8

“Oh It Was a Woman! Had I Known I Would Have Reacted Otherwise!”: Developing Digital Methods to Switch Identity-Related Properties in Order to Reveal Linguistic Stereotyping

**Mattias Lindvall-Östling, Mats Deutschmann,
and Anders Steinvall**

1 Introduction and Background

Introduction

Ever since John Money, Hampson, and Hampson (1955) introduced the concept of “gender roles”, that is, the idea that behaviours, activities and attributes that a given society considers appropriate for men and women are socially constructed rather than biologically determined, the notion of masculine and feminine behaviour as dichotomous and pre-determined

M. Lindvall-Östling (✉) • M. Deutschmann
School of Humanities, Education and Social Sciences, Örebro University,
Örebro, Sweden
e-mail: Mattias.Ostling@oru.se; Mats.deutschmann@oru.se

A. Steinvall
Department of Language Studies, Umeå University, Umeå, Sweden
e-mail: Anders.steinvall@umu.se

has been under intense sociological scrutiny (Butler, 1990; Messerschmidt, 2009; Pecis, 2016; Risman, 2009; Wallenberg & Thanem, 2016; West & Zimmerman, 1987). For example, in their seminal article “Doing Gender”, West and Zimmerman (1987) argue that gender roles are being shaped continually in all everyday interactions. In this process, individuals are continually assessed according to socially accepted conceptions of what appropriate male and female behaviour may be, but there is room for resistance and norms are not constant. Today, rather than being seen as a dichotomy, gender is increasingly being considered along a continuum, where, in addition to hegemonic masculinity and hegemonic femininity representing what is culturally normative (Connell, 1987, 2005), there is increased acceptance that there is an array of subordinate gender identities that do not conform to hegemonic gender constructions. In spite of this progress, it would be naive to assume that normative hegemonic views of gender roles have lost their power of influence. While we as researchers clearly distance ourselves from the idea that gender norms should determine who we are, we also recognise that centuries of gender construction based on a hegemonic dichotomy have left their mark. Accordingly, we would argue that most people’s perception of others, whether they are aware of it or not, is strongly influenced by hegemonic gender stereotypes. It is not until awareness of these mechanisms is raised that we can begin to break free from the hegemonic straightjacket, and this, we would argue, is an important educational mission.

Decades of research in social psychology has shown that we draw on pre-existing attitudes and stereotypical beliefs when forming initial impressions of others, and that these stereotypes have a deep impact on how we perceive the people we meet (Higgins & Bargh, 1987; Macrae & Bodenhausen, 2001). More important for this chapter is the research that shows that these stereotype-based categorisations also affect how we interpret and process speech (Hay, Warren, & Drager, 2006; Johnson, Strand, & D’Imperio, 1999; Niedzielski, 1999) and that language affects how individuals are judged in relation to, for example, intellect and empathy (Cavallaro & Ng, 2009; Fuertes, Gottdiener, Martin, Gilbert, & Giles, 2012).

Sociolinguistic research (e.g. Bucholtz & Hall, 2005, p. 607) suggests that identity can be viewed as “an intersubjectively achieved social and

cultural phenomenon" and is as such identifiable in discourse. This opens up for the theoretical position that social identity, as expressed through language for example, is something that is renegotiated during every meeting between humans (see Crawford, 1995). Stereotyping based on various social categories, such as gender, age, social class, ethnicity, sexuality or regional affiliation, serves to simplify how we perceive and process information about individuals and builds up expectations on how they are supposed to act, and language is at the heart of such mechanisms (see Talbot, 2003, p. 468). Thus, there is a conflict between the complex and dynamic negotiation of identity construction and the routine-like way we judge others. Accordingly, awareness of mechanisms of linguistic stereotyping and identity is of crucial importance in education, especially in the training of groups who will be working with people in their future profession, groups such as teachers, police, psychologists, nurses. Courses addressing these issues can fall in the trap of reinforcing old or introducing new stereotypes in focusing on, for example, statistical gender differences.

We argue that students, instead, need deeper insights into how they themselves are affected by such processes. The goal of the project RAVE (Raising Awareness through Virtual Experiencing) is to develop experiential pedagogic approaches aimed at raising sociolinguistic language awareness about conceived identity-related phenomena in language. More specifically, our ambition is to develop innovative methods for raising subjects' awareness of their own linguistic stereotyping, biases and prejudices, and to systematically explore the efficiency of these methods. The RAVE framework is based on traditional matched-guise methodology from sociolinguistics but uses new digital manipulation techniques for the guise-flipping. Here digital technology has opened up new possibilities for the manipulation of identity variables such as gender—the voice quality of a recording can be manipulated to sound like a man or a woman, for example. In this way, we can illustrate how we as listeners react differently to a speaker and what is being said depending on the perceived identity of the speaker. This insight can then be used as a starting point for self-reflection and awareness-raising activities. The primary focus of this chapter is to share our experiences so far in the development

of methods for raising awareness about gender and language issues conducted within this project.

Structure of the Chapter

In the section “[Methodological Background](#)”, we provide an overview of the theoretical and methodological frameworks that have informed RAVE. We also give a brief account of early experiments that inspired the project (section “[Early Experiments](#)”) and finally, in section “[RAVE—The Overall Model](#)”, we illustrate the overall methodological model we use. The aim of Sect. 2 is to give a chronological account of different aspects of our methods development to date. In section “[Overall Development Design](#)” we give an overview of the overall Action Research process we apply in the project. Section “[Identifying Focus](#)” discusses some of the theoretical basis for our choices. In section “[Planning, Building, Testing and Modifying the Cases](#)” we describe aspects related to various processes in the construction of our “cases”, the contextually adapted matched-guise experiment designs that we use in classroom contexts. Sections “[Delivery—Packaging](#)”, “[Ethical Issues of the Project](#)”, “[Reliability of Data—Obtaining Baselines and Dismissing Unwanted Variables](#)”, “[Debriefing](#)”, and “[How to Identify and Measure Awareness Raising](#)” deal with other important aspects of our method development such as digital packaging (Delivery—Packaging), ethical issues (Ethical Issues of the Project), securing reliable data (Reliability of Data—Obtaining Baselines and Dismissing Unwanted Variables), the framework for the classroom discussions that follow the experiments (Debriefing Debriefing) and the challenges involved in measuring whether awareness raising has taken place or not (How to Identify and Measure Awareness Raising). Finally, in Sect. 3, we summarise our findings and also look ahead.

Methodological Background

People often use stereotypical preconceptions ascribed to certain social groups when they attribute characteristics and traits to individuals, and in the defining of social belonging, aspects of speech, such as voice qual-

ity, dialect, and word choice, are major triggers (Johnson, 2000; Lippi-Green, 1997). Such stereotype-based categorisations are deeply embedded in our social makeup and affect how we interpret and process speech and language (Hay et al., 2006; Johnson et al., 1999; Niedzielski, 1999). Even if there is substantial documented knowledge about this phenomenon, we would argue that there is limited awareness among most of us about how these structures affect our own judgments and actions. This state of affairs motivates a shift of research focus from identifying linguistic dissimilarities between different social groupings, to exploring *beliefs* about the language behaviour of different social groupings and how these beliefs in turn affect our interpretations of “reality”.

The systematic enquiries into linguistic stereotyping began roughly half a century ago with Lambert et al. They proposed that even brief samples of speech varieties (e.g. accent, intonation, and minority language) associated with a certain social group can affect how an individual is judged on traits related to behaviour, personality, social status and character (Bradac, Cargile, & Hallett, 2001; Lambert, Hodgson, Gardner, & Fillenbaum, 1960). In order to test this hypothesis, the so-called matched-guise technique was developed. In a matched-guise set-up, the same text (normally spoken) is produced in two or more variants, where the manipulated variable is the perceived identity of the speaker as manifested through language—social or regional accent, for example. In the original set-ups, one bilingual or polyglot actor or actress would be used to produce different variants of a spoken text. The text was then played to respondents and the reactions elicited by each of the linguistic guises were compared. The method thus served to show how a speaker’s accent, speech patterns, intonation, etc. can serve as markers which respondents’ evaluation of speaker’s behaviour, personality, social status and character. In short, the experiment revealed how stereotypical language attitudes are used to evaluate speakers. The matched-guise test is still widely used today in social psychology, sociolinguistics, business research and medicine (Buchstaller, 2006; Cargile, 1997; Carson, Drummond, & Newton, 2004). One major critique of the method, however, has been that it is almost impossible to produce two texts where the only variable that differs is the accent, even when the same actor or actress is used. Speed, intonation, or pitch can have a significant impact on how something is

perceived (Tsalikis, DeShields, & LaTour, 1991). Also, if the same actor or actress is used, it is almost impossible to manipulate gender in a convincing manner, a fact that has excluded this social variable from matched-guise set-ups to date.

Early Experiments

The methods used under RAVE were initially inspired by experiments we conducted in virtual worlds, more specifically Second Life, where we were working with various pedagogical language learning projects at the time. The introduction plug-in software “MorphVOX pro” meant that you could alter your voice quality at the click of a button from, say, female to male and you could thus move around in the virtual world in a guise of the opposite sex. The idea of using this affordance for pedagogical purposes, in courses that dealt with issues of language and gender, was born. Over the period of one year, we experimented with various models based loosely on matched-guise designs in our language classes. These initial experiments highlighted challenges related to ethical issues, as well as the design of the learning activities so that focus was kept on what we wanted to illustrate. Ensuring reliability of the data and technical aspects related to voice-morphing were added difficulties that we became aware of.

Our current methods are inspired by classical matched-guise set-ups but instead of using different actors or actresses, we use digital software to manipulate the same recording in different ways. Using digital manipulations of recordings does not only allow us to eliminate influencing variables such as speed or intonation, but it also enables us to perform matched-guise tests with focus on gender without using two actors or actresses. We can thus filter out a great deal of noise from the data.

RAVE—The Overall Model

Using digital manipulation, the RAVE project seeks to explore and develop pedagogical methods for revealing sociolinguistic stereotyping among respondents with regard to identity-related properties such as

gender in order to foster a problematised view of language and stereotyping. The overall principles behind the methodological model discussed in this chapter are relatively simple (see Fig. 8.1 for an overview).

Based on a scripted dialogue between two virtual characters, let us say “Kim” and “Robin”, in which each character can be assigned presumed stereotypical properties, we produce a recorded dialogue—a “case”. Updating traditional matched-guise techniques with digital methods, we produce two property manipulated versions of the dialogue based on one singular recording. Thus, in one version, “Kim” may sound like a man, while the other recording has been manipulated for pitch and timbre so that “Kim” sounds like a woman. That there is a link between the perception of a voice as male or female and the trigger of stereotype inferences has been demonstrated in controlled studies. Ko, Judd, and Blair (2006) and Ko, Judd, and Stapel (2009), for example, could show that the perception of a voice as male or female functioned as an overall between-category source for gender stereotyping. Moreover, their research illustrated that voice quality itself (in terms of degree of femininity or masculinity) had some minor effects on within-category judgments when the gender identity of the person was known, but this did not match the overall effects of the gender perception of voices as male or female.

After detailing respondents’ background data on aspects such as gender and age in a pre-survey, they then listen to one of the two versions of the texts. Note that at this stage the respondents are unaware of the real purpose of the experiment and of the fact that there are two versions of the case. In an immediate post-exposure survey, the test subjects are asked to respond to questions related to linguistic behaviour and character traits of the interlocutors in the dialogue.

After analysis of the responses the class is then reassembled, and the design and real purpose of the case is revealed. Students can now see for themselves how the responses of the two groups differ. Our ambition here is to create what we call an “aha-moment”, that is, the realisation that we are all subconsciously affected by stereotyping in our interpretation of the social world around us. This subsequently constitutes the starting point for seminar discussions (a debriefing session). In using the students’ own results as a starting point for the debriefing, discussions of stereotypical categorisation acquire an additional and immediate urgency.

Fig. 8.1 Overview of the matched-guise set-up used in RAVE

After all, the results are based on their own judgments of what is essentially the same dialogue, where the only manipulated variable is the perceived gender of the speaker, as triggered by voice quality. Note that we confirm the gender perception of the voices as either male or female in the post-survey after the group reflection. As Chavez, Ferris, and Gibson (2011) have demonstrated, group reflection and careful use of probing questions make debriefing sessions like these into significant learning opportunities, and evaluations demonstrate that students value this experience positively. A post-survey provides feedback on the participants' experience and measures whether awareness raising has taken place employing qualitative and quantitative methods.

2 Method Development

Overall Development Design

Essentially our methods' development is based on an Action Research process whereby we work collaboratively as a research team with various stages of the process. These stages include identifying a focus based on previous research, planning, building, pre-testing and evaluating various constituents of the overall method, implementing the matched-guise simulation, and finally gathering and evaluating data in order to gain a basis for informed changes to the next cycle of the process in the method development (see Fig. 8.2).

The overall ambition is to create learning experiences that lead to maximum insights into how stereotyping affects our interpretation of language events, such as dialogues, debates, and statements, that surround us. Below we will give an account of some these processes so far in the project. These include identifying a focus, the production, framing and digital packaging of the cases, ensuring reliability of the data, the debriefing phase and measuring whether awareness raising has taken place or not. It should be noted that our account is restricted to the part of the project that deals with matched-guise experiments involving gender only.

Fig. 8.2 Cycles in methods development

Identifying a Focus

The purpose of our method is to raise awareness of how stereotypical gender expectations affect our interpretation of linguistic behaviour, and it has thus been essential to identify a focus on exactly which stereotypical features we are interested in and want to expose. In so doing, we have used the general concepts *hegemonic masculinity* (Connell, 1987, 2005) and *hegemonic femininity* (Connell & Messerschmidt, 2005; Pyke & Johnson, 2003; Schippers, 2007) as starting points. According to Connell, hegemonic masculinity is the image of masculinity which is, or at least has been, culturally normative and something everyone must relate themselves to. The characteristics and behaviours of hegemonic masculinity include: competitiveness, stoicism, courage, toughness, risk-taking, adventure and thrill-seeking, violence and aggression, as well as achievement and success

(Donaldson, 1993). Important to note, however, is the fact that although hegemonic masculinity is the "ideal", it is not necessarily the most common form of masculinity. While hegemonic femininity and hegemonic masculinity are similar in the sense that both are dominant gender constructions, they are dissimilar as hegemonic femininity does not hold dominance over masculinity. Schippers defines hegemonic femininity as "the characteristics defined as womanly that establish and legitimate a hierarchical and complementary relationship to hegemonic masculinity" (Schippers, 2007, p. 94). It is characterised by traits such as submissiveness, cooperativeness, and meekness. Again this is an idealised image of femininity and there are many examples of so-called pariah femininities that embody characteristics of hegemonic masculinity (aggression, competitiveness and authority) (Schippers, 2007). In summary, gender is a continuum rather than a dichotomy. For our purposes, we would argue, however, that gender stereotypes are based on the hegemonic models of masculinity and femininity. So how do these gendered models translate into stereotypic views on gender and language behaviour?

Although voices have been raised in favour of a "gender similarities" approach to sociolinguistic research, that is, supporting the hypothesis that holds that men and women are more alike than they are different in their language behaviour (Hyde, 2005), much research into gender and language to date has been informed by a "gender difference" approach (Gilligan, 1982). Accordingly, considerable effort has been expended to find gender differences rather than similarities in language behaviour. According to Kaiser, Haller, Schmitz, and Nitsch (2009), this approach inevitably leads to the detection of differences rather than similarities. For example, according to sociolinguists such as Cheshire and Trudgill (1998, p. 3), women and men have a statistical preference for different conversational styles. Women have a tendency to communicate in a manner that supports other speakers and signals solidarity, whereas men, on the other hand, use a number of conversational strategies that can be described as a competitive style, stressing their own individuality and emphasising the hierarchical relationships that they enter into with other people. Thus, according to Coates (2004, p. 126), when men converse they tend to seek power while women's style is based on support. We would argue, however, that while some data shows that men's language is competitive, and wom-

en's collaborative, we must also recognise that other data shows the opposite. In other words, individuals, both men and women, can use more than one speech style. Nevertheless, we would argue that the two speech styles "competitive" and "collaborative" are firmly gendered in the minds of many students.

These two speech styles described by Cheshire and Trudgill, and various others (see Coates, 2004; Holmes, 1995; Sunderland, 2006; Tannen, 1990, for example), are characterised by some key linguistic features. For the collaborative speech style, these include politeness and signalling interest in the other speaker through minimal responses, taking limited floor space and inviting the conversational partner into conversation, not asserting one's opinion, for example by signalling mitigation using hedging (*I think...*, *it might be...*, *probably...*), and complimenting the other speaker. Competitive speech features include taking a lot of conversational space, being forceful in the assertion of one's opinions, and contradicting and interrupting the other speaker. Important to note is that we are not claiming that these conversational styles actually are typically masculine or feminine behaviours, although much research effort has been devoted to trying to assert this. Rather we hypothesise that these claims, especially as formulated in textbooks, run the risk of inadvertently shaping stereotypical expectations on how men and women behave (or should behave) in conversations. There is thus a danger that research not only serves to confirm stereotypes but also runs the risk of legitimising them.

Linguistic features identified as typical for competitive and collaborative styles were the starting point for the building of our cases. They constituted the key areas of focus in the construction of our case conversations, and they were also the features we wanted to draw the students' attention to in the response questionnaires and in the debriefing sessions. This is of course deliberate, and we would argue that we cannot expose students to their own stereotypes without evoking them.

Planning, Building, Testing and Modifying the Cases

The planning of the cases has been guided by certain principles. First, it has been important to create cases that lead to large and reproducible dif-

ferences in response patterns between the two respondent sub-groups. Without this, the debriefing may not result in an “aha-moment”, and impact of the awareness raising design may be compromised. Second, it has been important that we create cases that we can justify and contextualise in believable (but actually false) learning contexts so that students do not suspect the real purpose of the experiment prior to the debriefing. Third, it has been essential that the students do not suspect that the recordings are manipulated, at least not as far as the gender identities of the speakers are concerned. Finally, it has been important to our design that the cases are of optimal length—long enough to accommodate the linguistic features we want to highlight and to give the respondents a chance to form an informed impression of the speakers, but not so long that respondents risk losing interest and focus.

Initial Attempts

In the first run of the methods-testing cycle, we opted to work with dialogue. Many of the stereotypical features we wanted to highlight could only be illustrated in a dialogue or multilogue. We also decided to create a conversation that was unbalanced in terms of styles. One speaker, Robin (note the gender-neutral nature of the names), was primarily collaborative in his or her conversational behaviour, while the other (Terry) was more competitive. The conversational topic was language and gender, and the conversation was supposed to represent an authentic recording of a discussion between two researchers. Our plan was to contextualise this as input for future discussions in sociolinguistics classes. Accordingly, the students were told that they would be asked questions on language pragmatic features in the conversation after they had listened to it (a topic also dealt with in the sociolinguistics class). We then worked in several language features that were typical for competitive and collaborative styles in the script of the conversation. For example, quantitatively the conversation was dominated by “Terry” who occupied 66% of the floor space. Terry also produced 86% of the interruptions and used forceful language, contradicting Robin on several occasions using several expletives. Robin, on the other hand, occupied far less floor space and was generally a better

listener, producing 81% of the supportive moves. Robin was also less assertive producing 64% of the hedges.

When we recorded the conversation, we used the same actor for both characters and then altered this voice to sound either more masculine, or more feminine depending on which version we were producing. Each supposed speaker in the conversation (Terry and Robin) was recorded on separate tracks and then MorphVOX pro software was used to manipulate the voices, thereby producing two versions of the original recording (see Table 8.1).

In this way, we produced one version that answered to hegemonic norm values (competitive male and submissive or collaborative female) and one version which represented alternative behaviours (competitive female and submissive or collaborative male).

The voice qualities of the manipulated voices, however, proved unsatisfactory when quality checked with colleagues. So, we decided to “camouflage” the shortcomings. Rather than saying that the recording was done in a face-to-face context, we reduced the sound quality further and claimed that it was a skype recording of poor quality. This reduced the negative reactions to the recording in the pre-quality checks significantly.

Our hypothesis at this stage was that respondents would react strongly to the behaviours of the alternative version (see Table 8.1) noticing features that did not match hegemonic norms. In the response questionnaire, we listed various linguistic features and asked the respondents to decide what proportion of these was produced by Terry (competitive) and Robin (collaborative). We also included a number of statements on traits and asked respondents to agree or disagree with these (5 point Likert scale where 1 was disagree completely).

When testing this model on our respondents (24 teacher trainees) we learnt several important lessons: first, the results from the response ques-

Table 8.1 Versions first case trial

Original	Version 1: Hegemonic norm	Version 2: Alternative norm
“Terry” (competitive)— Actor A	Lowered pitch = Terry male	Raised pitch = Terry female
“Robin” (collaborative)— Actor A	Raised pitch = Robin female	Lowered pitch = Robin male

tionnaires did not support our hypothesis—respondents did not take note of language behaviour that contradicted their stereotypical expectations. In fact, the opposite was the case. Respondents seemed to notice and overestimate behaviour that confirmed the stereotypical hegemonic discourse. Some of these tendencies are illustrated in Table 8.2.

As illustrated in the examples listed in Table 8.2, competitive hegemonic masculine language behaviour, such as taking space and interrupting, was overestimated by respondents who listened to the male guise of Terry (the competitive speaker), while language behaviour associated with collaborative style (hegemonic feminine), such as hedging and supportive listening, was underestimated. In addition, Robin was deemed as more sympathetic when speaking in the female guise than in the male guise, while Terry was seen as less sympathetic when speaking as a male guise. In short, it seemed that respondents noticed what they were looking for. Note, however, that the differences above were not large enough to be statistically significant given the limited number of respondents.

Second, we had serious problems with “believability”. Post-surveys revealed that some students had suspected that the voices had been manipulated for gender. One student even suspected that it was the same speaker who had produced both voices. In addition, many respondents questioned the authenticity of the recording. This was not a satisfactory result.

Third, the set-up was deemed too complex. Estimating proportional use of different features was simply too difficult, and many respondents

Table 8.2 Real and estimated impressions of some aspects of Terry’s (competitive speaker) conversational behaviour

Tendency	Floor space	Interruptions	Hedging	Supportive moves
Proportion of Terry (competitive speaker) in the actual recording	66%	86%	34%	19%
Estimates of group who listened to the male version of Terry	75%(+)	92%(+)	22%(–)	17%(–)
Estimates of group who listened to the female version of Terry	63%(–)	76%(–)	40%(+)	30%(+)

claimed that they just had guessed at random. They were not convinced that the results represented a real measure of their impressions. This had a negative impact on the “aha-effect” we were after. Similarly, asking students to focus on both speakers in the conversation was deemed to be distracting and confusing. In addition, the conversation was too long. Many respondents claimed that they could not maintain close focus for six minutes.

Further Trials—Modifications Based on Initial Lessons

The first trials of the method revealed a number of weaknesses and uncertainties in our design:

1. In our trials respondents seemed to especially notice features which confirm their stereotypical preconceptions. In other words, typically masculine behaviour was more noticeable when the respondents thought they were listening to a male speaker and vice versa. Was this tendency strengthened by the nature of the conversation (unbalanced)? This was something we wanted to explore further.
2. We needed to work on the “believability” of the cases. The quality of the voice morphing had to be improved and the feeling of authenticity had to be improved.
3. The case design had to be simplified to make it more focused on the aspects we wanted to highlight.
4. We needed many more respondents to confirm that the differences we saw in fact were “real” differences.

In order to see if the nature of the script affected responses, we decided to test different script structures. We produced two balanced dialogues—one where both speakers adhered to a more collaborative style and one where both were more competitive. The two speakers in the dialogues occupied equal floor space and used similar numbers of linguistic features typical for collaborative or competitive speech styles.

Furthermore, the quality of the voice morphing and the believability of the case needed improvement. In the new production we systemati-

cally tested different voices to see how they responded to the morphing. It turned out that some voices were much better suited than others producing more believable, less artificial sounding recordings. We produced test recordings, morphing 12 different voices, and we sent these out to 25 peers asking them whether (1) the recordings sounded natural and (2) whether they sounded convincing as male or female voices. Based on these responses, we chose the voices (actors or actresses) that were evaluated most positively. When producing the recordings, we also decided to use different software for the voice manipulations. They were first recorded using Avid Pro Tools HD 12.0.0 and then edited in the same software. Pitch shifting was processed manually with X-Form (Rendered Only) using Elastic Audio properties in Pro Tools.

The idea of contextualising the cases as authentic recordings was abandoned. Instead, we decided to present them as reproductions or representations of genuine conversations. This, we deemed, would not affect what we were testing but eliminated the risk of students reacting to the fact that the conversations may not have sounded entirely natural and thereby suspecting that something was afoot.

In order to simplify the cases, the conversations were shortened (from six to four minutes), and respondents were asked to focus on one of the speakers in the conversation only. Moreover, the response questionnaires were simplified to include simple statements on a more limited number of dimensions, five related to speech style and two related to characteristics or personality, which the respondents could rate on a seven-point Likert scale ranging from 1 (disagree completely) to 7 (agree completely).

In order to increase the number of respondents, we decided to run the experiment in several classes (four in all). To test if there were any "real" differences among the response groups, we also tested the recordings with respondents we "bought" from SurveyMonkey. SurveyMonkey found random respondents from the UK (Sweden was not an option here) with an age limit of 45 aiming for a 50/50 split between males and females. This group comprised 101 individuals in all, 48 males and 53 females. In all, we tested 170 respondents, a number which was deemed to be large enough to statistically confirm any differences in responses.

The results from these trials were encouraging. There were significant differences between how the guises were rated when it came to floor space

and contradictions. Interruptions approached significance. The competitive speech variables (interruptions, floor space, contradictions and forcefulness) correlated with each other, as did the collaborative (taking little floor space, signalling interest and sympathy). Using various multivariate analyses (MANCOVA, ANOVA, and ANCOVA. MANOVA) that took aspects such as the gender of the respondents into account (see section on Reliability of Data), we were able to confirm statistical differences between how the guises were rated. The female guises were rated higher on the collaborative variables and the male guises were rated higher on the competitive speech variables. These effects were particularly evident in the balanced, collaborative dialogue. We have been able to reproduce these results in subsequent trials, and the hypothesis that respondents notice behaviour that matches their stereotypical expectations and ignore behaviour that does not seems to hold. This adds to the credibility of the whole set-up, thereby increasing the impact of the “aha-moment” in the debriefing.

Delivery—Packaging

One aspect not dealt with above is the delivery and packaging of the cases. In the early attempts, we worked with text and oral instructions, separate sound packages, survey link entry points, etc. in learning management systems that the students had access to. This model proved to be “messy” in that it was not self-evident in which order and how things had to be done. We continually had to instruct respondents and differences in instructions could in turn affect the behaviour of the respondents.

We thus started exploring models for standardising the method by “packaging” all information from a one-point entry principle. This was also necessary in order to access respondents who were not physically present. The software Articulate Storyline (see <https://articulate.com/360/storyline>) has met many of our needs. It allows us to integrate instructions (text and oral recordings), pre-surveys, sound files with relevant illustrations (see below) and response surveys in one package which is accessible from a single URL link. Articulate Storyline has the added advantage of being flexible in terms of delivery mode—it can be accessed

Fig. 8.3 Drawing attention to one of the conversational participants using iconic symbols. Note how “Robin” is represented as male in one version and female in the other version

from a mobile phone, tablet or computer. In this way, we ensure that the packaging of the experiment does not become a variable that interferes with the results. An added benefit of this is that we can deliver the method to various groups outside university contexts.

In the latter trials, when respondents were asked to focus on one speaker only, efforts have been made to eliminate the risk of the respondents focusing on the wrong speaker. For this purpose, we have used iconic symbols to draw attention to a certain speaker in the conversation. Figure 8.3, for example, illustrates how attention is drawn to the speaker “Robin” using an image of the speaker, a name tag and a speech bubble.

More recently we have opted to use silhouette images only of the speakers to eliminate the potential of facial features, etc. affecting the respondents. Silhouette images have proved sufficient to signal the gender of the speaker.

Ethical Issues of the Project¹

Ensuring anonymity while at the same time creating a system whereby we can follow the respondents in different phases of the process (pre-survey, response survey and post-survey) is essential for the design of our method. If respondents are not confident that they are anonymous, they may not

¹ The methods of the project have been approved by the Swedish National Ethical Vetting Authority.

answer honestly. At the same time, we need to be able to track their responses. Accordingly, we have devised a system whereby respondents create a seven-digit code based on personal information which we have no access to (first letter of your mother's first name, for example). In this way, respondents can recreate their code easily even if they should forget it. We also make sure that we avoid working with groups of less than ten individuals where the identity of the respondents can be jeopardised.

Informed consent is another issue that we have had to consider. Obviously, we cannot inform the respondents about what we do prior to the matched-guise treatment. Instead, we give the chance for participants to withdraw their responses from the study in the post-survey. Results from participants who do not do the post-survey are not included.

Reliability of Data—Obtaining Baselines and Dismissing Unwanted Variables

Our method builds on splitting a group into two sub-groups which respond to different versions of the recorded dialogue. In this design, it is important to control for unwanted variables that may affect the results. In other words, we need to ensure, as far as possible, that the observed differences in responses are a direct result of the voice or identity morphing of the recordings and not imbalances in the make-up of the response groups. Aspects such as age, gender and cultural or national identity of the respondents are included in the pre-survey and can be controlled for in the statistical analysis of the results as potential variables affecting the results. In addition, if, as we hypothesise, the respondents' stereotypical preconceptions act as a filter which draws selective attention to some language features, then potential differences in preconceptions between the response groups also need to be taken into account. In an attempt to control for this variable, we have included three measures in the post-survey. Two established tools from social psychology, namely the Modern Sexism Scale (Ekehammar, Akrami, & Araya, 2000) and the Ambivalent Sexism Inventory (Glick & Fiske, 1996), attempt to measure sexism among the respondents. Both tests consist of a number of statements such as "Discrimination of women is no longer a problem", which the

respondents can agree or disagree with on a five-point Likert scale. Low values indicate little sexism. We have also started devising our own measure, the Linguistic Stereotyping Inventory, which we include in the post-survey. In this test we list a number of linguistic tendencies and ask respondents to rate these on a scale ranging from typically male (−2)—neutral (0) to typically female (+2). A typical response pattern among Swedish students is illustrated in Fig. 8.4.

Using these inventories, we can compare the nature and strength of the stereotypical preconceptions that the response sub-groups may have. So far, however, our experiences indicate that Swedish students hold fairly similar stereotypical pre-conceptions regarding sexism and language behaviour but we need to explore this further by refining the linguistic inventory and further testing the model on more heterogeneous groups.

Controlling for background variables is motivated. Using ANCOVA and MANCOVA multivariate analyses on the response data retrieved so far indicate that the gender of the respondent seems to affect how they

Fig. 8.4 Typical response pattern to Linguistic Stereotyping Inventory: (−2) indicates typically male; (0) indicates neutral; (+2) indicates typically female

rate some of the linguistic features of the guise (floor space and the signalling of interest). In addition, the differences found in the rating of guises can be partially explained by the respondents' sexism. Here, it is difficult to elucidate what is a gender effect and what is a sexism effect since these have a strong tendency to correlate (males are more sexist than females). So far, we have not been able to establish any effects related to linguistic stereotyping but this is mainly due to the fact that results here have been quite homogeneous. Pilot studies in other cultural contexts (the Seychelles, for example) do suggest that this is a very important background variable, however.

Debriefing

Debriefing is a pedagogical activity that has long been employed in social sciences such as psychology and medicine, where it has been used in educational contexts involving simulations, or other experiential teaching methods. Since the RAVE project essentially builds on a simulation activity, a debriefing session became a natural part of the design. More importantly, as the learning opportunity *par excellence* in the RAVE set-up, the nature of debriefing session is crucial for the learning outcome.

The internet site Debreifing.com describes the term *debriefing* as referring to “conversational sessions that revolve around the sharing and examining of information after a specific event has taken place”, and it can take many shapes as shown by Lederman (1992) and Dreifuers (2009), for instance. Two contexts, listed by Lederman (1992), in which the concept has been developed are particularly relevant to our setting. First, there is the use of debriefing in psychological studies in which participants were deceived in one way or the other in an experiment. The debriefing here serves the main purpose of telling participants what the experiment was all about. As pointed out by Lederman, the designers of such an experiment become the debriefers, who are in a powerful position, which may lead to them promoting their own explanations and labels at the expense of any ideas the participants may have. The second context for debriefing is an educational setting, in which debriefing is used after an experiential activity of some sort. The purpose here is to

help participants reflect and learn from their experience. Given our ambitions to raise self-awareness, the latter context is of primary interest in our method.

The procedure of the debriefing session in the RAVE project has a design which largely follows the steps outlined by Lederman (1992, pp. 151–152). Lederman lists three main phases in the debriefing: (1) Systematic reflection and analysis; (2) Intensification and personalisation; and (3) Generalisation and application.

In the RAVE application of this model, in phase (1) we start by organising students into groups of four or five individuals belonging to the two groups that experienced different simulations. Then the nature of the simulation is described, giving students a possibility to recollect what happened using brief snippets from both simulations, and, after this, the results from the response questionnaires are presented. We show how the two groups responded to the statements, and the attention of the participants is directed towards aspects where the greatest differences have been recorded. This step leads us over to the second phase in which the groups are encouraged to discuss their own experience of the simulation and possible explanations of the recorded differences in responses. The second step here leads to the third phase in which generalisations and implications are discussed. We partially use prepared questions in order to help students focus on important aspects—"what implications might these results have for your future professional roles?", for example. It is in this phase that we hope to achieve an "aha-moment" by developing an understanding of the simulations and their results and a transformation of this into metacognitive awareness of more general and applicable consequences of stereotyping. Further, rather than ending the debriefing with a discussion of "what did you learn from this experience?", where a few participants might formulate answers which might affect several students, we have this explicit question in the post-survey which is answered individually with the help of a digital tool at the very end of the session. It is with the help of this post-test that we hope to record and measure a possible change in awareness of the influence of stereotypes. Typically the class debriefing is achieved in a double lesson (2×45 minutes). The post-survey is administered individually and students are requested to answer this survey within a week after the debriefing.

In reframing the simulation as a case making no claims about its authenticity, we have been able to play down the deception aspect thereby aligning the goals of the debriefing activity more closely to that of educational settings. While still in more control than in classic educational debriefings, we try to adapt a position of guiding discussions rather than providing answers. For example, participants are encouraged to discuss and reflect on the results of the experiment in small groups comprising participants from both “treatment groups”. In common with Lederman’s (1992, p. 149) characterisations of debriefings in educational settings, we have created a setting where the teacher knows what kind of reaction the experiment was meant to generate, but where we are nevertheless interested in learning about the participants’ experiences and understandings. In this way, we are also able to help them go from the specific to the generic with the aim to raise self-awareness.

How to Identify and Measure Awareness Raising

The debriefing session has a crucial position in the RAVE set-up as the learning arena. The goal of the activity is to raise awareness about how stereotypical preconceptions affect our judgements in everyday contexts. An important aspect of the project is thus measuring success or failure in reaching this goal. The literature on awareness raising is quite barren in terms of discussions regarding how to measure raised awareness. In the development of the RAVE project, we have tried different methods, which are briefly described below. This is an area which needs further development as will be clear from the descriptions.

Initially the idea was to use the established Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998) to measure awareness raising. It is a test based on reaction time which is designed to measure the strength of unconscious associations between concepts, in our case certain linguistic behaviour and gender. The participating students performed an IAT test designed for this special purpose immediately before the RAVE simulation and after the debriefing. Because the IAT is designed to measure indirect and unconscious associations, we were from the very start concerned that it would not be a good tool for measuring awareness

raising. We believed, however, that it could at least serve as a reference point, a baseline, for the interpretation of the results. As it turned out, the IAT was problematic. For example, it was difficult to make sure that the circumstances were identical for all participants on the occasions they performed the IAT; since the test builds on reaction times for the presented associations, it is important that there are minimal distractions. This was difficult to achieve and meant that we had to work in computer labs, thereby excluding students working from a distance. The IAT was also quite time consuming and required full concentration. We could notice that students were less concentrated when performing activities following the IAT session. In addition, the results from the IATs turned out to be inconclusive thereby providing a baseline of questionable validity. We thus decided to drop this test. Furthermore, given our ambition to create an online packaging of our methods with easily interpretable feedback (see Summary and Looking Ahead below), there were more aspects favouring another solution.

We needed a reliable but far less exhaustive tool for measuring any change in awareness. An important aspect of the tool was also that it should not in any way prime the participant for the answer we wanted. Initially, we tried a solution based on an open-answer question, where the participants were invited to provide five explanations for the language behaviour of a speaker in a fictive scenario. The test was given before and after the experiment and care was taken so that there was nothing to suggest what the test was really meant to measure. The hope was that the answers could be categorised in such a way that we could see some progression in terms of awareness of how our own interpretation can colour our experience of an event. We wanted to see whether there was any systematic drift in the nature of the explanations provided by the participants in the pre-test and the post-debriefing tests. Here, we were looking for answers that included perspectives where participants' own interpretations of the event were questioned. For example, rather than saying that "John is pushy in the conversation because he is a man", we were hoping that the post-tests also would include explanations such as "I experience John as pushy because this is the sort of behaviour I expect from a man". As it turned, out there was some change in this respect. However, it was

too small to be significant, and more importantly, the categorisation of the answers was laborious and often difficult.

After the above attempts, a simplified model which combines quantitative and qualitative measures was opted for. In the post-survey, we simply ask the open-ended questions “What was your general experience of the experiment that you have just partaken in? Did you learn anything new?”. These questions are in line with what Lederman (1992) suggests for debriefings and can be analysed qualitatively. In addition, we also attempt to get a quantitative measure of awareness raising that allows us to compare the responses of each participant in the pre- and post-surveys. Here, we use a 0–100 Likert scale response option to the question “To what extent do you think that you are influenced by stereotypical preconceptions (conscious or unconscious) in your expectations and judgements of others?” Note that this question is “hidden” among dummy questions in the pre-survey in order not to draw attention to the real purpose of the activities. Our hypothesis is that the matched-guise experiment and the debriefing will make the participants more aware that they, in fact, are affected by stereotyping and that the values generated in the post-survey will be greater than in the pre-survey. The quantitative measure can also be triangulated with the qualitative data for each respondent adding to reliability.

Results so far have, however, been difficult to interpret. A cross-comparison between the qualitative and quantitative data reveals that there is considerable variation. For instance, a closer examination of 40 positive responses to the qualitative questions, which suggest that respondents have become more aware, and their corresponding responses to the quantitative measure suggest that the latter cannot be interpreted in a straightforward way. The average result of this group showed an increase by 6.9 units in the quantitative measure, thus indicating, in keeping with their comments, that they have increased their awareness that they are affected by stereotypical preconceptions. This result, however, is not statistically significant primarily due to a very large standard deviation, 17.90. A concrete example can provide some insights into problems with the measures. One student wrote the following comment: “I found it to be interesting, and it gave me food for thought. Even though I believed myself to be relatively free of prejudice I can’t help but wonder if I make

assumptions about personalities merely from hearing someone's voice”. It was apparent from the comment that the student had become more aware of the influence of stereotypes as a result of the simulation and the debriefing session. Yet, this same participant's indication in the quantitative measure gave a negative result of 35 units. In other words, the answer suggested that the participant in question was less aware of how he was influenced by stereotypical preconceptions in his judgements of others after the experiment than before the experiment. One way to interpret this paradox is that some participants may use the Likert scale to indicate that with an increased awareness of stereotypes, they actually become less affected by these. This is another illustration of the difficulty of measuring raised awareness in this context in a simple numeric way, a challenge we are still battling to resolve.

3 Summary and Looking Ahead

The RAVE project has an action research-like format in which we repetitively revisit a cycle of method development with the aim of improving it (see Fig. 8.2). This chapter has focused on this process. For example, initially, the tools we used for digitally manipulating the recordings were rudimentary, but with continuous trialling, based on the feedback of colleagues, reflection and explorations of new methods, we have now improved methods so that the manipulations sound believable in the contexts they are used. The contextualising of the cases has also undergone changes. Originally, we presented the recordings as authentic conversations between two researchers, whereas we now present them as reproductions of real conversations thereby decreasing the risk of respondents suspecting the real nature of the manipulations. We have rationalised the packaging of the method and we have also simplified the model, by instructing the respondents to focus on one speaker only, for example. Here, we also use various iconic signals to guide respondents. These measures have led to improved results and we are now in a position where we achieve reproducible statistically significant differences between the responses to the two manipulations. This, we argue, is a prerequisite for the achievement of the “aha-moment” leading to awareness raising.

With that said, there are still many parts of the method where there is room for improvement. For example, we are still battling with how to measure awareness raising in an effective way. Previously trialled methods have been time consuming, unsuitable for online data accumulation, complicated and unreliable. Our current method has issues too, as highlighted in the previous section. There are indications that the questions asked are open to misinterpretation, which makes analyses of the quantitative data difficult, as some respondents report a negative value when they most likely mean a positive value.

Furthermore, we need to develop the method for new implementations on more diverse arenas. We will end this chapter by pointing to three such areas of expansion, which will generate further development as well as interesting results. First, we have recently been implementing the model in other learning contexts. Awareness regarding the effect of stereotypes is arguably relevant in many professional practices and key knowledge for anyone working with human interaction. Thus, an expansion of the project into other people-oriented professions was always an ambition. In line with this ambition, we have been implementing the RAVE model in a course of social psychology where it is used in the context of a module on personality traits. In this context, the model is used to provide students with an opportunity to consider how their perception of a person's gender in a dialogue can affect their understanding of that person's personality traits. The implementation has required simulations of a different character, taking personality features into consideration and not only conversational features, and has also led to further pedagogical discussions with the teachers about how to smoothly integrate the model into students' course work and how to debrief with the best possible result. Thus, this collaboration is stimulating the development of packaging and the measurement of awareness among other things.

Another interesting area, which was also always part of the project's original ambition, is the testing of the model in other cultural contexts. Since stereotyping is a cognitive phenomenon that can be described as an automatic and reductive categorisation of groups of people, it follows that cultural contexts can have great influence on stereotypical assumptions. Although gender has been studied in many different cultural con-

texts, and occasionally with contrastive purposes, studies using the method described here (matched-guise techniques for gender) have not previously been conducted. With this in mind, we are currently piloting the collection of comparable material in the Seychelles, where we have good contacts and where we have been given access to language teacher-training classes. This implementation does not only provide us with new and exciting data but also motivates further development of the model and the pedagogy. For instance, it has been observed (Chung, Dieckmann, & Issenberg, 2013) that debriefings may have to be adjusted to suit participants in different cultural contexts. Thus, a model well-suited for students at a Swedish university may well leave much to be desired elsewhere. Moreover, in a context such as the Seychelles, it is not necessarily the case that online solutions work well, so, for a general applicability, it has become apparent that the project requires "low-tech" backup solutions.

Looking further ahead, a primary ambition is to create an open-access online resource based on the methods developed within the project, which, in its fully functional state, would be a resource that could be used for awareness-raising activities in various learning environments, including contexts where low-tech solutions are needed. These ambitions require further development and adaptations of the method in close contact with technical staff with the necessary knowledge of how to create interactive and intuitive online facilities easily accessible for users, while at the same time considering how low-tech adaptations of the same can be designed. Interesting times lie ahead.

References

- Bradac, J. J., Cargile, A. C., & Hallett, J. S. (2001). Language Attitudes: Retrospect, Conspect. In W. P. Robinson & H. Giles (Eds.), *The New Handbook of Language and Social Psychology* (pp. 137–158). Chichester: John Wiley.
- Bucholtz, M., & Hall, K. (2005). Identity and Interaction: A Sociocultural Linguistic Approach. *Discourse Studies*, 7(4–5), 585–614.
- Buchstaller, I. (2006). Social Stereotypes, Personality Traits and Regional Perception Displaced: Attitudes Towards the "New" Quotative in the UK. *Journal of Sociolinguistics*, 10, 362–381.

- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Cargile, A. (1997). Attitudes Toward Chinese-Accented Speech. *Journal of Language and Social Psychology*, 16, 434–443.
- Carson, L., Drummond, J., & Newton, J. (2004). Social Perception in the Clinical Dental Encounter: The Matched-Guise Technique Re-visited. *Psychology & Health*, 19, 667–683.
- Cavallaro, F., & Ng, B. C. (2009). Between Status and Solidarity in Singapore. *World Englishes*, 28(2), 143–159.
- Chavez, C. I., Ferris, W. P., & Gibson, L. A. (2011). Predicting Values and Beliefs: Learning That Can Last. *Journal of Management Education*, 35(4), 504–536.
- Cheshire, J., & Trudgill, P. (1998). Section I: Introduction. In J. Cheshire & P. Trudgill (Eds.), *The Sociolinguistics Reader: Volume 2: Gender and Discourse* (pp. 1–6). London: Arnold.
- Chung, H. S., Dieckmann, P., & Issenberg, S. B. (2013). It Is Time to Consider Cultural Differences in Debriefing. *Simulation in Healthcare*, 8, 166–170.
- Coates, J. (2004). *Women, Men, and Language: A Sociolinguistic Account of Gender Differences in Language*. New York: Pearson Education Limited.
- Connell, R. W. (1987). *Gender and Power: Society, the Person and Sexual Politics*. Malden: Polity Press.
- Connell, R. W. (2005). *Masculinities*. Cambridge: Polity Press.
- Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic Masculinity: Rethinking the Concept. *Gender and Society*, 6, 829–859.
- Crawford, M. (1995). *Talking Difference: On Gender and Language*. London: Sage. Darcy.
- Donaldson, M. (1993). What Is Hegemonic Masculinity? *Theory and Society*, 22(5), 643–657.
- Dreifuerst, K. T. (2009). The Essentials of Debriefing in Simulation Learning: A Concept Analysis. *Nursing Education Perspectives*, 30(2), 109–114.
- Ekehammar, B., Akrami, N., & Araya, T. (2000). Development and Validation of Swedish Classical and Modern Sexism Scales. *Scandinavian Journal of Psychology*, 41, 307–314.
- Fuertes, J., Gottdiener, W., Martin, H., Gilbert, T., & Giles, H. (2012). A Meta-Analysis of the Effects of Speakers' Accents on Interpersonal Evaluations. *European Journal of Social Psychology*, 42, 120–133.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *Journal of Personality and Social Psychology*, 70(3), 491–512.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Hay, J., Warren, P., & Drager, K. (2006). Factors Influencing Speech Perception in the Context of a Merger in Progress. *Journal of Phonetics*, 34, 458–484.
- Higgins, T. E., & Bargh, J. (1987). Social Cognition and Social Perception. *Annual Review of Psychology*, 38, 369–425.
- Holmes, J. (1995). *Women, Men and Politeness*. London: Longman.
- Hyde, J. S. (2005). The Gender Similarities Hypothesis. *American Psychologist*, 60(6), 581–592.
- Johnson, F. L. (2000). *Speaking Culturally: Language Diversity in the United States*. Thousand Oaks, CA: Sage.
- Johnson, K., Strand, E., & D'Imperio, M. (1999). Auditory-Visual Integration of Talker Gender in Vowel Perception. *Journal of Phonetics*, 27, 359–384.
- Kaiser, A., Haller, S., Schmitz, S., & Nitsch, C. (2009). On Sex/Gender Related Similarities and Differences in fMRI Language Research. *Brain Research Reviews*, 61(2), 49–59.
- Ko, S., Judd, C., & Blair, I. (2006). What the Voice Reveals: Within- and Between-Category Stereotyping on the Basis of Voice. *Personality and Social Psychology Bulletin*, 32(6), 806–819.
- Ko, S., Judd, C. M., & Stapel, D. A. (2009). Stereotyping Based on Voice in the Presence of Individuating Information: Vocal Femininity Affects Perceived Competence but Not Warmth. *Personality and Social Psychology Bulletin*, 35(2), 198–211.
- Lambert, W., Hodgson, R., Gardner, R., & Fillenbaum, S. (1960). Evaluational Reactions to Spoken Languages. *Journal of Abnormal and Social Psychology*, 60, 44–51.
- Lederman, L. C. (1992). Debriefing: Toward a Systematic Assessment of Theory and Practice. *SAGE Journals*, 23(2), 145–160.
- Lippi-Green, R. (1997). *English with an Accent: Language, Ideology, and Discrimination in the United States*. New York: Routledge.
- Macrae, N. C., & Bodenhausen, G. (2001). Social Cognition: Categorical Person Perception. *British Journal of Psychology*, 92, 239–255.
- Messerschmidt, J. W. (2009). "Doing Gender:" The Impact and Future of a Salient Sociological Concept. *Gender & Society*, 23(1), 85–88.

- Money, J., Hampson, J. G., & Hampson, J. (1955). An Examination of Some Basic Sexual Concepts: The Evidence of Human Hermaphroditism. *Bulletin of the Johns Hopkins Hospital*, 97(4), 301–319.
- Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, 18, 62–85.
- Pecis, L. (2016). Doing and Undoing Gender in Innovation: Femininities and Masculinities in Innovation Processes. *Human Relations*, 11, 2117–2140.
- Pyke, K. D., & Johnson, D. L. (2003). Asian American Women and Racialized Femininities: “Doing” Gender Across Cultural Worlds. *Gender and Society*, 17(1), 33–53.
- Risman, B. J. (2009). From Doing to Undoing: Gender as We Know It. *Gender & Society*, 23(1), 81–84.
- Schippers, M. (2007). Recovering the Feminine Other: Masculinity, Femininity, and Gender Hegemony. *Theory and Society*, 36(1), 85–102.
- Sunderland, J. (2006). *Language and Gender: An Advanced Resource Book*. London: Routledge.
- Talbot, M. (2003). Gender Stereotypes. Reproduction and Challenge. In J. Holmes & M. Meyerhof (Eds.), *The Handbook of Language and Gender* (pp. 468–486). Oxford: Blackwell.
- Tannen, D. (1990). *You Just Don't Understand: Women and Men in Conversation*. New York: Morrow.
- Tsalikis, J., DeShields, O. W., & LaTour, M. S. (1991). The Role of Accent on the Credibility and Effectiveness of the Salesperson. *The Journal of Personal Selling and Sales Management*, 11(1), 31–41.
- Wallenberg, L., & Thanem, T. (2016). Just Doing Gender? Transvestism and the Power of Underdoing Gender in Everyday Life and Work. *Organization*, 23(2), 250–271.
- West, C., & Zimmerman, D. (1987). Doing Gender. *Gender and Society*, 1(2), 125–151.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.