P. B., & Greeley, A. M. (1978). Attitudes toward racial integration. *Scientific American*, *238*, 42-49.
- Tusing, K. J., & Dillard, J. P. (2000). The sound of dominance: Vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, *26*, 148-171.
- Wegner, D. M., & Erber, R. (1992). The hyperaccessibility of suppressed thoughts. *Journal of Personality and Social Psychology*, *63*, 903-912.
- Wenzlaff, R. M., Wegener, D. M., & Klein, S. B. (1991). The role of thought suppression in the bonding of mood and thought. *Journal of Personality and Social Psychology*, *60*, 500-508.
- Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *Journal of the Acoustical Society of America*, *110*, 464-478.

- Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262-274.
- Yamada, N., Hakoda, Y., Yuda, E., & Kusuhara, A. (2000). Verification of impression voice in relation to occupational categories. *Psychological Reports, 86*, 1249-1263.
- Yogo, Y., Ando, M., Hashi, A., Tsutsui, S., & Yamada, N. (2000). Judgments of emotion by nurses and students given double-bind information on a patient's tone of voice and message content. *Perceptual & Motor Skills, 90*, 855-863.
- Zuckerman, M., Miyaki, K., & Hodgins, H. S. (1991). Cross-channel effects of vocal and physical attractiveness and their implications for interpersonal perception. *Journal of Personality and Social Psychology, 60*, 545-554.
- Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior, 13*, 67-82.

Summary

Given that the voice is our main form of communication, we know surprisingly little about how it impacts judgment and behavior. Furthermore, the modern advancement in telecommunication systems, such as cellular phones, has meant that a large proportion of our everyday interactions are conducted vocally and these interactions can have important consequences. For example, imagine two people, who met through an internet-dating service, having their first phone conversation or potential employees on phone interviews—what kind of stereotypic associations are elicited by the voices? And, consequently, how might these associations influence the dating relationship and whether the interviews are successful? These are some of the questions that motivate my research.

In Chapter 2, I report research that attempts to shift the traditional focus of visual cues to auditory cues as a basis for stereotyping. Moreover, our approach examines whether gender-signaling vocal cues lead not only to between-category, but also to within-category gender stereotyping. Study 2.1 showed that both males and females vary, within category, in how feminine their voices sound and that perceptions of vocal femininity are highly consensual. Furthermore, the measured acoustic characteristics that differed between gender were also related to perceptions of within-gender femininity. Subsequent studies demonstrated that variability in vocal femininity affects gender stereotyping when the targets are all of the same gender (Study 2.2) and when the targets are of different genders (Study 2.3). In the latter case, evidence of both category-based and feature-based stereotyping was found. Mediation analyses showed that the relationship between acoustics and stereotyping was in part due to femininity.

In Chapter 3, we examined the impact of applicants' voices in job interview situations. Specifically, in Study 3.1, using male and female speakers posing as job applicants, we investigated how applicants' vocal femininity cues and résumé information

impacted judgments on the dimensions of competence and warmth. Results showed that competence was solely affected by vocal femininity such that applicants with masculine sounding voices were rated as more competent than applicants with feminine sounding voices, regardless of the applicants' gender or résumé information. Warmth, on the other hand, was predominantly affected by résumés in expected ways—applicants with feminine résumés were rated as warmer than those with masculine résumés. The potent effect of vocal femininity on competence was replicated in a second study (Study 3.2) under more stringent and controlled conditions.

Given the dramatic shifts in societal norms to curb overt stereotyping and prejudice, these biases may leak out in more subtle ways than were apparent in the past. Accordingly, in Chapter 4 we present research showing how the suppression of stereotypes might affect post-suppression category-based stereotyping and the more subtle feature-based stereotyping. In support of our proposition, participants in the suppression condition used more feature-based, but less category-based stereotypes in their post-suppression task than participants in the control condition. Furthermore, a relation between post-suppression category-based and feature-based stereotyping existed in the suppression condition such that increases in feature-based stereotyping were associated with decreases in category-based stereotyping. Findings as a whole suggests that norms placed to reduce stereotypic biases may ironically lead people to be more vulnerable to biases as a function of within-category features.

Nederlandse Samenvatting

Gegeven dat de stem onze belangrijkste manier van communiceren is, weten we verrassend weinig over de manier waarop de stem beoordelingen en gedrag beïnvloedt. Als gevolg van de moderne vooruitgang in telecommunicatiesystemen, zoals de mobiele telefonie, verloopt een groot deel van onze dagelijkse interacties vocaal. Deze vocale interacties kunnen soms belangrijke gevolgen hebben. Denk bijvoorbeeld aan twee personen die elkaar hebben ontmoet via een internet datingservice en hun eerste telefoongesprek hebben, of aan een potentiële werknemer in een telefonisch sollicitatiegesprek. Welke stereotypische associaties worden er dan opgeroepen door de stemmen? En, als gevolg daarvan, hoe zouden deze associaties de liefdesrelatie kunnen beïnvloeden en het succes van de sollicitatie gesprekken? Dat zijn enkele van de vragen die centraal staan in mijn onderzoek.

In Hoofdstuk 2 rapporteer ik onderzoek dat tracht de traditionele nadruk op visuele kenmerken te verschuiven naar auditieve kenmerken als basis van stereotypering. Onze aanpak onderzoekt of vocale sekse kenmerken niet alleen leiden tot stereotypering tussen sekse categorieën, maar ook tot stereotypering binnen een sekse categorie. Studie 2.1 liet zien dat zowel mannen als vrouwen binnen elke categorie variëren in hoe vrouwelijk hun stemmen klinken en dat er een hoge mate van consensus bestaat in waarnemingen van vocale vrouwelijkheid. Daarnaast waren de gemeten akoestiek kenmerken die verschillen tussen de seksen ook gerelateerd aan waarnemingen van vrouwelijkheid binnen de sekse categorieën. Verdere studies toonden aan dat variabiliteit in vocale vrouwelijkheid van invloed is op sekse stereotypering wanneer de targets allen van dezelfde sekse zijn (Studie 2.2) en wanneer de targets van verschillende seksen zijn (Studie 2.3). In het laatste geval werd er bewijs gevonden voor zowel op categorie gebaseerde en op kenmerk gebaseerde stereotypering.

Mediatie analyses lieten zien dat de relatie tussen akoestiek en stereotypering voor een deel veroorzaakt werd door vrouwelijkheid.

In Hoofdstuk 3 hebben we de invloed van de stemmen van sollicitanten bij sollicitatie gesprekken onderzocht. Meer specifiek hebben we in Studie 3.1, gebruikmakend van vrouwelijke en mannelijke sprekers die zich voordeden als sollicitanten, onderzocht wat de invloed was van vrouwelijke vocale kenmerken en het curriculum vitae op de dimensies competentie en warmte. De resultaten lieten zien dat competentie alleen werd beïnvloed door vocale vrouwelijkheid waarbij sollicitanten met mannelijk klinkende stemmen als competentier werden beoordeeld dan sollicitanten met vrouwelijk klinkende stemmen, afgezien van de sekse of het curriculum vitae van de sollicitant. Warmte werd echter vooral op de verwachte manier beïnvloed door het curriculum vitae - Sollicitanten met een vrouwelijk curriculum vitae werden als warmer beoordeeld dan degenen met een mannelijk curriculum vitae. Het sterke effect van vocale vrouwelijkheid werd gerepliceerd in een tweede studie (Studie 3.2) onder striktere en meer gecontroleerde omstandigheden.

Gegeven de dramatische verschuivingen in maatschappelijke normen om openlijke stereotypering en vooroordelen om te buigen, kunnen dergelijke vertekeningen zich op subtielere manieren ten toon spreiden dan in het verleden is gebleken. In lijn hiermee presenteren we in Hoofdstuk 4 onderzoek dat laat zien hoe het onderdrukken van stereotypen van invloed is op categorie gebaseerde stereotypering en op subtielere vormen zoals op kenmerk gebaseerde stereotypering, na het onderdrukken van die stereotypen. Zoals verondersteld gebruikten deelnemers in de onderdrukconditie meer op kenmerk gebaseerde, maar minder op categorie gebaseerde stereotypen in de taak die volgde op het onderdrukken van stereotypen dan deelnemers in de controle conditie. Daarnaast bestond er een relatie tussen op categorie gebaseerde en op kenmerk gebaseerde stereotypering in de onderdrukconditie zodat meer op kenmerk gebaseerde stereotypering werd geassocieerd met

minder op categorie gebaseerde stereotypering. De bevindingen als geheel suggereren dat de normen die bestaan om stereotypische vertekeningen te verminderen, er ironisch genoeg toe kunnen leiden dat mensen kwetsbaarder worden voor vertekeningen als functie van kenmerken die binnen een categorie kunnen variëren.

Footnotes

¹ Chapter 2 is based on Ko, Judd, & Blair, 2006.

² For simplicity we characterize the vocal cues of interest as vocal femininity. However, our research shows that the dimension could as easily be characterized (in the opposite direction) as vocal masculinity (see footnote 3).

³ The voices were also rated on pleasantness and babyishness, as these figure importantly in research on visual cues (Montepare & Zebrowitz, 1998). Even though femininity and babyishness turned out to be positively correlated across these voices, the effects of femininity on judgments in subsequent studies remained the same regardless of whether babyishness and pleasantness were controlled. Accordingly, we present the simpler analyses that did not control for these other dimensions. In addition to these judgments, a different group of judges also rated the set of male voices on masculinity. Masculinity ratings correlated $-.92$ with femininity. Accordingly, we concluded that masculinity and femininity were opposite ends of the same dimension. For the remainder of this article, only judged femininity is used.

⁴ To confirm these differences, we asked pretest participants to rate the person described in each description for stereotypicality (male versus female) and valence (dislikeable versus likeable), both on 7 points scales varying from -3 to $+3$. The mean ratings for the descriptions on these two dimensions were as follows: Stereotypic male negative: -1.31 and -2.85 ; Stereotypic male positive: -0.70 and 1.69 ; Stereotypic female negative: 2.09 and -1.91 ; Stereotypic female positive: 1.53 and 2.62 . The full text of the descriptions may be obtained by writing the authors.

⁵ The judgments of the female voices and the judgments of the male voices were obtained at two different times with two separate groups of participants. We present the two groups together because all other aspects of the study were the same. Because the

participants were not randomly assigned to gender category, the data are analyzed separately by category and direct comparisons between the two sets of results should be made cautiously.

⁶ The pattern of findings in this study did not differ as a function of individual self-descriptions or the aggregate of the pairs of self-descriptions.

⁷ We tested for participant gender effects but found none here or in any of the other studies presented in this paper.

⁸ Separate analyses of the data for the male and female targets showed that this pattern occurred for both target groups, thereby not replicating the target gender difference suggested in Study 2.2.

⁹ We report these R-square differences rather than the more traditional standardized regression coefficients for this mediational analysis in order to assess the simultaneous effects of all four acoustic characteristics as a set.

¹⁰ Chapter 3 is based on Ko, Judd, & Stapel, 2007.

¹¹ The participants also rated the resumes on valence using a 1 (*very negative*) to 7 (*very positive*) scale. The valence mean ratings for the chosen resumes were as follows: feminine resumes were 5.18; masculine resumes were 4.74. Pairwise comparisons yielded differences between the feminine and masculine resumes, $t(1, 15) = 2.74, p < .05$.

¹² In our previous work (Ko et al., 2006) we had had separate participants scale voices on masculinity. The extremely high negative correlation of the resulting masculinity ratings with the femininity ratings ($r = -.92$) led us to conclude that vocal femininity and masculinity were opposite ends of the same dimension. Accordingly, voices low on femininity are voices high on masculinity and *vice versa*, voices high on femininity are voices low on masculinity.

¹³ We found no participant gender effects here or in any of the following results.

¹⁴ The contrast code for vocal category was +1 if female and 1 if male. The contrast code for the behaviors in the four competence experimental scenarios was +1 if competent and -1 if incompetent. The contrast code for the behaviors in the warmth experimental scenarios was +1 if warm and -1 if cold.

¹⁵ We did not find any participant gender effects in these data.

¹⁶ Chapter 4 is based on Ko, Muller, Judd, & Stapel, 2007.

¹⁷ The English descriptions were translated into Dutch.

¹⁸ For a detailed description of the vocal femininity scaling process, criteria used to select the final subset of voices, and procedure of the probability task see (2006).

¹⁹ Our results also revealed effects of less theoretical interest such as a valence main effect, $F(1,168) = 55.95, p < .001$. This valence effect was moderated by stereotypicality, $F(1,168) = 19.80, p < .001$, revealing that the difference between female and male negative self-descriptions ($M = 14.67$ vs. $M = -3.26$) was larger than the difference between female and male positive self-descriptions ($M = 2.71$ vs. $M = -6.30$). Finally, the condition by valence interaction, $F(1,168) = 5.43, p < .021$, showed that the difference between negative and positive self-descriptions was larger in the control condition ($M = 6.77$ vs. $M = -3.07$) than in the suppression condition ($M = 4.64$ vs. $M = -0.52$).

²⁰ Consistent with past work on vocal stereotyping (Ko, Judd, & Blair 2006; Ko, Judd, Stapel, 2007) we found no participant gender effects here or in any of the following results.

²¹ Of less theoretical relevance was a valence main effect, $F(1,168) = 4.28, p < .041$, indicating that more feminine voices were rated as more probable for negative self-descriptions ($M = -1.49$) than for positive ones ($M = -2.88$).

Table 2.1

Study 2.1: Statistics on the Four Acoustic Cues and Set of 47 Female Voices and 47 Male Voices

Means (and Standard Deviations)

Acoustic Cues	Gender of Voices	
	Female	Male
Pitch (F0)	197.37 Hz (18.74)	114.97 Hz (14.15)
Resonance (D_f)	1133.57 Hz (39.99)	1062.93 Hz (51.10)
Pitch Variability (<u>SD</u> of F0)	41.08 Hz (9.25)	19.01 Hz (6.32)
Duration (secs.)	8.57 (0.60)	8.54 (0.57)

Correlations and Partial Correlations (Partiallying Out the Other Three Acoustic Cues) Among the Four Acoustic Cues and Gender Category of Speakers

Acoustic Cues	r	Partial r
Pitch (F0)	.93**	.77**
Resonance (D_f)	.61**	.33*
Pitch Variability (<u>SD</u> of F0)	.82**	.14
Speech Duration	.02	-.13

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

NOTE : Coding for category of speaker was 1=female and 0=male.

Table 2.2

Study 2.1: Correlations and Partial Correlations (Partiallying Out the Other Three Acoustic Cues) Among the Four Acoustic Cues and Judged Femininity for the Female and Male Voices

Acoustic Cues	Gender of Voices			
	Female		Male	
	<i>r</i>	Partial <i>r</i>	<i>r</i>	Partial <i>r</i>
Pitch (F0)	.69**	.73**	.59**	.41**
Resonance (D_r)	.28*	.46**	-.14	-.09
Pitch Variability (<i>SD</i> of F0)	.26*	.04	.46**	.20
Speech Duration	-.06	-.09	.24	.00

* $p < .10$ ** $p < .01$.

Table 2.3

Study 2.2: Mean Slopes Predicting Self-Description Probability Ratings from Vocal

Femininity

Femininity Slope For:	Self-Description			
	Female Stereotypic		Male Stereotypic	
	Positive	Negative	Positive	Negative
Female Voices	4.22**	8.23**	-8.03**	-7.14**
Males Voices	11.02**	9.97**	-6.83**	-7.85**

**Slope significantly different from zero, $p < .001$.

Table 2.4

Study 2.3 (Mixed Voices): Mean Slopes Predicting Self-Description Probability Ratings from Gender Category and Vocal Femininity

Model 1: Gender Category as Sole Predictor

	Self-Description			
	Female Stereotypic		Male Stereotypic	
	Positive	Negative	Positive	Negative
Category Slopes	7.45**	17.72**	-4.59*	-8.97**

Model 2: Femininity and Gender Category as Predictors

	Self-Description			
	Female Stereotypic		Male Stereotypic	
	Positive	Negative	Positive	Negative
Category Slope	4.20*	16.48**	-2.48	-8.49**
Femininity Slope	7.94**	3.04**	-5.16**	-1.18

NOTE: For Category slopes, positive numbers indicate female voices judged as more probable than male voices.

*Slope significantly different from zero, $p < .05$. **Slope significantly different from zero, $p < .001$.

Table 3.1

Study 3.1: Text Presentation of Résumés. Means (and Standard Deviations) for the Composite Ratings on the Dimensions of Warmth and Competence, Broken Down by Applicant Gender and Résumé Type.

Résumé Type	Dimensions			
	Competence		Warmth	
	Female	Male	Female	Male
Feminine	4.87 (0.95)	4.54 (0.94)	5.63 (0.77)	5.68 (0.62)
Masculine	5.10 (0.78)	4.93 (0.88)	4.44 (0.78)	4.31 (0.93)

Table 3.2

Study 3.2: Mean Slopes Predicting Likelihood of Behavior in the Fifth Experimental Scenario from Vocal Category, Vocal Femininity, and Averaged Behaviors in the First Four Experimental Scenarios (Individuating Information), Broken Down by Condition.

	Condition	
	Competence	Warmth
Partial Slopes for:		
Vocal Femininity	-1.46**	-.23
Vocal Category	-1.81*	.58
Individuating Information	.64**	.68**

NOTE: For Category slopes, negative numbers indicate male voices judged as more likely than female voices.

*Slope significantly different from zero, $p < .01$. **Slope significantly different from zero, $p < .001$.

Figure Caption

Figure 2.1. Models of the Effects of Acoustic Characteristics on Stereotypic Inferences.

Figure 3.1. Oral Presentation: Composite Ratings of warmth and competence on the basis of voice category and vocal femininity for each résumé type.

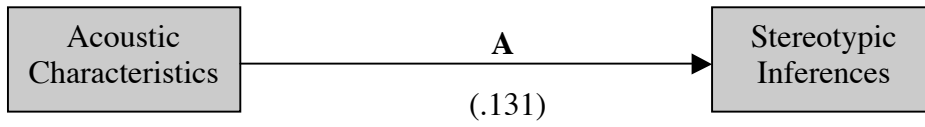
Figure 3.2. Model of the roles played by vocal and résumé cues in job interview contexts.

Dotted lines denote weak associations and solid lines denote strong associations.

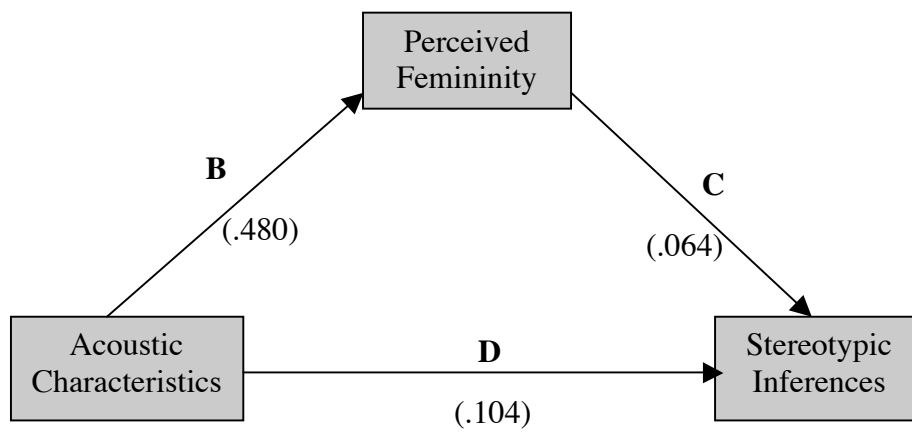
Figure 4.1. Category use as a function of condition and stereotypicality of self-descriptions (controlling for vocal femininity)

Figure 4.2. Vocal Femininity use as a function of condition and stereotypicality of self-descriptions (controlling for gender category)

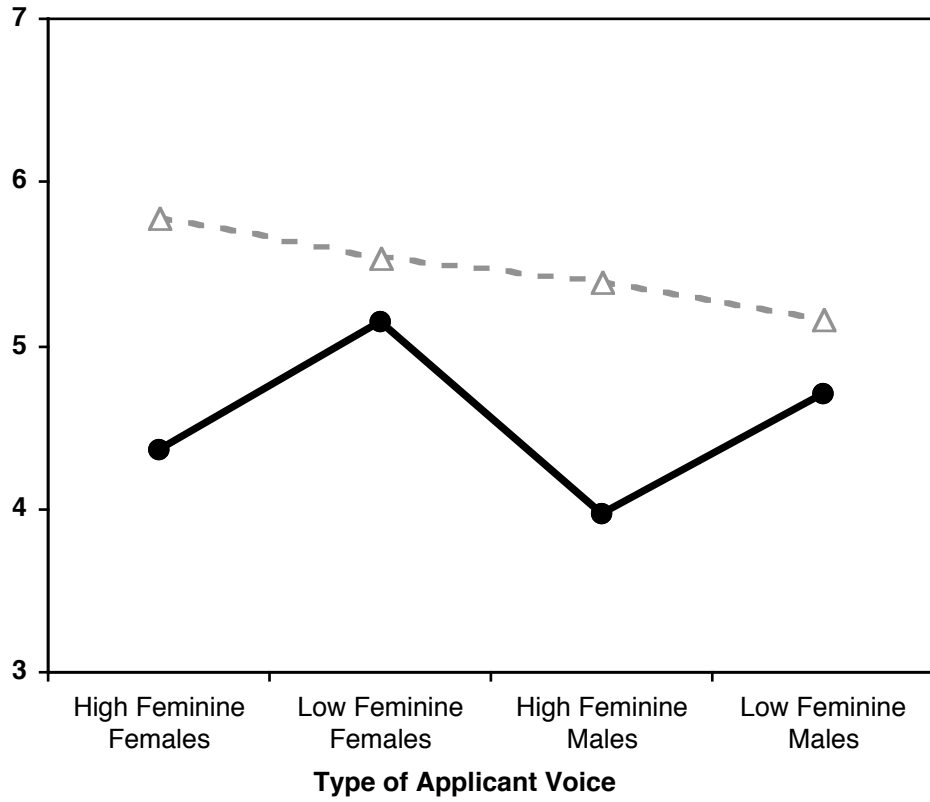
Simple Direct Effect Model



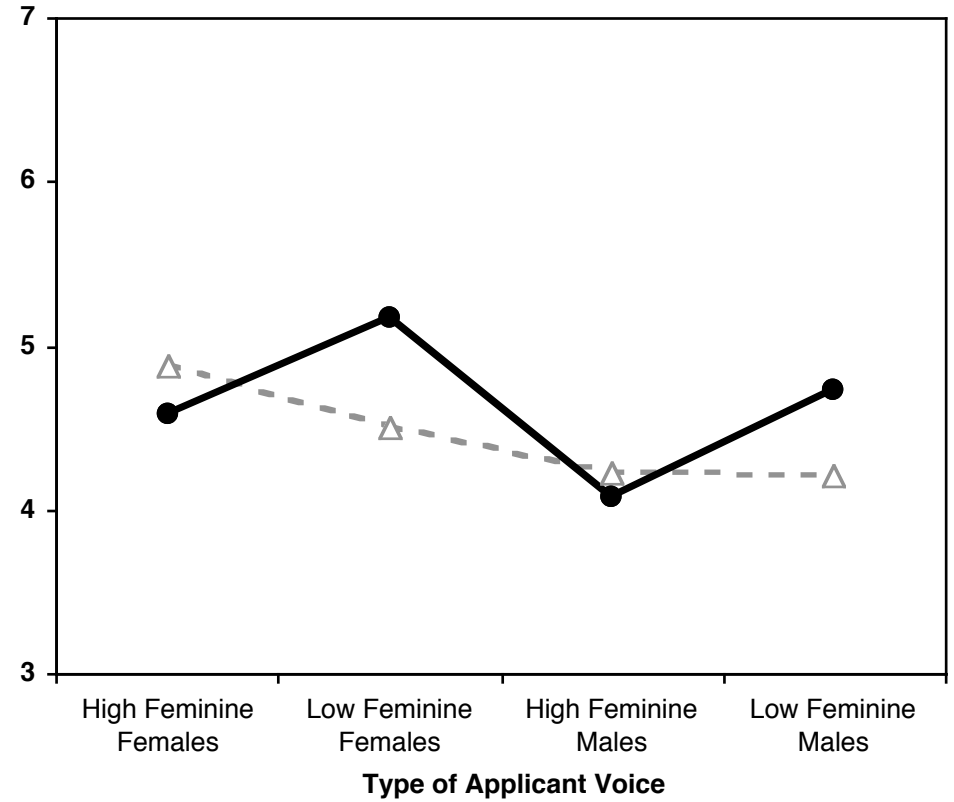
Partial Mediation Model



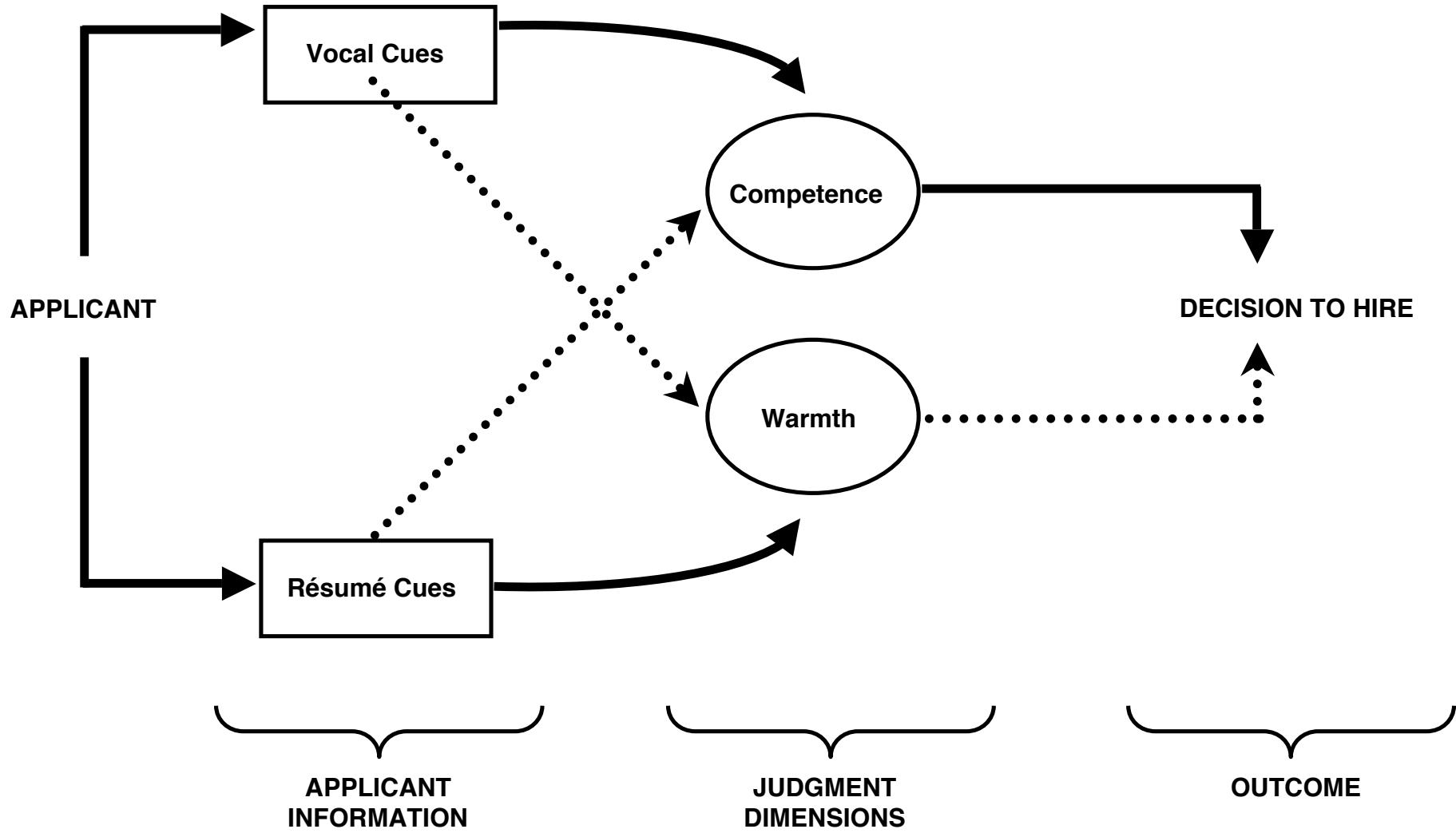
Feminine Résumés

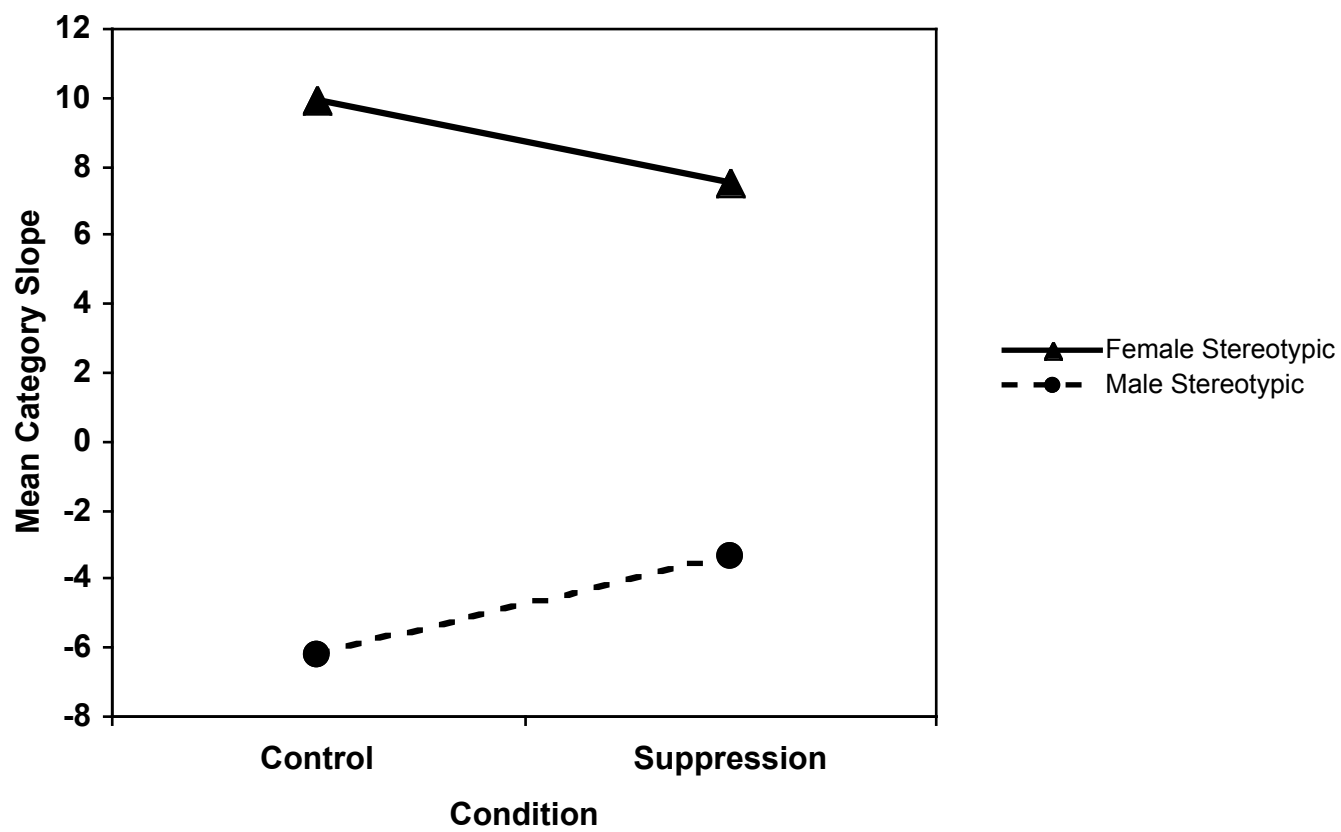


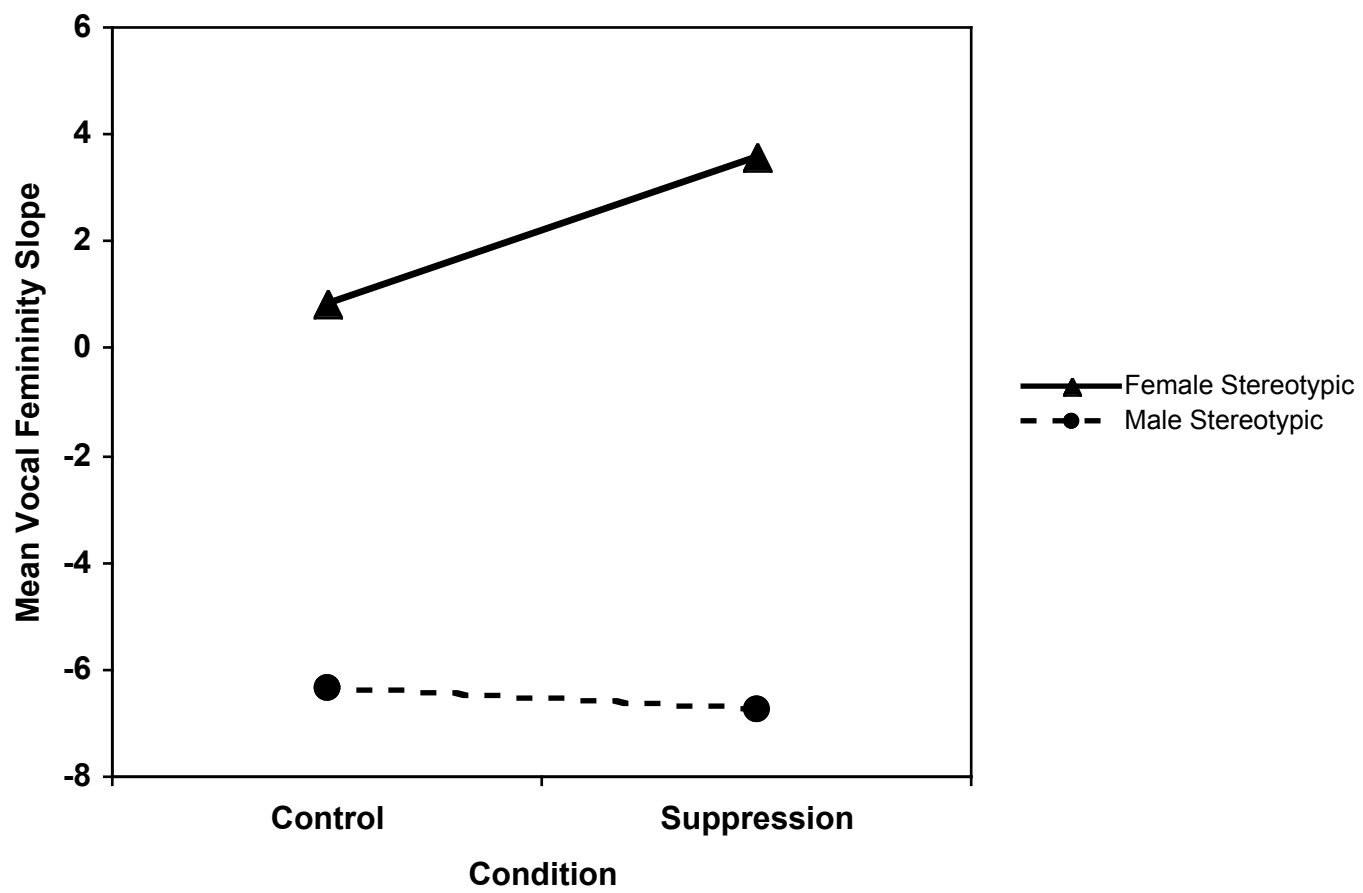
Masculine Résumés



-△- Sociability
 -●- Competence







Appendix A

Experimental Scenarios and Possible Behaviors per Condition

Competence Condition

Experimental Scenario 1: In dit experiment is het de taak van de deelnemer om een uitdagende computer puzzel te maken door de antwoorden op zeven woordproblemen op een juiste manier te achterhalen. Elk voorafgaand woordprobleem moet juist beantwoord worden voordat de deelnemer verder kan gaan met het volgende woordprobleem. Of de deelnemer rond de computer puzzel succesvol af binnen de toegekende tijdslimiet (competent gedrag), of de deelnemer heeft niet genoeg tijd om de computer puzzel af te ronden (incompetent gedrag).

Experimental Scenario 2 In dit experiment wordt de deelnemer gevraagd zich het volgende scenario in te beelden: 'je bent de enige overlevende van een vliegtuig crash. De temperatuur is onder het vriespunt. Er ligt sneeuw op de grond, het landschap is bebost, en je ziet verschillende riviertjes. Het meest nabijgelegen dorp is 32 kilometer ver. Je draagt stadskleding geschikt voor een zakenbespreking. Je slaagt erin om 12 attributen te redden uit het neergestorte vliegtuig.' De deelnemer moet van de 12 attributen een rangordening maken op volgorde van belangrijkheid. Of de deelnemer is in staat om het belang voor overleving in te zien van elk attribuut, en dus om de attributen in de juiste volgorde te zetten (competent gedrag), of de deelnemer is alleen in staat om het gebruikelijke nut van de attributen te zien, en zal de attributen dus niet in de juiste volgorde kunnen zetten (incompetent gedrag).

Experimental Scenario 3: In dit experiment wordt de deelnemer tot leider benoemd van een groep van zes werkers die de deelnemer nog nooit heeft ontmoet. Het is de taak van de deelnemer om de werkers ertoe te zetten om de zes problemen op te lossen. Elke werker bezit één unieke vaardigheid, die hem/haar in staat stelt om slechts één van de zes problemen op te lossen. De deelnemer weet niet welke werker de capaciteiten heeft om welk probleem op te lossen, maar krijgt 5 minuten de tijd om de werkers te observeren terwijl ze met elkaar interacteren. Op basis van deze observatie selecteert de deelnemer voor elke werker het juiste bijbehorende probleem (competent gedrag) of selecteert de deelnemer voor elke werker een onjuist probleem (incompetent gedrag).

Experimental Scenario 4: In dit experiment worden de geheugenvaardigheden van de deelnemer getest. De deelnemer krijgt een driedimensionale Lego sculptuur te zien. Na vijf minuten wordt de sculptuur verwijderd en krijgt de deelnemer een stapel losse Legoblokjes om de sculptuur te herbouwen. Of de deelnemer is in staat om de sculptuur succesvol te herbouwen (competent gedrag), of de deelnemer is niet in staat om de sculptuur te herbouwen door problemen met het zich herinneren van de details (incompetent gedrag).

Experimental Scenario 5: In dit experiment moeten deelnemers uitdagende wiskundige problemen oplossen, terwijl ze worden afgeleid door luide witte ruis. Deze witte ruis klinkt op willekeurige momenten door een luidspreker, waardoor het erg moeilijk wordt voor de deelnemers om zich op de wiskunde taak te concentreren. Mensen met ruimtelijk inzicht hebben waarschijnlijk weinig moeite om goed te presteren, zelfs onder zulke afleidende omstandigheden, omdat de vaardigheden die nodig zijn voor de wiskunde taak vooral automatisch zijn. Mensen zonder ruimtelijk inzicht presteren waarschijnlijk slecht omdat de vaardigheden die nodig zijn voor de taak veel cognitief vermogen vereisen, dat niet direct

beschikbaar is vanwege de afleidende ruis. De deelnemer haalt dus of een hoge wiskundescore ondanks de afleidende witte ruis (competent gedrag) of haalt een lage wiskundescore omdat de witte ruis de deelnemer afleidde (incompetent gedrag).

Warmth Condition

Experimental Scenario 1: In dit experiment wordt aan de deelnemer verteld dat hij of zij samen met iemand anders zal werken aan een computerprobleem. Het paar kan ofwel individueel aan het probleem werken of samen aan het probleem werken. Het paar mag kiezen welke manier volgens hen het meest effectief zal zijn voor het oplossen van het probleem. Het paar trekt strootjes om de bepalen wie beslist hoe ze het probleem gaan aanpakken. Maar de deelnemer weet niet dat er geknoeid is met de strootjes en dat de andere persoon de vertrouweling van de experimentator is zodat de deelnemer altijd wordt gekozen om de beslissing te nemen. De deelnemer beslist ofwel om samen te werken en de oplossing te zoeken door samen over het probleem te discussieren (vriendelijke reactie) ofwel om individueel aan het probleem te werken en te zien wie de oplossing kan vinden (onvriendelijke reactie).

Experimental Scenario 2: In dit experiment wordt de deelnemer groepsleider gemaakt en moet hij of zij 6 verschillende taken toewijzen aan de leden van zijn of haar groep. De aard van elke taak is zodanig dat mensen enige voorkennis nodig hebben op het gebied van de taak om deze interessant te maken voor de persoon, anders is de taak vervelend. De deelnemer wijst de taken toe ofwel door te discussieren met de groep hetgeen veel tijd in beslag neemt maar de kans vergroot dat de taken snel klaar zullen zijn omdat de groepsleden de taken interessant vinden (vriendelijke reactie) ofwel door de taken zelf toe te wijzen zonder overleg hetgeen discussietijd scheelt maar het risico met zich meebrengt dat de groepsleden langer bezig zullen zijn met de taken omdat ze ze oninteressant vinden (onvriendelijke reactie).

Experimental Scenario 3: In dit experiment wordt de deelnemer gevraagd om zich het volgende scenario in te beelden: “Jij en zes vreemden zijn overlevenden van een schipbreuk en zijn op een verlaten eiland gestrand. De enige voedselbron bestaat uit de vele verschillende soorten bonen die op het eiland groeien. Echter, de meeste bonen zijn giftig en alleen jij weet welke bonen niet giftig zijn. Jij slaagt erin om net voldoende eetbare bonen te vinden om een kleine zak te kunnen vullen. Je hebt geen idee wanneer jullie gered zullen worden”. De deelnemer beslist ofwel om de zak met bonen met de andere overlevenden te delen waardoor de deelnemer de tijd verkleint om te kunnen overleven alvorens gered te worden (vriendelijke reactie), ofwel om de zak met bonen voor zichzelf te bewaren zodat de tijd om te kunnen overleven verlengd wordt maar de deelnemer het risico loopt om alleen over te blijven omdat de andere overlevenden waarschijnlijk zullen sterven (onvriendelijke reactie).

Experimental Scenario 4: In dit experiment brengt een vrouwelijke experimentator de deelnemer naar een kleine ruimte om een vragenlijst in te vullen. De deelnemer kan door een open deur een bureau, een archiefkast en een boekenkast waarop papieren liggen opgestapeld zien staan. Zodra de deelnemer klaar is met de vragenlijst dient hij of zij op een knop te drukken om de tweede experimentator te roepen. De experimentator sluit een gordijn tussen de ruimte en haar kantoor en blijft in haar kantoor waarbij ze genoeg lawaai maakt om de deelnemer haar aanwezigheid te laten merken. Vier minuten later zet ze een geluidsband aan die het laat klinken alsof ze op een stoel geklommen is om wat papieren van de boekenkast te pakken. Dan hoort de deelnemer een harde klap alsof ze op de grond gevallen is. De

deelnemer hoort haar zeggen: “Oh mijn god, mijn voet... ik...ik...kan hem niet meer bewegen...” Op dit moment kan de deelnemer ofwel proberen de vrouw te helpen door haar te hulp te schieten (vriendelijke reactie) of blijven zitten en de knop indrukken op de muur om de tweede experimentator te waarschuwen (onvriendelijke reactie).

Experimental Scenario 5: In dit experiment wordt de deelnemer verteld dat hij of zij op video zal worden opgenomen tijdens maken van reclame voor een nieuw schoonmaakmiddel. Eigenlijk is dit verhaal niet waar, en is het alleen een manier om de deelnemer zich angstig te laten voelen. Het werkelijke doel van het experiment is om te onderzoeken hoe angstige mensen zich gedragen ten opzichte van andere mensen die in dezelfde situatie verkeren. De deelnemer wordt naar een wachtruimte gebracht waar een andere angstige deelnemer (een vertrouweling van de experimentator) ook wacht. Of de deelnemer begint een gesprek met de andere deelnemer om zijn of haar gevoelens te kunnen delen waardoor de angst vermindert (de vriendelijke reactie) of de deelnemer kiest ervoor om alleen te zitten en de andere persoon te negeren en te proberen om alleen met deze angst om te gaan (de onvriendelijke reactie).