On matching speaker (dis)guises – revisiting a methodological tradition

Barbara Soukup

University of Vienna [1]

Over the past fifty years, 'matched-guise'-type speaker assessment experiments have seemingly become the methodological bedrock of social psychological 'language attitude' research. In the study that pioneered this technique, Lambert, Hodgson, Gardner and Fillenbaum (1960) outline its basics as follows, in application to their own investigation of 'evaluative reactions' towards English and French in Canada:

> A 2 ½ min. passage of French prose of a philosophical nature was translated into fluent English and tape recordings were made of the voices of four male bilinguals each of whom read both French and English versions of the passage. Recordings were also made of the voices of two other men, one reading the passage in English, the other in French. There were, then, 10 taped voices, four of which were 'matched', each speaker using both languages, and two used as 'filler' voices and for practice. The 10 voices were presented to *S*s [subjects] in alternating French-English order starting with the two filler voices and allowing the maximum possible interval between successive presentations of the English and French guises of any speaker. Evaluational reactions to the matched voices only were examined. [...] *S*s were *not* told that they were going to hear some of the voices twice, but rather that they would hear 10 recorded male voices, all reading the same passage, five in French and five in English. [...] There was no indication that any *S* became aware of the fact that bilingual speakers were used. (Lambert et al. 1960: 44; italics in the original)

Lambert et al.'s artful protocol of having bilingual speakers read the same text in 'matching guises' was motivated by a desire to 'minimize the effects of both voice of the speaker and his message' on the assessment (Lambert et al. 1960: 44), and thus to keep language choice as the only experimental variable. The speaker and text being the same, any elicited rating differences between the English and French guises could presumably be attributed to respective differ-

[1] I thank the editors of this volume, Stefan Grondelaers and Tore Kristiansen, for their valuable feedback on previous versions of this chapter.

ences in 'language attitudes'.[2] Interestingly, however, the authors never explicitly state any reason for trying to keep their participants unaware of the speaker repetition.

Fifty years later, in the year 2010, I myself set out to conduct a research project on 'language attitudes' in the country of Oman towards Baluchi (an Indo-European minority language) vs. Arabic, using the matched-guise technique. At the time, however, I was faced with a recruitment dilemma that may sound familiar to some: on the one hand, I had access to only a handful of the bilingual informants I wanted to poll. On the other, I also still needed to pilot the experiment I was planning to do with them, but couldn't afford to 'lose' study participants this way. Ultimately, my female bilingual 'stimulus speaker'[3] and I decided to try out the experiment first with her own (adult) nephew, without telling him who the female 'voices' belonged to, my speaker having even convinced herself that he would not recognize her in the Baluchi guise (as they never actually spoke Baluchi together, and his competence in the language was mostly 'passive'). The experiment ran its course; the nephew never hesitated and turned in his ratings of the male and female Omani Arabic and Baluchi guises, via a classic response scheme of personality traits on semantic differential scales. Then, as we were going over the results and the study design together for feedback, he said quite casually, 'My aunt sounded like a totally different person when she spoke Baluchi that second time'. And indeed, this was reflected in his ratings – for example, he had put her down as sounding much less educated, though a bit more likeable, in her Baluchi guise.

I am relating this fieldwork anecdote here, because it was in fact the trigger for my present undertaking: to check up on the old and largely unquestioned premise of matched-guise studies as first established in Lambert et al. (1960), cited above, which holds that informants are to be kept ignorant of the fact that they are hearing the same speaker(s) over again using different accents, varieties, or languages. In the following, I start out by further elaborating on this premise and exploring its possible motivations. I then report an experiment that abandons this protocol, and instead openly presents the same speakers in different guises to informants. This experiment is embedded in research on 'language

[2] In this paper, I am using the term 'language attitude' in the sense I specify in my other chapter in this volume, as referring to the *social meanings* associated with language use and variation, rather than with reference to traditional social psychological definitions of 'attitude'.

[3] The experiment also included a male bilingual speaker of Baluchi and Omani Arabic, to investigate possible gender effects. First results of the study were presented in Soukup (2010a). I am grateful to my speakers for donating their time and effort.

attitudes' in the context of Austrian German; thus, a second purpose of this paper is to provide findings in this regard. I conclude with some more general reflections on the implications of my study and its application of what I have come to call the 'open-guise' technique.

THE GAMBIT: SPEAKER (DIS)GUISES

A quick search in the LLBA database (Linguistics and Language Behavior Abstracts - ProQuest) for the term 'matched guise' in citations and abstracts underscores the fact that the matched-guise technique in its classic form[4] still has a lot of currency in 'language attitude' research today. Over the past decade (2000–2010), at least thirty-six studies (or more than three per year) were published that feature matched guises either together with filler or 'distractor' voices (11 studies), without distractors (12), using written guises (6), or incorporating elements of both matched- and verbal-guise (7). Most of those studies that did not use fillers feature a considerable amount of speakers, so that the alternation per se serves as distraction. However, in four of these studies,[5] the guises were in fact produced by only one speaker, so that no alternation occurred at all in the battery, which arguably makes speaker recognition a real issue. Yet, the authors either eschew mentioning that fact, or still explicitly state that informants were deliberately kept ignorant of speaker repetitions. Thus, Lai (2007: 231), for example, writes, 'The fact that the speaker was, for all guises, the same person was not revealed to the [informants]. On the contrary, subjects were led to believe that the speakers were different persons as they were referred to as Speakers 1, 2 and 3.' Of course, whether or not this strategy was successful can no longer be determined, for lack of evidence. But in any case, and more generally, studies like these illustrate that the application of the matched-guise technique is still tenaciously assumed to pivot on the gambit of speaker *dis*guise. The question that poses itself, then, is what motivates this gambit in the first place, and whether it is at all necessary.

Going back to Lambert et al. (1960), as its originators, perhaps the most likely explanation for the procedural decision to keep their informants 'uninformed',

[4] as contrasted with, notably, its derivative form, the 'verbal-guise', which uses different speakers for the different varieties tested, mainly for reasons of authenticity or language competence (see e.g. Garrett 2010 for description and discussion).

[5] namely, Cavallaro and Ng (2009); Jie and Zhong (2008); Kitanaka (2007); Lai (2007)

as it were, about identical speakers is an underlying assumption that if participants knew that the speakers were the same, they would be influenced by this to the point of not producing any rating differences. If anything, however, this is refuted by my Omani anecdote reported above – in my case, the judge knew the speaker very intimately (they lived in the same house), and fully recognized her, but *still* rated her differently between the two guises. It should be noted also that Lambert et al.'s assurance that their informants remained unaware of hearing the same speakers twice is not actually supported by any direct evidence (e.g. having asked the informants whether this was true). Certainly, however, my experience in Oman shows that mere ratings differences *cannot* be taken as proof that a speaker was *not* recognized as the same across recordings. All in all, then, from whichever end one looks at the matter, there does not seem to be any a priori causal link between speaker disguise and rating differentiation.

Alternatively, one could argue that what was being tested by Lambert et al. (1960) were informants' 'language attitudes' regarding supposedly *monolingual* English and French speakers, not bilinguals – or at least regarding speakers with limited command of the respective other language. Indeed, the authors state that their interest lay in eliciting assessments of members of informants' own and of the other 'language group' (English and French speakers respectively, as represented also in their informant sample; note, however, that the speakers' language group status was apparently not made explicit as a point of reference in the protocol). By force of this argument, though, it seems that the decision of whether or not the speakers are made known to be multi-lectal is a function of the specific research question asked, rather than of the methodology per se. Thus, there does again not follow any intrinsic necessity for the disguising ploy in the application of the matched-guise technique.

But of course, the very notion of using one particular way of speaking to represent one particular social group is debatable in and of itself. It is prone to essentialize a monolithic (stereotypic) link between a certain type of language use and a specific social group category; and, concomitantly, to suppress questions of agency in code choice relative to social situations. In this line, the matched-guise study paradigm has often been criticized for largely ignoring the phenomenon of linguistic variation *within* social groups or even individuals – the latter counter to the famous axiom established by variationist research which holds that 'there are no single-style speakers' (Labov 1972: 208). Similarly, Agheyisi and Fishman (1970), in their now classic review of methodologies in 'language attitude' research, reprimanded matched-guise studies for typically presupposing

'that each population is characterized or identifiable by a single language variety. However, when we examine bilingual speech communities and networks, [...] a lot of switching is found to go on [...]. So questions of speech repertoire [...] become very important and must be reckoned with rather than ruled out' (Agheyisi and Fishman 1970: 146).

Yet studies reviewed in Giles and Bourhis (1976) demonstrate that the matched-guise technique can be 'fixed' so as to address the issue of intra-speaker variation, notably by incorporating it as an assessment factor in the test design. Thus, the French Canadian informants in Bourhis, Giles and Lambert (1975) were asked to rate a speaker who could deliberately be heard to shift her accent between European and Canadian French across two interview speech events. Giles, Taylor and Bourhis (1973) had English Canadian informants assess a speaker who was known to be bilingual in French and English, based on his language selection in an interactional context. In both studies, differences in the speaker's variety selection strategy generated differences in the ratings outcome.

Studies like these suggest very generally that matched-guise experimentation is a more flexible tool than it may be given credit for, and one which can indeed accommodate research questions regarding the assessment fall-out of a particular speaker's code-switching/ style-shifting behavior. Furthermore, they drive home what seems like an almost gratuitous point, namely that informants do not seem to have any inherent problem with making sense of one speaker's use of multiple varieties.[6] Typically, however, such studies have still proceeded by playing the various instantiations of a speaker's acts of style-shifting/ code-switching (usually in a context of speech accommodation) to *different* informant groups, and having each group provide one general assessment. In other words, the disguising gambit is actually kept up, in the sense that presenting the *same* participants with different forms of the *same* individual speaker's language shifting behavior, in an outright and direct fashion, is still avoided and a one-dimensional linking of speaker identity and delivery upheld.

Yet, such one-dimensionality runs counter to an ever increasing amount of sociolinguistic research attesting that individuals do routinely, agentively, and above all very openly vary their linguistic behavior, even within the same interaction and with the same audience, *precisely* for the purpose of projecting *multi-*

[6] This is not surprising, in light of the above-mentioned sociolinguistic axiom about the inexistence of single-style speakers. Linguistic variation is a fundamental fact of life, and thus featured in everyone's (including informants'!) experience.

ple local identities and relationships (see e.g. Auer 2007; Coupland 2007; Schilling-Estes 2004). What's more, such identity projections seemingly draw on the social meanings associated with particular varieties – in other words, the respective 'language attitudes' (see Soukup 2009, 2010a, in application of i.a. Gumperz 1982). By using a particular linguistic style, then, speakers may strategically 'contextualize' (Gumperz 1982) their utterances in terms of its social associations, making these relevant to utterance interpretation by the listener, and thus effecting identity and relationship projections in a 'dialogic' process (see also Bakthin 1986 [1952–1953]). For example, Baluchi/ Arabic bilinguals have been found to switch from Arabic into Baluchi to express contempt, which is achieved by drawing on stereotypes that associate Baluchi with sounding less intelligent (see Al Zidjaly 2008; Soukup 2010a).

Despite this, studies describing the phenomenon of evoking and changing local identities via the strategic use of linguistic variation in interaction (or, the phenomenon of 'Speaker Design', as Schilling-Estes 2002 puts it) do not (yet) routinely adduce, let alone generate, empirical social psychological evidence regarding the nature and activation of the social meanings of styles in listeners, to support claims about respective interactional outcomes. But as I would argue, such evidence can in fact be quite conveniently collected via a speaker assessment experiment in which one and the same speaker can be *openly* heard in different 'guises' and is evaluated accordingly by listener-informants (who thus know that they are rating the same person in different versions). Such a procedure can recreate and simulate the process of conversational 'contextualization' operant in strategic language shifting: informants are asked to actively assess and interpret the use of different linguistic varieties in juxtaposition in the experiment, similar to when speakers use different varieties in the same conversation for utterance contextualization. In both cases, listeners are called upon to activate culturally shared social meanings attaching to the particular language varieties they hear being used, for the purposes of interpreting 'what is going on' in the activity they are engaging in (interactional inferencing / experimental responding).

To recap my argument so far, then, there appears to be no inherent necessity that drives the disguising ploy in matched-guise research. Thus, abandoning the ploy and instead applying what I have come to call an 'open-guise' method can be fully expected to 'work', in the sense of generating rating differences (see my discussion above in the context of the Baluchi-Arabic study). But what's more, it can also boast some considerable benefits and address research questions in

ways that the traditional matched-guise technique cannot. In addition to rendering any kind of artful 'smokes and mirrors' strategy obsolete, easing the workload for both investigators and informants alike, the open-guise method can actually fill an apparent empirical gap in present-day social constructionist research on Speaker Design. One could even flip the argument and go so far as to say that if there were *no* rating differences brought out in an open-guise protocol, the very claim that linguistic shifting has interactional bearings on persona and relationship projections would be rather difficult to uphold. In other words, if listeners are *not* found to call up different social meanings in connection with hearing different linguistic varieties, no rhetorical effects can be achieved via shifting between these varieties in interaction either.

In return, such application of speaker assessment studies can give attitudinal elicitation a new sense of purpose, consequence, and direction that may take it beyond the contextual and motivational 'vacuum' social psychological experimentation has been accused of in the past (Tajfel 1981: 23; see also Soukup 2010b for discussion of bearings on experimental design; see Soukup [this volume] for discussion of speaker assessment experiments under a social constructionist perspective).

In the light of these considerations, what follows below is an attempt to provide more than mere anecdotal evidence in support of an open-guise approach. I present a corresponding experiment that was carried out under the agenda of an interactional discourse analysis of Speaker Design in Austrian German, or, more specifically, speakers' strategic shifting from standard into Bavarian-Austrian dialect in TV political discussions (see Soukup 2009, 2012). Such shifting can be found to serve the expression of antagonistic interactional 'footings' (Goffman 1981) and of (negative) identities for opponents ('other-positioning' – van Langenhove and Harré 1999) in the given setting. In order to provide a sound empirical basis for my claims about these contextualization processes, then, the experiment reported here elicited the social meanings an Austrian audience is likely to call up when hearing a speaker use dialect vs. standard.

My discussion of this experiment is intended to answer two main questions. First, it tests quantitatively whether an open-guise design can in fact elicit significantly differentiated responses from informants, or whether, contrary to my expectations derived from my experience in Oman and the findings from my discourse analysis of the Austrian TV data, no assessment differences emerge if the speakers are known to be identical across guises. Secondly, if successful, the experiment should yield the common social meanings Austrian natives associate

with the use of Bavarian-Austrian dialect and standard Austrian German respectively in juxtaposition, thus outlining some basic aspects of the current 'language attitudinal' landscape in the country. While due to my research protocol my results hold most convincingly in the context of bidialectalism, the latter is, however, assumed to apply to all Austrians at least passively (see also Moosmüller 1991), allowing, arguably, for broader generalization of the outcome.

PROCEDURE OF THE OPEN-GUISE EXPERIMENT [7]

The open-guise experiment reported here consists of a speaker assessment study carried out in the spring of 2012, in which Austrian university students listened to a set of six speech samples and rated each sample via a list of twenty-two five-point bipolar semantic differential scales (Osgood, Suci and Tannenbaum 1957) provided in a questionnaire. 123 students at the University of Vienna were polled; the total (convenience) sample is 76% female (n=94) and 24% male (n=29).[8] The informants' age range was 18–30 (median: 21; mean: 21.15). All students had grown up in Austria and lived there at the time, and had at least one Austrian parent. 32% (n=39) hailed from the province of Lower Austria, 29% (n=35) from Vienna, 15% (n=18) from Upper Austria, 8% (n=10) from Salzburg, 7% (n=8) from Styria, 5% (n=6) from Carinthia, 2% (n=3) from Burgenland, and 1% each (n=2) from the Tyrol and Vorarlberg (percentages rounded). Thus, all nine Austrian provinces were represented in the sample, though a vast majority of informants came from the Middle Bavarian-Austrian dialect area in the Austrian north, which is the most populous area (comprising Upper Austria, Lower Austria, and Vienna), and from which I also recruited my speakers for

[7] This study is a spin-off from the verbal-guise experiment reported in Soukup (2009). Thus, see there for further details on experimental design, including text selection, the linguistic variation involved, as well as study setting and assessment scheme. That experiment (and the present one in analogy) was in fact designed so as to inform an interactional sociolinguistic analysis of TV discussion show data; its configuration was therefore tailored to the situational context of this show.

[8] The unequal gender-distribution in the sample is due to informant recruitment in female-student-dominated courses (though the sample still covers a broad array of subjects of study). My past speaker assessment experiments (reported in Soukup 2001, 2009) have shown, however, that informant gender has typically merely a low effect on ratings, and, if any, a predictable one, namely that females tend to give 'kinder' ratings across the board.

I cordially thank all my informants for their participation, and Manfred Glauninger at the University of Vienna for facilitating recruitment.

the experiment (see below). 32% (n=40) of the informants indicated their 'mother tongue' as *österreichischer Dialekt*, 28% (n=34) as *österreichische Hochsprache* (Austrian standard), and 40% (n=49) as both *Dialekt* and *Hochsprache*; but note again that all native speakers of Austrian German can be assumed to have some competence in both varieties.[9]

The six speech samples used in the experiment consisted of recordings by one male and two female bidialectal speakers from Upper Austria (subsequently called speaker 'M', the male, and speakers 'K' and 'S', the females).[10] All three were between thirty and forty years of age at the time of recording, with a middle class background and at least a few years of university education. Each performed the same text (a one-minute argumentative piece on genetically engineered food) once in standard Austrian German, and once in Upper Austrian dialect, which is part of Middle Bavarian-Austrian German (see above).[11]

The list of adjectives used to assess the speech samples contained the following items, compiled mainly on the basis of interviews and existing literature regarding Austrian 'language attitudes', so as to test the most commonly cited social associations of standard and dialect (English translations in italics):

sympathisch	-	unsympathisch	*likeable*	-	*not likeable*
gebildet	-	ungebildet	*educated*	-	*uneducated*
vertrauenswürdig	-	nicht vertrauenswürdig	*trustworthy*	-	*not trustworthy*
höflich	-	unhöflich	*polite*	-	*impolite*
intelligent	-	unintelligent	*intelligent*	-	*unintelligent*
freundlich	-	unfreundlich	*friendly*	-	*unfriendly*
ehrlich	-	unehrlich	*honest*	-	*dishonest*
selbstbewusst	-	nicht selbstbewusst	*self-confident*	-	*not self-confident*
kompetent	-	nicht kompetent	*competent*	-	*not competent*
fleißig	-	faul	*industrious*	-	*lazy*
natürlich	-	gekünstelt	*natural*	-	*artificial*
viel Sinn für Humor	-	kein Sinn für Humor	*good sense of humor*	-	*no sense of humor*

[9] For reference on standard and dialectal Austrian German and the linguistic situation in Austria in general, see e.g. Dressler and Wodak (1982); Ebner (2008); Hornung and Roitinger 2000 [1950]; Moosmüller (1991); Soukup (2009); Wiesinger (2006). See furthermore the website of the Austrian Academy of Sciences for a dialect map of Austria (http://www.oeaw.ac.at/dinamlex/Dialektgebiete.html – accessed June 30, 2013).

[10] A second, matching male speaker was unfortunately not available at the time of polling. As including another two samples would furthermore have added considerably to the task length, and past verbal guise research on Austrian standard and dialect had mainly shown strong correlations between male and female speakers (Soukup 2009), it was decided to carry out the study with the present speaker set. Once more, I cordially thank my speakers in this experiment for their invaluable help.

[11] My use of the term 'dialect' in the subsequent analysis of results from the experiment therefore always means a Middle Bavarian-Austrian variety.

schlau	-	nicht schlau	*clever*	-	*not clever*
emotional	-	unemotional	*emotional*	-	*unemotional*
locker	-	nicht locker	*relaxed*	-	*not relaxed*
ernst	-	unernst	*serious*	-	*non-serious*
aggressiv	-	nicht aggressiv	*aggressive*	-	*not aggressive*
streng	-	nicht streng	*strict*	-	*not strict*
konservativ	-	aufgeschlossen	*conservative*	-	*open-minded*
grob	-	sanftmütig	*rough*	-	*gentle*
arrogant	-	unarrogant	*arrogant*	-	*non-arrogant*
derb	-	vornehm	*coarse*	-	*refined*

At the beginning of the experiment, I specifically told the informants that they were about to hear the same speakers in two recordings each, presenting the same text in two different versions (‘auf zwei verschiedene Arten’). Their task was to provide feedback regarding how the speakers would come across to a public audience in each of the two ways of presenting the text (‘Wie kommen [die Personen] mit ihrer jeweiligen Art, diesen Text vorzutragen, bei einem öffentlichen Publikum an?’). The experiment was applied to two different groups of informants (n=74 and n=49); the order of speakers was switched up between those sessions so as to control for potential ordering effects (the first order of speakers being M-K-S, the second S-K-M; for each speaker, the standard version was always played before the dialectal one). Subsequent to collection, data were compiled and analyzed using SPSS for Windows (v.17.0).

RESULTS OF THE OPEN-GUISE

A series of paired sample *t* tests were carried out to compare the average ratings of each speaker between her or his respective standard and dialect guise, as elicited via the semantic differential scales.[12] The statistical results are presented in detail in Table 1 (female speaker ‘K’), Table 2 (female speaker ‘S’), and Table 3 (male speaker ‘M’) on the next pages.

[12] Parametric tests were chosen under the considerations of a sufficiently large sample and of the repeated-measures design, where homogeneity of variances can be assumed. See also Himmelfarb (1993) for discussion of using parametric tests with attitudinal scales. See furthermore e.g. Aron, Aron and Coups (2009: ch. 8) for discussion of the complexities of carrying out a large number of *t* tests and how this may increase the likelihood of Type I errors. Here, my approach is to focus my subsequent presentation of results mainly on those cases where significant mean differences occur together with at least medium effect sizes, so that the basis for my claims seems fairly solid.

Table 1: Means, standard deviations, and results from the paired-samples *t* tests for the ratings of female speaker 'K', including Cohen's *d* as measure of effect size.[13]

	Speaker 'K' standard guise		**Speaker 'K' dialect guise**				
	Mean	***StD***	***Mean***	***StD***	***N***	***t***	***Cohen's d***
educated	**4.09**	0.747	3.27	0.811	123	10.371*	**1.1**
intelligent	**4.03**	0.757	3.50	0.803	123	6.490*	**0.7**
serious	**3.86**	0.823	3.24	0.924	123	6.124*	**0.7**
industrious	**3.92**	0.822	3.48	0.754	121	5.037*	**0.6**
competent	**4.02**	0.936	3.44	0.919	122	4.907*	**0.6**
strict	**3.17**	1.143	2.54	1.096	123	4.513*	**0.6**
arrogant	**2.98**	1.112	2.43	1.079	123	3.942*	**0.5**
coarse	2.23	0.777	**3.28**	0.728	123	-10.869*	**-1.4**
relaxed	2.65	1.012	**3.69**	1.021	122	-7.955*	**-1.0**
natural	3.27	1.208	**4.13**	1.109	123	-5.877*	**-0.7**
sense of humor	2.46	0.880	**3.08**	0.946	123	-5.767*	**-0.7**
honest	3.70	0.946	**4.12**	0.868	121	-3.600*	**-0.5**
clever	**3.70**	0.802	3.41	0.769	122	3.251*	0.4
polite	**3.92**	0.988	3.67	0.945	123	2.187*	0.3
emotional	3.15	1.010	**3.48**	0.938	122	-2.705*	-0.3
friendly	3.67	1.032	**3.93**	1.035	120	-2.279*	-0.3
likeable	3.32	1.111	3.55	1.161	123	-1.719	
self-confident	4.20	0.881	4.08	0.862	121	1.281	
trustworthy	3.77	0.930	3.63	0.962	123	1.251	
conservative	3.12	1.045	2.97	0.975	123	1.187	
aggressive	2.34	1.070	2.23	1.007	123	1.129	
rough	2.64	0.919	2.66	0.924	122	-0.219	

* indicates statistically significant difference of means at p<.05, two-tailed
bold print indicates higher mean (in case of significant difference)

[13] where a Cohen's *d* of 0.2 traditionally designates a small, 0.5 a medium, and 0.8 a large effect size (see e.g. Coolican 2009). Cohen's *d* values were computed using Becker's online effect size calculator (http://www.uccs.edu/~faculty/lbecker/ – last accessed June 21st, 2013).

Table 2: Means, standard deviations, and results from the paired-samples *t* tests for the ratings of female speaker 'S', including Cohen's *d* as measure of effect size.

	Speaker 'S' standard guise		**Speaker 'S' dialect guise**				
	Mean	***StD***	***Mean***	***StD***	***N***	***t***	***Cohen's d***
arrogant	**3.46**	1.042	2.27	1.049	123	9.198*	**1.1**
strict	**3.47**	1.270	2.67	1.198	123	5.474*	**0.7**
educated	**3.42**	0.932	2.95	0.886	123	5.054*	**0.5**
conservative	**3.43**	0.979	2.89	1.100	122	4.436*	**0.5**
natural	2.37	1.231	**4.26**	1.055	123	-13.653*	**-1.7**
relaxed	2.10	1.036	**3.59**	1.207	123	-11.305*	**-1.3**
likeable	2.49	1.133	**3.61**	1.053	123	-9.690*	**-1.0**
sense of humor	2.22	0.966	**3.13**	1.036	122	-7.487*	**-0.9**
honest	3.43	0.917	**4.14**	0.826	122	-7.514*	**-0.8**
coarse	2.77	0.916	**3.42**	0.791	122	-7.247*	**-0.8**
friendly	3.02	1.169	**3.79**	1.008	121	-6.538*	**-0.7**
trustworthy	2.98	0.927	**3.46**	0.986	123	-4.628*	**-0.5**
emotional	3.48	1.059	**3.95**	0.808	123	-4.298*	**-0.5**
aggressive	**3.06**	1.237	2.57	1.153	123	3.789*	0.4
serious	**3.60**	1.010	3.19	0.956	122	3.636*	0.4
rough	**3.21**	0.917	2.88	0.826	123	3.195*	0.4
industrious	3.73	0.904	**3.55**	0.752	121	2.037*	0.2
polite	3.12	1.025	**3.42**	0.995	122	-2.858*	-0.3
self-confident	4.07	1.038	**4.27**	0.747	123	-2.442*	-0.2
clever	3.22	0.949	3.35	0.732	119	-1.534	
intelligent	3.35	0.890	3.30	0.726	122	0.587	
competent	3.28	0.979	3.34	0.895	123	-0.572	

* indicates statistically significant difference of means at p<.05, two-tailed
bold print indicates highest mean (if significantly different)

Table 3: Means, standard deviations, and results from the paired-samples *t* tests for the ratings of male speaker 'M', including Cohen's *d* as measure of effect size.

	Speaker 'M' standard guise		**Speaker 'M' dialect guise**				
	Mean	***StD***	***Mean***	***StD***	***N***	***t***	***Cohen's d***
educated	**3.78**	0.795	2.90	0.882	123	9.053*	**1.1**
intelligent	**3.77**	0.780	3.21	0.805	122	6.396*	**0.7**
polite	**4.11**	0.770	3.56	0.968	123	5.450*	**0.6**
arrogant	**2.70**	1.113	2.11	0.964	122	4.815*	**0.6**
serious	**3.56**	0.891	3.10	0.913	122	3.902*	**0.5**
natural	2.87	1.248	**4.43**	0.879	123	-12.492*	**-1.5**
coarse	2.29	0.827	**3.38**	0.835	123	-10.550*	**-1.3**
relaxed	2.78	1.132	**4.11**	0.943	122	-10.527*	**-1.3**
emotional	2.47	1.051	**3.68**	0.961	123	-10.808*	**-1.2**
sense of humor	2.60	0.985	**3.40**	0.897	122	-8.330*	**-0.9**
honest	3.78	0.958	**4.25**	0.742	122	-4.511*	**-0.6**
industrious	**3.73**	0.860	3.37	0.792	123	4.016*	0.4
competent	**3.70**	0.975	3.28	0.988	123	3.611*	0.4
clever	**3.51**	0.884	3.21	0.763	122	3.254*	0.4
aggressive	1.84	0.953	**2.17**	1.099	123	-2.796*	-0.3
self-confident	4.04	0.909	**4.26**	0.699	123	-2.297*	-0.3
rough	2.38	0.894	**2.63**	0.955	122	-2.269*	-0.3
likeable	3.45	1.013	3.69	1.061	122	-1.819	
friendly	3.86	0.925	3.98	0.908	121	-1.032	
strict	2.51	1.169	2.37	1.042	123	1.025	
conservative	3.06	1.070	2.93	1.100	122	1.000	
trustworthy	3.70	0.946	3.64	0.930	121	0.519	

* indicates statistically significant difference of means at p<.05, two-tailed
bold print indicates higher mean (in case of significant difference)

As it turns out, then, for each of the three speakers, the informants did indeed produce diverging ratings between the two guises on the majority of scale items. In fact, all three speakers were rated as sounding significantly (at p< .5) more *educated* and *arrogant* when speaking in the standard. By contrast, they were indicated to sound more *natural*, *relaxed*, *honest*, and as having more *sense of humor*, but also as sounding noticeably *coarser* when speaking in the dialect.

The effect sizes for all these items were at least medium for each individual speaker (Cohen's $d \geq 0.5$); the evidence therefore seems quite robust.

In the case of the items *serious* ('ernst'), *industrious*, and *emotional*, effect sizes differ so that for one or two speakers the effect is small, but there is at least one of the others for whom it is large. It seems reasonable to take these items also into account, so that there is an additional tendency by which the use of dialect makes a speaker sound more *emotional*, but the standard more *serious* and *industrious*.

Two salient patterns of rating 'inconsistencies' occur in the data.[14] First, both female speakers show a large effect for sounding much *stricter* in the standard, while the male speaker shows no significant rating difference here. At the same time, neither of the three speakers' ratings differ for the dialect. This might point towards a possible interaction of language use with gender, whereby using the standard has more 'negative' social consequences for females than for males with regards to perceptions of sternness. Such a hypothesis would, however, have to be subjected to much further testing. Secondly, there are recurring instances where female speaker 'S' looks like 'the odd one out' in terms of the ratings she received. Thus, speaker 'S' was said to sound significantly less *likeable* and *trustworthy* but more *conservative* in the standard, while these items did not come out significant in any way for 'K' or 'M', nor for all three speakers in their dialect guises. Furthermore, where 'K' and 'M' were assessed as significantly more *polite* in the standard, 'S', in contrast, was so in the dialect. Finally, both speakers 'K' and 'M' were indicated to sound significantly more *intelligent*, *competent*, and *clever* when using the standard (with a Cohen's $d \geq 0.4$; in fact, speaker 'K' outscores the others in both guises in this regard); however, the ratings are inconclusive here for speaker 'S'.

Very generally this seems to suggest that there was a (negative) bias in the assessment of speaker 'S' for these items with regards to her standard guise. This is supported by some open comments by informants holding that she sounded much pleasanter ('angenehmer') in the dialect overall. In the same line, an analysis of the correlation patterns regarding the three standard guises shows that the coefficient is much higher for speakers 'K' and 'M' ($r(20) = .92$, $p<$

[14] Items concerned were subjected to post-hoc analysis to compare mean ratings across the three speakers, using repeated-measures ANOVAs and paired-samples t tests in hierarchical order of means. Only those results that were found to show statistical significance are reported here. Details of the statistical analysis are not provided for space considerations – contact the author for the relevant details (barbara.soukup @univie.ac.at).

.001) than for speakers ‘K’ and ‘S’ (r (20) = .56, p< .01), while for speakers ‘M’ and ‘S’ there is in fact *no* significant correlation pattern. Meanwhile, correlation is consistently very strong for the dialect guises (speakers ‘K’ and ‘M’: r (20) = .96, p< .001; speakers ‘K’ and ‘S’: r (20) = .93, p< .001; speakers ‘M’ and ‘S’: r (20) = .96, p< .001). All in all, the standard performance by speaker ‘S’ thus seems to have featured some confounding factors that led to a different ratings outcome than for the ‘majority’ of the other two speakers in some respects. Arguably, the general, stereotypical social meanings associated with standard use are therefore better represented in the results for speakers ‘K’ and ‘M’. With this in mind, there could be additional trends whereby standard is indeed held to sound more *intelligent*, *competent, clever* and perhaps even *polite* than the dialect by Austrian listeners.

DISCUSSION AND CONCLUSION(S)

To sum up, then, the results as presented above now provide a resounding confirmation of the fact that the open-guise technique actually ‘works’: my informants had no problem at all in making sense of the fact that they were hearing the same speakers twice, using different linguistic varieties. They adjusted their assessments accordingly, and, crucially, still generated ratings that differentiated between the dialect and standard guises for many items. Further, the outcome now represents the basic patterns of the social meanings associated with standard and dialectal Austrian German in juxtaposition, or, in other words, a basic outline of the identities between which a speaker can move and which s/he can index via style-shifting. Middle Bavarian-Austrian dialect (the dominant dialectal variety in Austria) is thus associated with sounding less *educated*, *serious*, *industrious*, and *refined* than standard, but also with being perceived as less *arrogant* as well as more *natural*, *honest, emotional, relaxed*, and ‘fun’. This, particularly (but arguably not only) in contexts of style-shifting: at least for those items just listed, for which the three speakers’ ratings were consistent and showed considerable effect sizes, the elicited social associations can be argued to be fairly robust and generalizeable.

The findings thus seem to conclusively answer the two main questions explored in this paper, in the sense that an open-guise experiment can indeed generate contrastive ratings (at least in contexts such as the Austrian, where linguistically differentiated stereotypes exist), and that, as a reliable tool for checking

on evaluational activity, it provides a solid empirical basis for analyses of interactional contextualization. Further, the outcome now sketches the basic patterning of the 'language attitudinal' landscape in Austria.

In other respects, however, it is especially the patterns of rating divergences across speakers which now bring a question to point that any speaker assessment experiment has to face: to what extent are the rating differences a function of the language variety used, or rather of a particular speaker's performance? Where the ratings of the two females coincide but differ from those of the male speaker, a gender effect can arguably be assumed. Where, on the other hand, the assessment of speaker 'S''s standard guise diverges so saliently from that of her peers, it may be necessary to look for reasons beyond common stereotypes associated with the use of a particular language variety (whether by a male or female), to more individual factors of delivery and performance such as pitch or tone of voice, to explain the ratings. A comparative open guise study of even more speakers would be required to provide the necessary evidence here; this, however, must at the present remain a suggestion for further research.

On the one hand, then, Lambert et al.'s (1960) argument that matching guises can 'minimize the effects of both voice of the speaker and his message' on the ratings (Lambert et al. 1960:44) puts the results from my open-guise on quite firm ground in terms of eliciting common stereotypes regarding the social meaning of language varieties in juxtaposition (and certainly on firmer ground than any verbal guise study in this regard). However, as notably the differences *across* the three speakers' ratings have shown, this does not take the individual entirely out of the equation. Parameters like speakers' tone of voice, speech rate, and prosody must still be factored in as potential influences on ratings (see e.g. Brown, Strong, and Rencher 1975), as should perhaps others we are not yet aware of. The comparison of findings across similar studies, as well as the inclusion of calculations of effect sizes in the statistical battery, may eventually go some way towards helping us assess how robust and reliable (how stereotypical!) the 'language attitudes' we find in our experiments really are, even beyond variation in speakers.

While an open-guise approach may thus not resolve the 'variety vs. speaker effect' issue entirely in and of itself, what it does thoroughly attest to, in any case, is that one and the same speaker can indeed put on different 'coats' of identity, openly and unabashedly, by taking on different linguistic varieties. And listeners can make sense of this without problem - being fully aware of the process - by calling up respectively contrasting social meanings, whether these are

ad hoc, individual, or more stable. Ultimately, abandoning any pretense of speaker disguise in our 'language attitude' methodology may unclutter our experimental protocols and reconfigure our research questions in ways that allow us to more fully explore the true multi-dimensionality of the linkages between a speaker's social (interactional) persona and linguistic delivery.

REFERENCES

Agheyisi, R., and J. A. Fishman. 1970. Language attitude studies: A brief survey of methodological approaches. *Anthropological Linguistics* 12(5): 137–157.

Al Zidjaly, N. 2008. A mediated discourse analysis approach to language attitude research in Oman. Paper presented at Sociolinguistics Symposium 17, Amsterdam.

Aron, A., E. N. Aron and E. J. Coups. 2009. *Statistics for Psychology*. 5th ed. Upper Saddle River, NJ: Pearson.

Auer, P. (ed.) 2007. *Style and Social Identities: Alternative Approaches to Linguistic Heterogeneity*. Berlin: Walter de Gruyter.

Bakhtin, M. 1986 [1952–53]. The problem of speech genres. In C. Emerson and M. Holquist (eds.) *Speech Genres and Other Late Essays*, transl. by V. W. McGee. Austin: The University of Texas Press. 60–102.

Bourhis, R. Y., H. Giles and W. E. Lambert. 1975. Social consequences of accommodating one's speech style: A cross-national investigation. *International Journal of the Sociology of Language* 6: 55–71.

Brown, B. L., W. J. Strong and A. C. Rencher. 1975. Acoustic determinants of perceptions of personality from speech. *International Journal of the Sociology of Language* 6: 11–32.

Cavallaro, F. and B. C. Ng. 2009. Between status and solidarity in Singapore. *World Englishes* 28(2): 143–159.

Coolican, H. 2009. *Research Methods and Statistics in Psychology*. 5th ed. London: Hodder Education.

Coupland, N. 2007. *Style: Language Variation and Identity*. New York: Cambridge University Press.

Dressler, W. U. and R. Wodak. 1982. Sociophonological methods in the study of sociolinguistic variation in Viennese German. *Language in Society* 2: 339–370.

Ebner, J. 2008. *Duden – Österreichisches Deutsch*. Mannheim: Dudenverlag.

Garrett, P. 2010. *Attitudes to Language*. Cambridge, UK: Cambridge University Press.

Giles, H. and R. Y. Bourhis. 1976. Methodological issues in dialect perception: Some social psychological perspectives. *Anthropological Linguistics* 18: 94–304.

Giles, H., D. M. Taylor and R. Bourhis. 1973. Towards a theory of Interpersonal accommodation through language: Some Canadian data. *Language in Society* 2(2): 177–192.

Goffman, E. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Gumperz, J. J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press.

Himmelfarb, S. 1993. The measurement of attitudes. In A. H. Eagly and S. Chaiken (eds.) *The Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich. 23–87.

Hornung, M. and F. Roitinger. 2000 [1950]. *Die österreichischen Mundarten: Eine Einführung*. Newly revised and edited by Gerhard Zeilinger. Vienna: öbv & hpt.

Jie, X. and Y. Zhong. 2008. Locating users of interpretation in the court: An impact analysis of literal and meaningful renditions in a mock court situation. *Babel* 54(4): 327–342.

Kitanaka, K. 2007. Vocabulary recognition, English loan words and English learners. PhD Dissertation, California State University, Long Beach, CA.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lai, M. L. 2007. Exploring language stereotypes in post-colonial Hong Kong through the matched-guise test. *Journal of Asian Pacific Communication* 17(2): 225–244.

Lambert, W. E., R. Hodgson, R. C. Gardner and S. Fillenbaum. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60(1): 44–51.

Moosmüller, S. 1991. *Hochsprache und Dialekt in Österreich*. Vienna: Böhlau.

Osgood, C. E., G. J. Suci and P. H. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.

Schilling-Estes, N. 2002. Investigating stylistic variation. In J. K. Chambers, P. Trudgill and N. Schilling-Estes (eds.) *The Handbook of Language Variation and Change*. Malden: Blackwell. 375–401.

Schilling-Estes, N. 2004. Constructing ethnicity in interaction. *Journal of Sociolinguistics* 8(2): 163–195.

Soukup, B. 2001. 'Y'all come back now, y'hear!?' Language attitudes in the United States towards Southern American English. *VIEWS (VIenna English Working paperS)* 10(2): 56–68.

Soukup, B. 2009. *Dialect Use as Interaction Strategy*. Vienna: Braumüller.

Soukup, B. 2010a. 'Speaker Design': Language variation as a contextualization device. Paper presented at Masaryk University Brno, Czech Republic, December 3rd.

Soukup, B. 2010b. Matched-guise experimentation as a speech event: A recast and its implications. Paper presented at the Sociolinguistics Symposium 18, University of Southampton, September 1–4.

Soukup, B. 2012. Speaker design in Austrian TV political discussions. In J. M. Hernández-Campoy and J. A. Cutillas-Espinosa (eds.) *Style-shifting in Public: New Perspectives on Stylistic Variation*. Amsterdam: John Benjamins. 81–99.

Tajfel, H. 1981. *Human Groups and Social Categories*. Cambridge, UK: Cambridge University Press.

van Langenhove, L. and R. Harré. 1999. Introducing Positioning Theory. In R. Harré and L. van Langenhove (eds.) *Positioning Theory*. Oxford: Blackwell. 14–31.

Wiesinger, P. 2006. *Das österreichische Deutsch in Gegenwart und Geschichte*. Vienna: LIT Verlag.