Chapter 4

Non-categorical constraints in perception

4.1 Introduction

The previous chapter investigated the hard (or categorical) constraints on AAVE copula absence. AAVE copula absence is also subject to soft (or non-categorical) constraints. As noted, copulaless sentences like (1) almost always have counterparts with full (2) or contracted (3) forms of the copula.

(1) She my piano teacher.

(2) She is my piano teacher.

(3) She's my piano teacher.

The distribution of these three variants is systematically constrained by grammatical factors. These systematic yet violable constraints are well studied in the variationist tradition of sociolinguistics. However, with some important exceptions (see Chapter 5), they are not generally countenanced or even considered by competence models. Of course, it is not obvious that such constraints should be within the purview of competence grammar. The patterns found in production studies could well be the result of systematic performance factors. Therefore, this chapter presents an experiment designed to show that speakers have knowledge of a particular non-categorical constraint on copula absence. The issue of whether this knowledge is to be distinguished from knowledge of grammar is taken up in Chapter 6.

The structure of this chapter is as follows. Section 4.2 briefly reviews the notion of a sociolinguistic variable, its relationship to other linguistic systems, and the body of work on AAVE copula absence as a sociolinguistic variable. Section 4.3 presents the hypothesis to be tested in the experiment. Section 4.4 describes the methodology employed. Section 4.5 presents the results.

4.2 Non-categorical constraints in production

This section provides some background on AAVE copula absence as a sociolinguistic variable. The first question is why the alternation between overt and zero forms of the copula should be seen as a linguistic unit of any sort. To answer this, Section 4.2.1 reviews some definitions of linguistic variables. Section 4.2.2 applies those definitions to AAVE copula absence. Sections 4.2.3 and 4.2.4 discuss the properties that make the alternation a sociolinguistic variable: its correlations with social and grammatical factors.

4.2.1 Linguistic variables

This section reviews the notion of a linguistic variable with an eye towards whether variables constitute units that should be represented in a competence grammar. Many phonological variables that have been studied involve alternations between allophones of the same phoneme.[1] Thus, to the extent that the phoneme constitutes a unit of grammar, the variable does as well.

However, variationist studies have not been limited to alternations between allophones. In the words of Hudson (1996:169) "[a] variable is a collection of alternatives which have something in common." This is a notion that is applicable at pretty much

[1] Variables such as (r)—the alternation between [ɹ] and [ϕ]—are possibly better conceptualized as the realization or non realization of a phoneme, depending on whether the [ϕ] variant also involves some effect on the preceding vowel.

any level of linguistic structure. Wolfram (1991:23) enumerates six typical ways in which variants are seen to have something in common:

- Structural category realization: The variants are the realization and the non-realization of e.g., a morpheme or phoneme.
- Allo-forms within a structural category: The variants are allophones or allomorphs.
- Processes affecting linguistic classes of units: The variants are the application or non-application of linguistic 'processes', such as word-final consonant-cluster reduction.
- Co-occurrence relationships: The variants are the 'obeying' or 'disobeying' of co-occurrence restrictions, as in variable negative concord or variable subject-verb agreement.
- Item permutation: The variants are different possible orders of linguistic elements as in variable metathesis or variable word order.
- Lexical choice: The variants are (near) synonyms, usually function words.

Note, in regards to the last item, that Lavandera (1978) and Romaine (1981) argue that in variables where the variants bear 'referential' or 'ideational' meaning, one never finds exact equivalence—the variants are different ways of saying similar things.

With respect to representation in the grammar, these types of relatedness form a heterogeneous set. In the case of "process" variables, the process itself would be a thing in the grammar. In the case of lexical choice variables, one wouldn't want to say that the choice between the lexical items is a thing in the grammar. Rather, it is inherent in the grammar, i.e., inherent in the similarities between the two lexical items that allow them to substitute for one another. Thus, depending on the type of variable, the alternation may or may not correspond to something in the grammar. Each variant (e.g., lexical item, allophone, etc.), of course, is something in the grammar.

There are also methodological issues that remove variables somewhat from the domain of individual competence. As Wolfram and Hudson both point out, the identification of variables and of instances of the variants in a corpus is never completely unproblematic. For example, Wolfram's formulation of the co-occurence relationships type of variable makes it clear that, in order to conceive of such a variable, one has to establish the co-occurence restrictions on the basis of something else, perhaps a related variety without the variation.

Hudson gives the example of variable realization of /h/ word-initially (an example of structural category realization). In all varieties of English, there are words spelled with *h* which are never pronounced with *h* (e.g., *hour* in standard English). But why should the membership of this word class be assumed to be the same for all speakers? That is, if a speaker of a different variety of English is heard to say *house* as [aus], one cannot tell on the basis of that one instance whether or not *house* belongs to the *hour* class FOR THAT SPEAKER. Since it is not usually feasible to determine for every speaker studied which words have an underlying /h/, some compromise must be made. Most commonly, according to Hudson, researchers rely on their own knowledge of some related variety (usually a more standard variety).

It is this kind of compromise that distances the linguistic variable from studies of individuals' linguistic competence. Indeed, Wolfram writes:

> Is the cognitive basis of language patterning really one and the same with the sociopsychological patterning that determines the social distribution of linguistic variables? ... It may just be that the patterning of linguistic and social covariation is most adequately indicated by a construct which is not confined to a single linguistic rule or a sequence of rules but is instead a unique sociolinguistic construct. (1991:26)

Wolfram's comments are largely in response to the development of variable rules (Labov 1969), which tied sociolinguistic variables to individual rules in an SPE-style (Chomsky and Halle 1968) system of phonology. Thus a variable rule was a representation of the alternation within the grammar, and a locus for the application of non-categorical constraints (on which, see §4.2.4 below). Wolfram suggests that this conception of linguistic variables as a matter of (individual competence) grammar is a mistake.

However, before we throw the baby out with the bathwater, I would like to suggest that just because the alternation is not a matter of grammar doesn't mean that the constraints are also necessarily outside the grammar. The general picture I have in mind is this: The grammar provides choices in that it provides different ways of saying similar things. Such choices get associated with social differences or social meaning, and get exploited by speakers on the basis of that social meaning.[2] Non-categorical grammatical constraints are therefore associated directly with each variant, as factors that can affect how likely it is to be chosen.

To summarize, in most cases it is awkward to represent a linguistic variable in a grammar, but each variant is represented. Representing non-categorical constraints (discussed below) does not require a representation of the variable, as they can attach to each variant. Of course 'factors affecting how likely a variant is to be chosen' need not necessarily be a matter of grammar. The experiment reported in this chapter provides preliminary evidence that they are at least a matter of linguistic knowledge. The remainder of this section reviews the collection of variants that make up the variable of AAVE copula absence, and the constraints that affect their distribution.

4.2.2 The AAVE copula as a linguistic variable

The variable to be studied here is composed of a subset of instances of the copula in AAVE, in particular, those that appear in finite, non-past, non-habitual, non-first singular contexts. The forms that appear in those contexts are full forms (*is* and *are*), contracted forms (/z/ and /r/) and 'zero' forms, as illustrated in (1) to (2) above. In research on this variable, tokens that occur in past tense or non-finite contexts are excluded, as both contraction and the zero form are generally thought to be impossible in these environments.[3] Tokens that occur with first singular subjects are also excluded, as the zero form is again impossible (although both contracted and full forms are possible). Finally, 'habitual' or 'invariant' *be*, as in (4) from Wolfram (1969:180), is excluded on the grounds that it doesn't mean the same thing as the

[2]Such use constitutes a feedback loop wherein socially meaningful uses of variables are part of the construction of the meaning itself. See Eckert 2000, §5.1.2 of Chapter 5 and §6.2 of Chapter 6.

[3]See, however, the discussion of the zero copula and past tense in §3.3.2 of Chapter 3.

conjugated or zero forms.

(4) That's why I wonder why I don't see him—he usually be 'round.

Which of Wolfram's categories does this variable fall into? As a first pass, it appears to be an example of both 'structural category realization' and 'allo-forms within a structural category'. That is, the difference between the full and contracted forms would be a matter of allomorphy, while the zero form could be a matter of non-realization of the 'structural category' copula. However, there are still other ways of looking at it. Labov's (1969) analysis was in terms of two linguistic processes: deletion of the initial vowel of *is* or *are* (contraction) followed by deletion of the remaining consonant. The silent verb analysis in Chapter 3 suggests that the difference between the zero form and other forms of the copula is really a matter of lexical choice. The final constructional analysis of Chapter 3 makes it instead a matter of constructional choice (between an ordinary *head-comp-ph* and the *silent-verb-phrase*).

Labov's treatment of this variable as a phonological variable assumes semantic equivalence, as does a treatment in terms of straight allomorphy. However, if it is a matter of lexical or constructional choice, the possibility of semantic differences comes into play. In fact, there is some slight evidence for a semantic difference in terms of tense: Recall that, while the overt forms are marked for present tense, the zero form may in fact be unmarked for tense or underspecified between past and present tense, at least for some speakers. This is shown by sentences such as (5).[4]

(5) I dug being militant cause I was good. It was something I could do. Rap to people. Whip a righteous message on em. People knew my name. They'd listen. And I'd steady take care of business. **This was when Rap Brown and Stokeley and Bobby Seale and them ϕ on TV**. I identified with those cats...

There is also the possibility of semantic difference in the case of tokens with emphatic stress on the copula. Labov (1969) notes that this is one context with categorical copula presence (indeed, of full forms). As such, it should be excluded from studies of non-categorical constraints on variation. However, it is doubtful that

[4] From John Edgar Wideman's *Brothers and Keepers*, 1984, p. 114, cited in Green (in preparation).

all such tokens are successfully identified and eliminated in production studies of this variable. The perception experiment reported in this chapter contrasted only contracted and zero forms of the copula, largely because the full form introduces possible intonation differences as well as whatever semantic effect might follow from emphasis. The decision to contrast only zero and contracted forms in the experiment is supported by Wolfram's finding that the proportion of full forms was relatively constant across the four social groups of AAVE speakers that he studied in Detroit, ranging from 18.1% for the lower working class speakers to 27.4% for the upper middle class speakers. The white speakers from Detroit included in the study had 20.2% full forms. This contrasts with rates of zero forms ranging from 4.7% for the upper middle class speakers to 56.9% for the lower working class speakers. Wolfram concludes that "[t]he social significance of this variable is found in the different ratios of zero realization to contraction." (1969:168) In the rest of this chapter, I will refer to the zero variant as either the 'zero form' or as 'copula absence', using 'copula presence' or 'overt forms' to refer to either the contracted or full forms.

In order to consider copula presence/absence to be a variable of AAVE, it is necessary to establish that the forms with copula presence are not simply evidence of code-switching into a more standard variety. If they are, then the variation is not at the level of the form of the copula but rather at the level of choice of code. However, as discussed in Section 3.2 of Chapter 3, there is good reason to believe that both copula presence and copula absence are alternatives provided by the grammar of AAVE.

A final consideration with respect to this variable has to do with the conflation of copula and auxiliary uses of the verb *be*. That is, most studies of this variable do not exclude tokens on the basis of the syntactic category of the complement of *be*. Instances of the copula supporting nominal predicates are considered on a par with instances of the auxiliary before *ing*-form verbs and before *gon(na)*. As discussed in §4.2.4 below, the category of the complement is analyzed as a non-categorical constraint on the distribution of this variable, and it is precisely this constraint that I will be studying here. What motivates treating these different uses of *be* as one variable, when in other languages the copula is not also used as a tense auxiliary (and in fact, different kinds of non-verbal predicates can require different copulas)? The main motivation is the forms that appear: when the copula is present (i.e., not zero), it takes the forms *is* and *are*, or /z/ and /r/, across all of the different following environments. (And the formal similarities of course hold in the other parts of the paradigm of *be*—past tense, *am*.) Further, AAVE and other varieties of English are not alone in using the same form across these categories. In their company, one finds Bengali and Nahuatl,[5] and if compound tenses other than progressives are allowed, several Romance languages as well. Finally, there is syntactic evidence (discussed in §2.5 of Chapter 2) for treating copula and auxiliary uses of *be* as the same lexeme.

4.2.3 Social constraints on the AAVE copula

For Weinreich et al., a linguistic variable is not merely a choice-point in a linguistic structure:

> A linguistic variable must be defined under strict conditions if it is to be a part of a linguistic structure; otherwise, one would simply be opening the door wide to rules in which "frequently," "occasionally," or "sometimes" apply. Quantitative evidence for *covariation* between the variable in question and some other linguistic or extralinguistic element provides a necessary condition for admitting such a structural unit. (1968:169)

This and the following subsection discuss some of the quantitative evidence pertaining to the covariation of forms of the copula in AAVE with other linguistic and extralinguistic elements.

As with most sociolinguistic variables, the realization of the AAVE copula is sensitive to some social variables and not to others. The most thorough investigation of the AAVE copula across different social groups is that of Wolfram 1969. The 48 speakers in this study were balanced across four class groups (lower working class, upper working class, lower middle class, and upper middle class), three age groups (10-12 years, 14-17 years, and 30-55 years), and both genders.[6] Wolfram found that the use

[5] John McWhorter, copula project files

[6] The interviews were carried out by researchers who were not African American. As Wolfram notes, this had the result of eliciting the style that these speakers use to address respected strangers.

of the zero forms correlated inversely with social class (working class speakers use more zero forms), gender (men use more zero forms) and age (children use more zero forms than teenagers or adults, with the exception of the lower working class group in which the teenagers had the same high rate of zero forms as the children). The one social variable in his study that did not correlate with copula absence/presence was racial isolation—the amount of contact a speaker had with white speakers.

Rickford and McNair-Knox (1994) consider another kind of social variable: addressee and topic-influenced style shift. They examined two interviews with the same speaker (Foxy Boston) in East Palo Alto, one by McNair-Knox who, in addition to being African American, was also a member of Foxy's community, and one by a European American interviewer. An additional difference was that McNair-Knox's daughter (who was a teenager, like Foxy) accompanied McNair-Knox on her interview. Foxy's use of several AAVE variables was sensitive to both the addressee and the topic of conversation. Her rate of copula absence was 70% with the African-American interviewer, compared with 40% with the European-American interviewer. Within the first interview, copula absence ranged from 43% in a discussion of graduation and college/career plans to 86% when the topic moved to boy-girl conflicts and relations. Within the second interview, the topic with the least copula absence was recreation at 9% while the topic with the most was "wives and slamming partners" (a discussion of boy-girl relationships) at 82%. College and career plans came in at 17% in this interview.

In summary, rates of copula absence in AAVE have been shown to vary with socioeconomic class, gender, age, topic of conversation, and the addressee.

4.2.4 Grammatical constraints on the AAVE copula

The realization of the AAVE copula is also sensitive to a number of grammatical factors, and some of these effects have been remarkably consistent across studies in different communities. Labov's (1969) study of copula absence among young speakers in South Harlem identified effects of the type of subject (pronoun favors copula absence, full NP favors copula presence), the following grammatical environment (i.e., the part of speech of the predicate, discussed in detail below), and the preceding phonological environment (a preceding vowel favors contraction and a preceding consonant favors the zero form). These constraints differ from those discussed in Chapter 3 in that they are not absolute: copula absence is not ungrammatical but only uncommon in a disfavoring grammatical environment. Therefore I will be referring to these constraints as non-categorical (grammatical) constraints on AAVE copula absence.

The effect of a preceding NP vs. pronoun was confirmed in Detroit by Wolfram (1969), in Los Angeles by Baugh (1979) and in East Palo Alto by Rickford et al. (1991). The effect of the preceding phonological environment is less straightforward. As discussed in Section 3.3.3 of Chapter 3, Rickford et al. (1991) show that it is dependent on how one calculates the rates of each variant.

Wolfram (1969) found a further effect of person and number, comparing *is* deletion to *are* deletion.[7] (Labov (1969) only studied *is* deletion.) *Are* environments favor the zero form more strongly than *is* environments do in Wolfram's data. Bailey and Maynor (1987) confirm this pattern with data from children and 'folk speakers' (adults over 70 with less than a grade school education) in East-Central Texas. However, in Rickford et al.'s (1991) East Palo Alto data, the effect is reversed, with *is* environments favoring the zero form and *are* environments favoring overt forms. This difference could have to do with r-lessness. Wolfram's speakers have variable r-lessness, and the speakers with more r-lessness (working class) also have higher rates of *are* absence and bigger absolute differences in their rates of *are* absence and *is* absence. Bailey and Maynor (1987) and Rickford et al. (1991) do not indicate whether their speakers have r-less or r-ful dialects, but it is likely that Bailey and Maynor's speakers had r-lessness while Rickford et al.'s did not.

With so many factors, it is difficult to be certain of the exact contribution of each one by eye-balling frequency data. The form of statistical analysis called Varbrul (which runs a step-wise multiple regression) was developed to answer this kind of problem (by Cedergren (1973), Sankoff (1975), and Rousseau and Sankoff (1978)).

[7] The scare quotes are there because AAVE does allow *is* with non-third singular subjects, so that when plural subjects co-occur with the zero form, it's not clear that *are* would be the overt form. However, Wolfram reports that in his data, such non-agreement cases account for less than 5% of the overt forms.

Varbrul ascertains the effects of the various independent variables on the distribution of the dependent variable. In particular, it returns an input value (the general propensity that the speakers had for producing the distinguished variant) and factor weight for each factor in each factor group that was found to have a statistically significant effect. For example, the type of subject forms a factor group in studies of AAVE copula absence. The factors are NP subject and personal pronoun subject. A factor weight of above .5 (e.g., the factor weight of .62 that Rickford et al. (1991) found for the factor personal pronoun[8]) indicates that the factor favors the distinguished variant (here, copula absence). Similarly, a factor weight of below .5 (e.g., Rickford et al.'s (1991) factor weight of .42 for NP subject) indicates that the factor disfavors the distinguished variant. The higher/lower the factor weight, the stronger the favoring/disfavoring effect.[9]

The single most widely studied non-categorical grammatical constraint on copula absence, and the one that is the focus of the experiment, is the effect of the following grammatical environment. Table 4.1, excerpted from Rickford's Table 6.16 (1998:190), provides a summary of nine data sets from studies of copula absence in various AAVE speaking communities. The left half of the table gives information about the data set: whether it concerns *is*, *are*, or both; the population studied; and the author. The right hand side of the table gives information about the effect of following grammatical environment on copula absence. With the exception of Wolfram's study (#4), the values are Varbrul weights, with copula absence as the distinguished variant. Wolfram's results were reported are raw frequency data, given in percentages. The five environments considered are noun phrase (__NP), locative (__Loc), adjective (__Adj), *-in(g)* form verb (__V+ing), and *gon(na)* (__gon). These environments are arranged according to the ranking in (6), where environments to the left are the least favorable to copula absences and the environments to the right are the most favorable. (7) to (11) illustrate each of the environments with examples from Wolfram 1969.

(6) __NP < __Loc < __Adj < __V+ing < __gon

[8] Table 6, p. 117, Straight Deletion, *is* and *are* combined, input probability .35.

[9] For more on Varbrul, see Guy 1988.

Studies				Environments				
#	Form	Place	Source	__NP	__Loc	__Adj	__V+ing	__gon
1	*is*	NYC Thunderbirds	Labov 1969	.2	.36	.48	.66	.88
2	*is*	NYC Jets	Labov 1969	.32	.52	.36	.74	.93
3	*is*	NYC Cobras	Baugh 1979	.14	.31	.72	.59	.78
4	*is+are*	Detroit WC	Wolfram 1969	37%	44%	47%	50%	79%
5	*is*	LA	Baugh 1979	.32	.29	.56	.66	.69
6	*are*	LA	Baugh 1979	.25	.69	.35	.62	.64
7	*is+are*	Texas kids	Bailey & Maynor 1987	.12	.19	.25	.41	.89
8	*is+are*	Texas adults	Bailey & Maynor 1987	.09	.15	.14	.73	.68
9	*is+are*	EPA	Rickford et al. 1991	.29	.42	.47	.66	.77

Table 4.1: Copula absence in AAVE in different communities, by following grammatical environment. (Values given are Varbrul factor weights, with the exception of the data from Wolfram (1969), which is given in percentages.)

(7) __NP: She a nurse.

(8) __Loc: They out there in space.

(9) __Adj: She real nice.

(10) __V+ing: Do anything if you fighting.

(11) __gon: I really don't think John gonna make it.

While the pattern of non-categorical constraints in (6) is robust, it is not perfectly consistent across all of the studies. Four of the nine data sets (#s 1,4,7,9) do have exactly this pattern. Of the others, three are off by the ordering of one pair of environments: in #2, __Loc shows more copula absence than __Adj, in #3, __Adj shows more copula absence than __V+ing, and in #5, __NP shows more copula absence than __Loc. The remaining two data sets have more extensive differences. In #6, __Loc has the most copula absence, skipping ahead three slots in the ranking as compared to the most common pattern. In #8, __Loc shows more copula absence

Pair	n	Pair	n
NP < Loc	8	Loc < V+ing	8
NP < Adj	9	Loc < gon	8
NP < V+ing	9	Adj < V+ing	8
NP < gon	9	Adj < gon	9
Loc < Adj	6	V+ing < gon	8

Table 4.2: Pairwise rankings of environments

than __Adj (although they are extremely close), and __V+ing shows more copula absence than __gon.

It is important to note that these data are not exactly comparable. Some of the studies considered *is* only or *is* and *are* separately, while others treated *is* and *are* together. It is perhaps noteworthy that three of the four data sets exhibiting the ranking in (6) are of the *is+are* type, and three of the four data sets of the *is+are* type show the same pattern.[10]

Also, while the nine data sets do not all have the same ranking, they are still similar. One way to look at the similarity of these patterns is by pairwise ranking. Table 4.2 gives the number of data sets that conform to each of the pairwise rankings in the ranking given in (6). Only the ranking Loc < Adj has as few as 6 of the 9 datasets attesting it. All of the others have either 8 or 9. Looked at this way, the pattern appears quite robust.

Within each of these studies, the pattern found was argued to be robust (statistically significant)—i.e., something more than an accident of which data happened to be recorded. That (roughly) the same pattern is found across many communities is a strong confirmation. This pattern of non-categorical constraints on copula absence appears to be a property of AAVE, one that must either follow from something else in the language, follow from functional constraints on performance, or be learned, for there is no other way for it to be reproduced in community after community.

There is one important caveat to this argument: just because an effect can be

[10]Rickford et al. (1991) compare the results *is* and *are* deletion in their study and conclude that the constraints are similar enough in both cases to warrant treating them together as one variable.

found by considering the following grammatical environment doesn't mean that that's really what's going on. It could be that the correlation is between semantic properties and rates of copula absence, and that the strength of this effect together with a reasonably close pairing of syntax and semantics are enough to let the effect shine through even though we've been counting the 'wrong' thing. One semantic distinction that may be having an effect is that between individual and stage level predicates. Individual level predicates (e.g., *tall* in *Kim tall*) denote properties that do not change easily over the course of an entity's existence. Stage level predicates (e.g., *busy* in *Kim busy*) denote properties that tend to be temporary, i.e., are true of 'stages' of an individual. Another semantic distinction to investigate is the subdivision of NP predicates into identity sentences (e.g., *Samuel Clemens Mark Twain*) and predicational sentences (e.g., *Kim a doctor*). Both of these distinctions would seem to be equally likely candidates for non-categorical constraints as the syntactic distinctions studied so far. However, semantic constraints would still be grammatical constraints. Further, it seems unlikely that a reanalysis of the production data in terms of semantic categories would turn up only categorical constraints, given minimal sets such as in (1–2). Although the syntactic effect shown in Table 4.1 may only be a reflection of a semantic effect, the fact remains that it is robust. Syntactic or semantic, it must therefore either be a part of the grammar of AAVE or follow from something else in the grammar or in performance.

4.3 Hypothesis

This leaves us with the question of what (if anything) in the grammar of AAVE underlies the synchronic pattern of copula absence. The various existing theories of non-categorical constraints on variation will be discussed at length in Chapter 5. In brief, there are three possibilities in the literature: Variable rule accounts (e.g., Labov 1969) posit optional context-sensitive rules as part of the grammar. These optional rules have a probability of application that is affected by the grammatical context. Optimality Theoretic (OT) accounts (e.g., Anttila 1997) posit partially unranked grammars. The distribution of a variant (or candidate, in OT terms) is determined by the number of complete rankings that select that variant. The third possibility is functional accounts (e.g., Kiparsky 1972) where the disfavoring effect of certain grammatical environments is attributed to functional pressures on performance. The variable rules model, but not the others, requires that speakers have direct (if tacit) knowledge of non-categorical constraints on variation.

If speakers have knowledge of non-categorical constraints on variation, it is almost certainly tacit and inaccessible to introspection. Fortunately, Labov's (1963) finding that sociolinguistic variation is socially meaningful provides a jumping off point for constructing an experiment. If sociolinguistic variation is socially meaningful, then the social value of variants might interact with the non-categorical constraints. In particular, Wolfram makes the following observation concerning the data in Table 4.3 (from Wolfram 1969:172):[11]

> The relatively high frequency with which zero realization is found preceding intentional future *gonna* among middle-class informants suggests that zero realization preceding *gonna* is less stigmatized than zero realization in other environments. (1969:172–173)

	Following grammatical environment				
Speaker group	__Adj	__NP	__Loc	__V+ing	__gon
Middle class	1.6%	4.2%	13.3%	11.3%	33.3%
Working class	36.5%	47.3%	44.4%	50.0%	78.9%

Table 4.3: Wolfram's copula absence by following grammatical environment and social class

Generalizing to allow for social values other than stigmatization, I propose the following two-part hypothesis:

I Copula absence/presence in AAVE is associated with some social value.

[11]See also Sylva and Zwicky's (1975) observation that the stylistically marked syntactic rules they discuss can be more or less stylistically marked depending on the lexical items they interact with. For example, they observe that existential *there* is relatively 'formal', but only with verbs other than *be*.

II Copula absence/presence in AAVE is more strongly associated with that social value the more marked the environment is for each variant.

For example, if copula absence sounds confident, then copula absence before a noun should sound especially confident and copula absence before a verb somewhat less so.

Note that Part II of this hypothesis entails that speakers have knowledge of non-categorical constraints. However, what is at issue here is more than just the knowledge of the constraint. Previous formal approaches to constraints on variation, if they consider social constraints at all, treat them as separate from grammatical constraints. The hypothesis being tested here is that they interact. That is, social constraints are conceptualized as the social meaning of the variable, and grammatical constraints as the intensifying or attenuating effect of the grammatical environment on the social meaning or social value of the variable.

4.4 Methodology

The experimental design was based on the matched-guise methodology of Lambert et al. 1975.[12] Lambert et al. were looking at language attitudes in a French-American community, and in particular the social evaluation of speakers based on language variety. The varieties represented in the study were European French, middle class Canadian French, lower class Canadian French, middle class Madawaskan (local) French, lower class Madawaskan French, middle class Madawaskan English and middle class non-regional English.

They presented to listeners from various groups within the community recordings of the same passage spoken in the different language varieties. In some cases, bilingual or bidialectal talkers[13] rendered multiple versions of the passages, i.e., appeared in

[12]See also Lambert 1967, Giles 1971, Anisfeld and Lambert 1964, Lambert et al. 1966, S. Lambert 1973, and Lambert and Tucker 1975.

[13]I will use the term 'talker' for the speakers who made the recordings in a matched-guise experiment because the term 'speaker' is used to refer to a person with competence in a given language, regardless of what s/he is using the competence to do in the situation being considered. The participants who listened to the recorded stimuli will be referred to as 'listeners' in general, but also as 'speakers' when it is important to highlight their competence of a particular variety.

multiple guises in the study.[14] The talkers each read the same 20-second passage from *Le Petit Prince* by Saint-Exupéry, translated or adapted into the different varieties. Lambert et al. made their experimental tape by repeating each speech sample end-to-end, and randomly ordering the resulting 40-second speech samples.

They then presented this recording to listeners, and asked them to rate talkers on 20 7-point scales, to rate (again on 7-point scales) how much they felt they resembled the talkers and how much they would like to be like the talkers, and to guess at the occupation of the talker. Table 4.4 lists the scales used by Lambert et al. in their studies with high-school and college-age participants. These scales were chosen on the basis of preliminary studies with the target populations aimed at discovering which personality qualities were valued.

1.	intelligent – not-intelligent	11.	unsociable – sociable
2.	active – passive	12.	short – tall
3.	unfair – fair	13.	ambitious – not ambitious
4.	truthful – untruthful	14.	friendly – unfriendly
5.	good-looking – ugly	15.	not religious – religious
6.	not comical – comical	16.	strong – weak
7.	not courageous – courageous	17.	impolite – polite
8.	unsure – confident	18.	happy – sad
9.	likeable – hateful	19.	selfish – not selfish
10.	reliable – unreliable	20.	determined – not determined

Table 4.4: Rating scales used by Lambert et al.

Lambert et al. found that their listeners have definite social evaluations of the different language varieties. For example, the high school students in their study middle class Madawaskan French as favorably as European French and the English varieties on the traits that Lambert et al. related to social attractiveness (fair, likeable), but not as favorably on scales related to competence (intelligent, determined).

For the experiment reported in this chapter, I extended the matched-guise methodology to test the social evaluation of an individual linguistic feature. This entailed

[14]In some of the earlier studies (Lambert 1967, Giles 1971) the methodology was purely matched-guise, in the sense that each talker recorded the passage in each variety being considered.

certain modifications: a drastic reduction in the length of the each speech sample, and a corresponding reduction in the number of scales used. These changes are described in further detail and motivated in the following subsection.

4.4.1 Stimuli

Test sentences

As the purpose of this study was to test the social evaluation of a particular linguistic feature (copula absence) and the effect of the grammatical context on that evaluation, the speech samples had to be much shorter than those used in Lambert et al.'s study. This is because the longer the speech sample, the more likely the introduction of confounding variables. Similarly, the sentences were designed to avoid AAVE-and SAE-specific features as much as possible. Even if held constant across all of these stimuli, a feature other than copula presence/absence that is strongly stereotyped could have flooded the effect of the variable under consideration. Nonetheless, since I was looking at the evaluation of copula absence within the system of AAVE, it was important to choose sentences that are a part of AAVE, and to make sure that their actual production conformed to AAVE phonology, etc. Further, in testing for the effect of the non-categorical constraint in question, it was important to control for all other known non-categorical constraints.

The non-categorical constraint to be tested is the effect of the following grammatical environment. Recall that production studies have established that the ordering of following grammatical environments, from least favoring of copula absence to most favoring, is as in (12).

(12) __NP < __Loc < __Adj < __V+ing < __gon

Ideally, one would want to test all of the environments to see if they are ordered in perception as they are in production. However, in order to keep the experimental task reasonably short, only one pair was tested. *Gon* is somewhat problematic because it is itself an AAVE-specific feature and the high rates of copula absence before *gon* suggest that it may be in the process of being reanalyzed as a modal.[15] Leaving out *gon*, the V+ing and NP environments allow for the greatest contrast.

The test sentences, designed to meet all of these considerations, are as follows:[16]

(13) a. Yeah I know her. She's teachin me piano at Music World.
b. Yeah I know her. She's my piano teacher at Music World.
c. Yeah I know her. She teachin me piano at Music World.
d. Yeah I know her. She my piano teacher at Music World.

Sentences (13a) and (13b) represent the copula presence condition, while (13c) and (13d) represent the copula absence condition. (13a) and (13c) represent the V+ing condition, and (13b) and (13d) represent the NP condition.

The sentences are about as close as they can be in phonological and semantic content given the requirement of setting up the different conditions, and are constant with respect to all other known constraints: the subject is always a pronoun, and always phonologically identical. There is no known effect of following stop vs. nasal.

The initial and final phrases (*Yeah I know her* and *at Music World*) are included to give the sentences some length.

In addition to the test sentences, the following sentences were also recorded as filler stimuli:

(14) a. Yeah I know her. She useta teach me piano at Music World.
b. Yeah I know her. She useta be my piano teacher at Music World.
c. Yeah I know her. She taught me piano at Music World.
d. Yeah I know her. She was my piano teacher at Music World.

The filler sentences match the test sentences fairly closely in their semantics, but not as closely as the test sentences match each other. Likewise, there is more variation in phonological content.

The talkers that I recruited to record the stimuli confirmed that all of these sentences would sound natural in the middle of an AAVE conversation.

[15] Although not one with all of the properties of a true auxiliary: for example, it doesn't invert. (Salikoko Mufwene, p.c., Jan., 2000). Perhaps it is more like *better* in SAE *You better not do that.*

[16] Ideally, such an experiment would involve multiple similar test sets. However, in order to keep the listeners' task to a manageable length, only the test set in (13) was used.

Recording the stimuli

Eight talkers were recruited from the Stanford community to record the sentences. All were (at least) bidialectal, commanding both AAVE and SAE.[17] All were female. While it would certainly be interesting to look into the effect of gender of the talker on the evaluation of copula absence, gender was excluded as a variable in this study. The primary reason for this is that including both male and female talkers in the stimulus set would have made it easier for listeners to recognize individual talkers when they heard them for a second, third or fourth time. This could have led to listeners trying to be consistent in the way they rated different talkers.

The sentences were presented to the talkers in written form. This immediately raised the issue of how to avoid reading style and get something that more closely approximated natural AAVE speech. The issue was further exacerbated by the presence of the white researcher, the use of recording equipment, and the surroundings: the sound-proof room in the Stanford phonetics laboratory.

The solution adopted was to have the talkers do the recording session together in groups as much as possible. In the initial session, three talkers were brought together. They first read the sentences out loud to each other and discussed different ways of saying them. I tried to background myself as much as possible during this interaction. Then the talkers took turns recording all eight sentences (i.e., both the test and filler sentences) sometimes just saying them all in a row, sometimes with a prompt from another one of the talkers ("You know her?"). Subsequent talkers came in pairs or alone, and were all played the existing tape to give them an idea of how the initial group had decided the sentences should sound. This method succeeded in producing relatively natural sounding recordings despite the conditions and in minimizing the variation across talkers in features such as intonation. It is important to emphasize at this point that both sentences with copula absence and sentences with copula presence are possible AAVE sentences, and were produced with similar intonation and other phonological features. That is, it is NOT the case that the copula presence sentences represent standard English while the copula absence sentences represent AAVE.

[17] I did not systematically collect geographical data on the talkers. Although all were currently living in California, some at least were from other regions of the country.

All of the talkers were recorded saying each sentence multiple times. Each instance was digitized using SoundEdit on a Macintosh and the actual stimuli were selected on the basis of clarity, lack of hesitation or extraneous sounds like laughter, and consistency of intonation across the different sentences.[18] The four talkers with the best recordings were then designated as test-talkers and the other four as filler-talkers. The test-sentence recordings from the test-talkers and the filler-sentence recordings from the filler-talkers were combined to make 32 stimuli.

Finally, in order to avoid any effect of extraneous variation, a recording of *Yeah I know her* was selected for each talker from one of the recordings not chosen as stimuli and pasted onto the beginning of each of that talker's test or filler stimuli followed by a standardized pause of 0.15 seconds. The *Yeah I know her* was chosen to prosodically match the the rest of the stimuli as closely as possible and so usually came from the same reading of the sentences as the stimuli (e.g., from a filler sentence for a test-talker and vice versa). This added a certain unnaturalness to the stimuli, but it was equally unnatural for each stimulus. This was possible because the *Yeah I know her* part was prosodically independent of the rest of the sentence in all of the recordings chosen as stimuli. To do the same thing with the end of the sentences (*at Music World*) would have introduced too much unnaturalness.

The resulting test stimuli ranged from 2.482 seconds to 2.698 seconds in length, with an average of 2.592 seconds. These lengths include 0.1 seconds of silence at the beginning and end of each stimulus. The filler stimuli ranged from 2.36 to 3.208 seconds, with an average of 2.711 seconds. It was decided not to follow Lambert et al.'s strategy of lengthening the stimuli by repeating them on the grounds that this would give listeners more of a chance to consciously identify the variation in the realization of the copula.

Test scales

Lambert et al. were researching the language attitudes of members of the community they studied, and indirectly the ethnic identification of people growing up in a bicultural environment. Accordingly, they used the scales to assess the stigma or prestige

[18] All judged impressionistically by the experimenter.

associated with each language variety represented in their recordings. In their analysis, they group the scales into subgroups associated with competence (intelligence, determination, etc.) and social attractiveness (fair, likeable, etc.).

Since I was focusing on one feature (copula absence) there was the possibility that it has some more specific social value, akin to the way *ing*/*in* is associated with formality/informality (Fischer 1958), and that one of the scales chosen would be close enough to this social value to reflect it. This argues for including as many scales as possible. On the other hand, too many scales would end up making the experiment too long and could also lead to the last several scales being presented too long after the listener heard the stimulus. In the end, I settled on the seven scales listed in Table 4.5.

1.	comical – not comical
2.	confident – not confident
3.	well educated – not well educated
4.	good job – not a good job
5.	likeable – not likeable
6.	polite – impolite
7.	reliable – unreliable

Table 4.5: Scales used for the experiment

Each scale was presented with a question for context: "How likeable/reliable/well educated does this person sound?" "How polite/comical is this person being?" "How confident do you think this person is feeling?" and "How good of a job do you think this person has?"[19] All of the scales had seven points from which the listeners could choose.

These particular scales were chosen so as to cover as much ground as possible in the hopes that at least one of them would be relevant to copula absence. Further, they were meant to be traits that are not necessarily correlated. In particular, someone could sound confident without sounding well educated. Also, the scale 'reliable –

[19]The form of this last question is not grammatical for all speakers. Whether it was grammatical for all of the speakers in this study and, if not, whether it would have had any effect on the results, is unclear.

unreliable' was included in the hopes that it would index in-groupness, that the listeners would interpret it as reliable in the sense of 'a friend you can count on'.

Language attitudes questionnaire

Lambert et al. used the ratings as a measure of their participants' attitudes towards the language varieties they were considering. In this study, I was investigating the social evaluation of a particular feature and the effect of the surrounding grammatical environment on that feature. Overall attitudes towards AAVE certainly have an effect on the evaluation of AAVE features. Combining ratings from speakers with a positive stance towards AAVE and from speakers with a negative stance could potentially mask interesting effects. Therefore, a language attitudes questionnaire was included in the study.

The questionnaire consisted of five statements. Listeners were asked to specify whether they agreed or disagreed with each statement, and to what extent. The five statements are given in Table 4.6.[20] The word AAVE was explained in the instructions for this questionnaire.

S1	There are some situations in which it is better to speak AAVE than Standard English.
S2	A Black person who doesn't speak AAVE isn't really Black.
S3	AAVE is bad English and should never be used in any situation.
S4	It would be weird for me to speak Standard English instead of AAVE with my Black friends.
S5	When I hear a stranger speaking AAVE, I assume they are less educated.

Table 4.6: Statements for language attitudes questionnaire

[20]These statements were designed for an intended participant pool of African Americans, and were somewhat awkward for the non-African-American participants.

4.4.2 Participants

Participants were recruited from a introductory psychology course at a community college in a community in California with a large African-American population. Participants earned extra credit in their course for their participation and were also paid \$5.

Since participants could earn extra credit, the experiment had to be open to anyone in the class. The people who chose to participate fell into the five ethnically and linguistically defined groups given in Table 4.7. These groups will form the basis of the analysis below, and will be referred to by the numbers assigned to them in Table 4.7.

I	AAVE speakers (all African American)	N = 11
II	African Americans who do not identify as AAVE speakers, but who are familiar with AAVE	N = 5
III	participants who are familiar with AAVE but are not African American	N = 6
IV	native speakers of English with no familiarity with AAVE	N = 6
V	non-native speakers of English with no familiarity with AAVE	N = 7

Table 4.7: Groups of listeners

Familiarity with AAVE was determined in an exit interview as described in §4.4.3 below. Whether or not a person spoke (any variety of) English natively was determined on the basis of the demographic questionnaire. Any listener who grew up in a non-English-speaking country was counted as non-native. Note that membership in Groups II and III only requires familiarity with AAVE and not native-speaker status in English. In fact, one listener in Group III grew up in the Philippines, and a speaker who grew up in Haiti was included in Group II.[21] Groups II and III are distinguished on the basis that the African Americans' experience with AAVE is substantially different from the others'. Group III listeners reported that they were familiar with AAVE because they went to predominantly African American high schools and/or participate in Hip Hop culture.

[21] This speaker learned both Standard English and AAVE in this country. He was included in Group II on the basis that while he didn't have native command of AAVE, he did have extensive experience with it. His responses also pattern with those of the other speakers in Group II.

The demographic questionnaire consisted of the questions listed in Table 4.8. The ethnicity question was added only in the second session, as I had originally expected only to recruit African American participants. Participants' first answers were accepted, so that if father's occupation was reported as "retired" or "deceased" no further information was collected unless volunteered by the participant.

Are you (male/female)?
What is your current occupation?
What was your previous occupation (if any)?
What is your father's occupation?
What is your mother's occupation?
Where did you live between the ages of 2 and 18?
When did you move to the Bay Area?
How old are you?
Ethnicity

Table 4.8: Demographic questionnaire

The demographic information is summarized in the next few tables. Table 4.9 summarizes the information on age and gender by group. Table 4.10 gives the ethnic self-identification of the participants as well as information on where they grew up.[22]

Group	age range	median age	# men	# women
I	19–55	26	3	8
II	24–33	26	2	3
III	18–27	20.5	1	5
IV	19–40	25	2	4
V	18–34	21	2	5

Table 4.9: Age and gender information, by group

[22]Some of the African American participants gave their ethnicity as 'Black' and some as 'African American'. However, the software used for this portion of the experiment allowed participants to see the responses of the previous participant, and they often seemed to choose based on what was already there. That is, if the previous participant had said African American, another African American participant would say the same. If the previous participant was, say, Mexican American, an African American participant might say either 'Black' or 'African American'.

Group	ethnic self-identification	home state/country[a]	N
I	(didn't ask)	California	4
	African American	California	3
	Black	California	2
	African American	Florida	1
	Black	Louisiana, California	1
II	African American	California	2
	African American	Alabama	1
	Mixed	California	1
	Black	Haiti	1
III	White	California	1
	Jewish	Texas	1
	Filipino	Philippines, California	1
	White	California	1
	Caucasian	Ohio, New Mexico, California	1
	Asian	Illinois, California	1
IV	(didn't ask)	California	2
	Indian	US Virgin Islands, British Virgin Islands, California	1
	Latina	California	1
	White	California	1
	Mexican American	California	1
V	Chinese	China, California	3
	Asian Pacific	Korea	1
	Asian	China	1
	Vietnamese	Vietnam, California	1
	African American	Nigeria	1

[a]Where the listener lived between the ages of 2 and 18

Table 4.10: Ethnicity and home state/country, by group

Tables 4.11 and 4.12 give the responses to the questions about occupation, which were included to gather information about the socio-economic status of the participants. Rather than summarize this information based on a scale such as that used by Labov (1966), I have given the actual answers from each participant. The main reason for this is that these data were not collected in a consistent manner: in many cases participants gave job titles from which it is difficult to deduce occupations, or simply employers ("Pac Bell") or industries ("fast food"). There are further reasons that these responses are probably not reliable indicators of socio-economic status: Some of the participants' parents probably still live in other countries (notably, participant N in Group II and some of the participants in Group V), where the status of different occupations may not be the same as in the US. A second consideration is the age range of the participants. For the younger participants, occupations such as food service are more likely to be temporary part-time jobs. Finally, there are many blank cells in this table. In many cases, the occupation of a parent was unknown, or simply given as "retired" or "deceased". Blank cells in the second column (previous occupation) indicate that the current occupation is the participant's first.

L	Current Occupation	Previous Occupation	Father's Occupation	Mother's Occupation
Group I				
A	Bank teller	Student	Clerk	Nurse
B	Receptionist	Clerk		Security
C	Resident Management Officer	Program Monitor	Welder	Licensed Vocational nurse
D	Hair stylist			Pac Bell
E	Mortgage Consultant	Student	Coach	Psychologist
F	File clerk	Material handler		Computer programmer
G	Dispatcher	Certified nurse's assistant		Nurse
H	Executive asst.	Principle Secretary	Deputy sheriff	Job developer
I	Medical asst.			
J	Student	Accounting	Police man	Housewife
K	House keeping	Dishwasher		House keeping
Group II				
L	Auditor	Teller		
M	Industrial sales	Industrial management	Industrial work	Social work
N	Security	Student	Farmer	Housewife
O	Meeting planner	Administrative assistant	(pilot?)	Administrative assistant
P	State employee			County employee

Table 4.11: Current, former, mother's and father's occupation, Groups I-II

L	Current Occupation	Previous Occupation	Father's Occupation	Mother's Occupation
Group III				
Q	Student	Pharmacy intern	Janitor	
R	Student		Manager	School teacher
S	Student	Waitress	Lawyer	Housewife
T	Student	Phone operator	City manager	Paralegal
U	Student		Engineer	Housewife
V	Medical coordinator	Cashier	Teacher	Housewife
Group IV				
W	Human resources coordinator	Insurance representative		Instructional asst. (elementary school)
X	Sales		Doctor	Doctor's asst.
Y	Student		Engineer	Cosmetology
Z	Student		Teacher	Accountant
a	Student		House painter	USPS
b	Stay at home mom	Appointment setter	Chemical engineer	Works for social secretary
Group V				
c	Nurse		Doctor	Nurse
d	Student		Food service worker	Housewife
e	Student		Construction worker	Homemaker
f	Student	Fast food	Engineer (car design)	Home support service
g	Food service Student	Environmental services (water sampling)	Food service	Clothes manufacturing
h	Sales associate	Waitress	Cook	Cook
i	Student		Manager	works at a jewelry store

Table 4.12: Current, former, mother's and father's occupation, Groups III-V

4.4.3 Running of the experiment

Lambert et al. give the following description of how they introduced the task:

> [L]isteners were led to believe that they were being tested on their ability to infer the personality characteristics of a speaker using speech cues only; the language spoken or the dialect variation heard was said to be insignificant and to be overlooked; they were told that the task was analogous to estimating another's personality from a phone conversation or from overhearing but not seeing a communicator. (1975:135)

This introduction of the task requires participants to believe that there is some way that the experimenters were able to objectively determine the personality characteristics of the talkers so as to have something to compare the listeners' responses to. In my estimation, it is quite likely that the community college students I was recruiting as participants as well as the college students in Lambert et al.'s study would be suspicious of this.

Instead, I introduced the study by saying that I was researching how consistently people judge other people on the basis of their voice. I went on to say that I was particularly interested in how this worked in the AAVE-speaking community. I described AAVE as the language variety used by some African Americans in some particularly informal situations, and emphasized that not all African Americans speak AAVE, nor do AAVE speakers always use AAVE, most also speaking at least one other variety. I particularly encouraged anyone who spoke AAVE to participate, but welcomed anyone in the class as non-AAVE speakers who would form my control group(s). This introduction was given when I went to the psychology class, and given the patterns of attendance, it is not certain that all of the people who eventually participated were there to hear it.

The experiment took place in an instructor's office at the community college, with one to two participants at a time (there were two computers). The session began with the participants reading and signing a consent form. For any participants who missed the in-class introduction, the consent form provided the first information on the objective of the study:

> You are invited to participate in a research study on how consistently people judge other people on the basis of their voices. First you will be asked some questions about yourself (such as your age and occupation). Then you will be asked to listen to 32 short recordings of people speaking and rate each speaker on 7 scales representing personality characteristics, such as reliable/unreliable or polite/impolite. When you are done, you will be asked to complete a survey about different speech styles.

Once the demographic questionnaire was completed, the participants donned headphones attached to the computers they would use. The stimuli were presented to the subjects by the program PsyScope (Cohen et al. 1993), beginning with the instructions shown in Figure 4.1.

In this experiment, you will hear 32 sentence spoken by different people. After each sentence, you will be presented with 7 questions on the screen. Each question will have a scale to use in your answer. For instance, if the question is "How kind does this person sound?", the scale might look like this:

Kind • • • • • • • Unkind

If you think the person sounded very kind, click on the point closest to kind. If you think they sounded neither particularly kind nor unkind, click on the point in the middle, etc.

As soon as you click on a point, the scale will go away and you will be presented with the next scale. When you have finished all 7 scales, you will hear the next voice.

If you don't want to answer one of the questions, you can wait for it to time out.

The first trial will be a practice trial. When you are ready to start the practice trial, click the mouse button.

Figure 4.1: Instructions for main experimental task

The participants completed one practice trial and then were given a chance to ask questions before beginning the actual trials.

Each trial began with the message "loading speaker" being displayed for 0.5 seconds to give the participants a signal that they should get ready to listen. Then the screen went blank and one of the 32 test or filler stimuli was played over the headphones. The stimuli were presented in a near-random order: it was important to minimize the probability that stimuli from the same talker would be presented one after the other. The actual orderings were produced by randomly selecting one stimulus from each talker and presenting those eight stimuli in a random order before selecting another stimulus from each talker, etc. While this did lead to back-to-back stimuli from the same talker on occasion, it was less frequent than it would have been with a true random order. The order of presentation was determined separately for each listener.

Each recording was followed by the seven scales. The scales were presented in a random order, with the position of the positive and negative poles selected randomly (i.e., sometimes the negative adjective was on the left, and sometimes it was on the right). The presentation of the scales, together with the contextualizing question, was as in Figure 4.2.

How good of a job do you think this person has?

Good job	•	•	•	•	•	•	•	Not a good job

Figure 4.2: Presentation of scales

The question and scale remained on the screen until the participant put the mouse on or near one of the seven bullets and clicked,[23] or until the question timed out at 20 seconds. The timeout was included as a mechanism for allowing participants to

[23]Participants were able to use either a track pad or an external mouse.

pass on individual questions, as specified in the consent form. A click on or near the bullet closest to the negative end of the scale (e.g., 'not a good job') was coded automatically by PsyScope as 1, the next one over as 2, etc. After all seven scales were presented in this fashion, the next trial began with the "Loading speaker..." message.

The last part of the experiment to be presented on the computer was the language attitudes questionnaire. It was prefaced by another instructions screen, shown in Figure 4.3. Each of the five questions in Table 4.6 above was presented followed by the scale of responses shown in Figure 4.3. As with the rating scales, each statement and response scale stayed on the screen until either the participant clicked on one of the responses or it timed out. The time out period for these questions was originally set at 20 seconds, but extended to 50 seconds after some of the listeners who participated on the first day found 20 seconds to be too short.

Finally, when the computer-based portion was complete, there was an exit interview in which I informed the participants of the purpose of the study and determined their level of familiarity with AAVE. Whenever possible, the exit interview was done with two participants at a time. The participants were shown a list of the eight sentences (test and filler) and the exit interview proceeded roughly according to the following script:

> These are the sentences that the people were saying. How many different speakers did you think there were? ... Actually, there were eight speakers. Four said each of these sentences [the test sentences] and four said each of these sentences [the filler sentences]. These sentences [the test sentences] are the ones I'm interested in. The rest are filler sentences. As you can see, these two have the verb *is* in them and these two don't. In AAVE, both ways of saying it are grammatical. I was interested to see if the choice between saying *is* and leaving it out would affect how people rated the speakers.
>
> Also, in these two sentences the predicate is a verb, *teachin me piano*, while in these two it's a noun, *my piano teacher*. When you record people speaking AAVE, it turns out that they're much more likely to leave the *is* out when the predicate is a verb, like *teachin*, than when it is a noun, like *piano teacher*. So I am interested to see if that will have any effect on how people rate the speakers. My hypothesis was that if, for example, it

The second part of this experiment is a quick survey about your opinions about the language variety spoken by many African Americans in informal situations. Linguists call this variety African American Vernacular English, or AAVE for short. Other names include Black English, Ebonics, and Jive.

In this survey, you will be presented with 5 statements. Each statement will be accompanied by a scale that looks like this:

Disagree Strongly	Disagree Somewhat	Neutral	Agree Somewhat	Agree Strongly

Click on the point in the scale that best fits how you feel about the statement. As soon as you click on a point in the scale, the next statement will appear.

As before, if you don't want to answer one of the questions, you can wait for it to time out and a new one will appear.

If you have any questions about this task, please ask the experimenter now.

If not, please click the mouse button to continue.

Figure 4.3: Instructions for language attitudes questionnaire

> sounds confident to leave *is* out, it should sound even more confident to leave it out in the unusual case.
>
> Of course, this should only be true for listeners who are familiar with AAVE. Are you familiar with AAVE?

The phrasing of the last question was fortuitous. I chose to ask about familiarity rather than actual speaking because I didn't want anyone to feel like I was accusing them of speaking a non-standard variety and because in the first exit interview I had one African-American and one non-African-American participant. By asking the question this way, it was applicable to both of them. As it turns out, some African-American participants responded to this question by saying something like "Yeah, I'm familiar with it. We talk that way at home." Other African-American participants said, "I don't talk like that, but I hear it from … " One man said he "winced" at the preacher's bad grammar at his church. All of the African-American participants were at least familiar with AAVE. Also, some of the non-African-American participants said they were familiar with AAVE, most commonly from Hip Hop culture or from going to high schools with large African-American populations. The differences between these three groups are very interesting and will be highlighted in the discussion below.

Difficulties and missing data

Some technical difficulties and other problems led to a small amount of missing data, as described in this section.

First, a speaker setting on one computer combined with a different sampling rate for one of the test talkers caused the stimuli for that talker to not be played (i.e., the computer would show "Loading speaker..." followed by silence followed by the first scale) or to cut off in mid-sentence. When this was discovered, the program was altered so as not to try to play that talker's stimuli at all. Somewhat later, the speaker setting on the computer was fixed and the program returned to its original state. All together, seven listeners got the incomplete or silent stimuli. It was possible to tell from the output of the program which ones these were, and responses to the scales for those stimuli were discarded. Three additional listeners were only presented with stimuli from seven talkers. None of these listeners are African American, but some were familiar with AAVE nonetheless. For perhaps related reasons, the computer simply crashed part way through for one of the participants who got the incomplete or silent stimuli (Listener U). Whereas the first problem led to data being systematically missing for one test talker, the crash part way through meant that there is data missing for this listener across talkers.

In addition, some listeners didn't answer specific scales during the experiment, and the scales timed out. It's not clear whether the listeners deliberately decided they didn't want to answer these questions, or if they just hesitated too long and had the scale time out. However, the response times for the next scale in each case are long enough (at least 5 seconds) that it's clear they didn't try to click just as the scale was changing. 8 data points are missing for this reason, distributed among 7 listeners.

Further, the time out on the language attitudes questions was set at too short a setting at first. This led to a few cases where the point on the response scale that the listener chose was recorded as their response to the next question. This was evident because the response times recorded for the following question in these cases were very short, and the results were corrected accordingly.

Although the consent form stated that "you have the right to withdraw your participation at any time without penalty" and I reiterated this to the participants after they signed the forms, only one person (Listener L, an African-American participant) opted out part way through. Like the listener for whom the computer crashed, there is data missing for this listener across different talkers. She was willing to do the language attitudes questionnaire. Although I did not get a chance to do the complete debriefing with her, I did find out that she does not speak AAVE. Also, the responses she did give fit in most closely with those of Group II.

The two speakers who did not complete the main experimental task were unable to do the language attitudes questionnaire directly on the computer. Instead, I used the computer to show them the questions, but had them report their answers to me out loud. In one other case, a listener told me that he had clicked on the wrong answer for one of the language attitude questions and I recorded his corrected response.

In a couple of cases, listeners asked what the word "AAVE" meant when they saw it in the first language attitudes question. The other listeners were apparently satisfied by the definition given in the instructions for this part of the experiment.

Finally, because I had originally expected to only be recruiting AAVE speakers (and hence African-American participants), the original demographic questionnaire did not contain a question about ethnicity. This was fixed for the second day of testing, but I did not get ethnic self-identifications from the first six participants.

4.5 Results

This section presents the results of the experience. Section 4.5.1 briefly reports the results of the language attitudes questionnaire, which turned out not to correlate with responses on the judgment task. Sections 4.5.2 and 4.5.3 report the results of the judgment task and show that they confirm both parts of the hypothesis. The hypothesis is repeated here:

> I Copula absence/presence in AAVE is associated with some social value.
>
> II Copula absence/presence in AAVE is more strongly associated with that social value the more marked the environment is for each variant.

The results of the judgment task show that most listeners associated copula absence with some social value. That is, for some scales, their responses were sensitive to the presence or absence of the copula. For the two African-American groups (Groups I and II), there is a further effect of the grammatical environment, such that marked cases were rated more to the extremes of the scales. This effect is not present for Groups III-V.

4.5.1 Language attitudes questionnaire

The responses to the language attitudes questionnaire did not strongly correlate with differences between the groups or with responses on the experimental task. In fact, the most interesting language attitudes information did not come from the questionnaire, but rather from the debriefing: The self-identification as an AAVE speaker or non-AAVE speaker was a far stronger predictor of responses on the experimental task than any of the language attitudes questions. (See §4.5.3 below.) A summary of the responses to the language attitudes questionnaire are given in Table 4.14 for completeness.[24] Table 4.6 is repeated here as Table 4.13 for reference.

[24]The total number of responses for each question does not always add up to the number of listeners in a group as some listeners did not answer all of the questions.

S1	There are some situations in which it is better to speak AAVE than Standard English.
S2	A Black person who doesn't speak AAVE isn't really Black.
S3	AAVE is bad English and should never be used in any situation.
S4	It would be weird for me to speak Standard English instead of AAVE with my Black friends.
S5	When I hear a stranger speaking AAVE, I assume they are less educated.

Table 4.13: Statements for language attitudes questionnaire (reprise)

Group	statement	agree strongly	agree somewhat	neutral	disagree strongly	disagree somewhat	avg[a]
I	S1	2	3	1	2		3.62
	S2				3	8	1.27
	S3				6	5	1.55
	S4	1	1	3	2	3	2.50
	S5		3	1		6	2.10
II	S1	1	1	1	2		3.20
	S2			1		4	1.40
	S3			1	2	2	1.80
	S4			1		4	1.40
	S5	1	1			3	2.40
III	S1	3	1		1	1	3.67
	S2					6	1.00
	S3				2	4	1.33
	S4		1	2	1	2	2.33
	S5	1	4		1		3.83
IV	S1	3	1		1		4.20
	S2				2	4	1.33
	S3		1	1	1	3	2.00
	S4		1	2	1	2	2.33
	S5	1	2		1	2	2.83
V	S1		4	1		1	3.33
	S2		1	1	3	2	2.14
	S3			3	2	2	2.14
	S4		1	1	3	2	2.14
	S5	1	1	4	1		3.23

[a]Average response calculated by assigning 'agree strongly' a value of 5, 'agree somewhat' a value of 4, etc.

Table 4.14: Responses to language attitudes questionnaire

4.5.2 Copula presence vs. copula absence

This section presents the results relevant to Part I of the hypothesis, that copula absence/presence in AAVE is associated with some social value. In this analysis, the listeners were divided into the five groups shown in Table 4.7, according to their experience with AAVE and other varieties of English. Each listener heard each sentence four times and rated it each time on seven scales.[25] The sentences are repeated here in (15). In the labels for the sentences, P stands for copula presence, A for copula absence, V for following verb and N for following noun.

(15) a. PV: Yeah I know her. She's teachin me piano at Music World.
b. PN: Yeah I know her. She's my piano teacher at Music World.
c. AV: Yeah I know her. She teachin me piano at Music World.
d. AN: Yeah I know her. She my piano teacher at Music World.

To evaluate the hypothesis, I averaged the ratings across talkers, within listeners, sentences and scales. This gives each listener's average rating on each scale of each sentence. The four average ratings define an order of the sentences on each scale by each listener. Since there were 35 listeners and 7 scales, this gives a total of 245 such observations. Figure 4.4 gives six example observations, one for each possible order of two P sentences and two A sentences, where no sentences are rated exactly the same.[26]

To determine if a listener considered a scale relevant to the social value of copula absence, I compared their scores for the A sentences and P sentences. In the orders AAPP and PPAA, the sentences are differentiated according to copula absence/presence. If a listener's ratings placed the sentences in one of these two orders, the scale was considered relevant for that listener.

Out of 245 observations, AAPP appears 10 times, PPAA appears 115 times.[27] The remaining 120 observations found some other order (including orders not shown

[25] Some data is missing, as discussed in §4.4.3.

[26] In this figure and in the discussion that follows, I have regularized the scales so that the negative end is always on the left.

[27] These numbers include cases where both A sentences or both P sentences were rated the same.

AAPP: Group II, Listener N:

```
                                AN  AV      PV      PN
not     .       .       .       .       .       .       . conf
conf    1       2       3       4       5       6       7
```

APPA: Group I, Listener E:

```
                                 AN     PN PV     AV
not     .       .       .       .       .       .       . like
like    1       2       3       4       5       6       7
```

APAP: Group IV, Listener Y:

```
                        AV      PN  AN          PV
unrel   .       .       .       .       .       .       . rel
        1       2       3       4       5       6       7
```

PAAP: Group I, Listener K:

```
               PV               AV      AN PN
unrel   .       .       .       .       .       .       . rel
        1       2       3       4       5       6       7
```

PAPA: Group IV, Listener b:

```
                     PV    AV   PN  AN
not     .       .       .       .       .       .       . com
com     1       2       3       4       5       6       7
```

PPAA: Group III, Listener R:

```
           PV    PN AV                        AN
not     .       .       .       .       .       .       . com
com     1       2       3       4       5       6       7
```

Figure 4.4: Sample observations with no sentences rated exactly the same

in Figure 4.4 where one or more A sentences were rated the same as one or more P sentences).

Table 4.15 summarizes the 10 cases where the A sentences were both rated strictly higher than the P sentences (i.e., the 10 PPAA cases). In this table, the letters in parentheses are the listener codes. They show, for example, that only Group I's listener K rated both A sentences higher than both P sentences on two scales ('confident' and 'job').[28]

Scale	Group				
	I n=11	II n=5	III n=6	IV n=6	V n=7
comical	0%	0%	33% (R,T)	17% (a)	14% (h)
confident	9% (K)	0%	0%	17% (b)	14% (f)
educated	0%	0%	0%	0%	14% (i)
job	9% (K)	0%	0%	0%	0%
likeable	0%	0%	0%	0%	0%
polite	9% (J)	0%	0%	0%	0%
reliable	0%	0%	0%	0%	0%

Table 4.15: A* > P*, by group and scale (A* stands for AN or AV, P* stands for PN or PV. % are % of listeners in each group ranking the sentences such that A* > P*.)

Table 4.15 considers only those cases where the A sentences were ranked strictly higher than the P sentences. These scales were relevant to the social value of copula absence for these listeners. However, a slightly broader definition of relevance would allow one A sentence to overlap with one P sentence, as in (a) in Figure 4.5. In this ordering, the A and P sentences are not strictly distinguished, but the A sentences tend toward the higher end of the scale, and no A sentence intervenes between the P sentences or vice versa. Allowing observations that match the pattern in (a) of Figure 4.5 (but not those in (b) and (c) of the figure), adds four more listener/scale pairs to Table 4.15. This is reflected in Table 4.16. Although Table 4.16 remains sparse, it is interesting that the scale 'comical' produced the most PPAA orderings.

[28] In the caption of this and similar tables, A* stands for AN and AV and P* stands for PN and PV.

(a): Group I, Listener I:

		AN												
	PV	PN		AV										
not	.		.		.		.		.		.		.	com
com	1		2		3		4		5		6		7	

(b): Group I, Listener J:

							PN							
							AN							
						PV	AV							
not	.		.		.		.		.		.		.	com
com	1		2		3		4		5		6		7	

(c): Group I, Listener E:

								AV						
								PN	AN	PV				
impol	.		.		.		.		.		.		.	pol
	1		2		3		4		5		6		7	

Figure 4.5: Sample observations with some sentences rated exactly the same

Scale	Group				
	I n=11	II n=5	III n=6	IV n=6	V n=7
comical	9% (I)	20% (P)	33% (R,T)	17% (a)	14% (h)
confident	9% (K)	0%	0%	17% (b)	14% (f)
educated	9% (K)	0%	0%	0%	14% (i)
job	9% (K)	0%	0%	0%	0%
likeable	0%	0%	0%	0%	0%
polite	9% (I)	0%	0%	0%	0%
reliable	0%	0%	0%	0%	14% (i)

Table 4.16: A* '$\geq$' P*, by group and scale

Since copula absence was rated strictly higher than copula presence in only 10 cases (and higher in only 14 cases), I will focus in the remainder of this analysis on the cases where copula presence was rated higher than copula absence. Table 4.17 gives the percentage of listeners in each group who rated the P sentences strictly greater than the A sentences on that scale (order AAPP in Figure 4.4).

Scale	Group				
	I n=11	II n=5	III n=6	IV n=6	V n=7
comical	18%	40%	17%	17%	29%
confident	27%	60%	17%	50%	14%
educated	64%	80%	100%	67%	43%
job	45%	100%	67%	67%	29%
likeable	45%	40%	17%	67%	29%
polite	27%	100%	50%	50%	29%
reliable	55%	100%	67%	50%	14%

Table 4.17: P* > A*, by group and scale

From Table 4.17, one can conclude the following: Group II listeners were most likely to judge a scale to be relevant to the social value of copula absence, Group V listeners were the least likely to do so. The scale 'educated' was relevant for the most speakers, across groups, while the scale 'comical' was the least relevant.

Tables 4.18 to 4.22 give the breakdown by listener in each group. In these tables, an X indicates that the listener rated the P sentences strictly higher than the A sentences. An / indicates that the listener rated the P sentences higher than the A sentences, but not strictly higher. (That is, the listener produced an order that is the mirror image of (a) in Figure 4.5.)

Starting with Group I (the AAVE speakers), Table 4.18 shows that while there wasn't absolute agreement as to which scales were relevant to copula absence, the scales 'job' and 'educated' were selected by most of the listeners in this group. The table also shows wide variation in the number of scales that a listener found relevant, with listener A selecting all 7 scales and listeners J and K selecting none.[29]

	Listener											
Scale	A	B	C	D	E	F	G	H	I	J	K	Total
job	X	X	X		X	X	/	/	/			8
educated	X	X	X	X	X		X	X				7
reliable	X	X	X	X	X		X					6
likeable	X	X	X	X		X						5
polite	X		X	X								3
confident	X			X		X						3
comical	X	X										2
Total	7	5	5	5	3	3	3	2	1	0	0	34

Table 4.18: P* > A* by listener and scale, Group I

In contrast with the results of Lambert et al.'s (1975) study, there is scant evidence for a positive evaluation of copula absence among AAVE speakers. While it was to be expected that AAVE speakers should rate the talkers in the copula presence condition as sounding more educated and like they have better jobs, the same isn't necessarily true for the other scales.[30] As mentioned above, the scale 'reliable' was included in the hopes that it would index in-groupness. It patterned with 'job' and 'educated' instead. One possible explanation for this is that the participants interpreted 'reliable'

[29]Note, however, that listener K rated the A sentences above the P sentences on three scales (cf. Table 4.16 above).

[30]In fact, Listener K provides an exception to this result, selecting 'job' and 'education' with the other PPAA.

as having to do with corporate reliability—i.e., reliable in the eyes of one's employer. However, these data alone do not support any firm conclusions. One might also expect listeners to score the more vernacular sentences higher on the scales 'likeable' and 'confident', but this didn't happen. Although fewer listeners produced the AAPP order for these scales than for 'job' or 'educated', only one (Listener K) produced the other order (PPAA) for 'confident' and none did for 'likeable'.

It is possible that these results reflect 'linguistic insecurity' (Labov 1966), the stigmatization of AAVE features by AAVE speakers. However, it is also possible that they reflect the influence of the experimental environment. The experiment was carried out at the community college the participants were attending, by a white experimenter and with a computer. The listeners in Group I may have been evaluating the sentences they heard with respect to the norms of the wider community.

Table 4.19 shows that Group II (African Americans familiar with AAVE, non-AAVE speakers) is much more consistent than Group I: All five speakers selected the scales 'job', 'reliable', and 'polite' as relevant, rating the P sentences strictly higher on those scales than the A sentences. Over half of the listeners in this group selected each of the scales 'confident', 'educated' and 'likeable' as relevant.

	Listener					
Scale	L	M	N	O	P	
job	X	X	X	X	X	5
reliable	X	X	X	X	X	5
polite	X	X	X	X	X	5
confident	X	X	X		/	4
educated	X	X		X	X	4
likeable	X	X		X		3
comical	X		X			2
Total	7	6	5	5	5	

Table 4.19: P* > A* by listener and scale, Group II

The listeners in Group II all indicated that they were familiar with AAVE, but then distanced themselves from the variety. They did this by saying things like "I wince whenever my preacher splits his infinitives" or "I'm trying to teach my daughter to speak Standard English, but she's picked up AAVE at school." It seems that these speakers may have decided at some point to use Standard English rather than AAVE, although both may have been available to them. Having made such a decision, they would probably have negative associations with AAVE features, and such features would probably be very salient for them.

As shown in Table 4.20, while Group III (non-African Americans familiar with AAVE, non-AAVE speakers) selected fewer scales than Group II, there is still complete agreement on one scale ('educated'), and fairly good agreement on one other ('polite'). As with Group I, there is rather wide variation in how many scales a given listener selected.

	Listener						
Scale	Q	R	S	T	U	V	
educated	X	X	X	X	X	X	6
polite	X	X	X	/	/		5
job	X	X	X	X			4
reliable	X	X	X	X			4
likeable		X				/	2
confident	X						1
comical	X						1
Total	6	5	4	4	2	2	

Table 4.20: P* > A* by listener and scale, Group III

Table 4.21 gives the results for Group IV (native speakers of English, no familiarity with AAVE). In this group, 'job', 'educated' and 'likeable' are selected the most often, but none by all 6 listeners. There is also wide variation in how many scales a given listener selected. Listener W selected all 7, and Listener b only one, with AAPP (see Table 4.15 above).

Finally, Table 4.22 gives the results for Group V (non-native speakers of English, no familiarity with AAVE). This table is the sparsest of all, although there is some agreement on 'educated' as a relevant scale.

The general picture that emerges from these tables is as follows: Group II had the most agreement about which scales were relevant, but it's not clear that this

	Listener						
Scale	W	X	Y	Z	a	b	
job	X	X	X	X			4
educated	X	X	X	X			4
likeable	X	X		X	X		4
confident	X	X	X				3
polite	X	X	X				3
reliable	X	X		X			3
comical	X		/				2
Total	7	6	5	4	1	0	

Table 4.21: P* > A* by listener and scale, Group IV

	Listener							
Scale	c	d	e	f	g	h	i	
educated	/	X	X	/	X	/		6
likeable	X	X		/				3
job	X	X			/			3
polite	X		X					2
confident	X					/		2
comical				X			X	2
reliable			X					1
Total	5	3	3	3	2	2	1	

Table 4.22: P* > A* by listener and scale, Group V

should be interpreted as this group having a strong convention of associating a social value such as 'reliable' with copula absence. It is also possible that copula absence was simply the most salient for this group, and generally evaluated negatively. For Group III, where fewer scales were selected overall, it is more striking that the scale 'educated' was selected by all of the speakers. Group V listeners selected the fewest scales as relevant, and Groups I, III and IV fall somewhere between Groups II and V. The fact that Group I didn't select more scales as relevant, and especially the fact that they didn't rate the A sentences higher than the P sentences on all but a few scales, may have been the result of a conflict between their own attitudes towards

AAVE and the normative attitudes towards AAVE that they were sensitive to in the experimental session.

Nonetheless, the evidence presented in this section provides support for part I of the hypothesis: In the experimental task, listeners based their social evaluation of the talkers in part on the presence vs. absence of the copula. The heterogeneity in the pattern of scales selected (especially in Group I) raises intriguing further questions, but the fact that the majority of listeners selected some scales is sufficient to allow for testing of part II of the hypothesis.

4.5.3 Effect of the following grammatical environment

This section reports the results that pertain to part II of the hypothesis: that there should be an interaction between the grammatical environment and the social value of the variable. In particular, the hypothesis states that copula absence/presence in AAVE is more strongly associated with its social value the more marked the environment is for each variant. That is, part II of the hypothesis will be confirmed, if, for any scale treated as relevant by a listener, copula absence before a noun (AN, the more marked case) is rated lower than copula absence before a verb (AV, the less marked case) or copula presence before a verb (PV, marked case) is rated above copula presence before a noun (PN, marked case). If a talker sounded relatively uneducated to a listener in the copula absence condition, they should have sounded especially uneducated when using copula absence before a noun. To put it yet differently, to the extent that a disfavorable social value is associated with copula absence, that disfavorable social value should be heightened in the AN sentence with respect to the AV sentence. Conversely, to the extent that a favorable social value is associated with copula presence, that favorable social value should be heightened in the PV sentence with respect to the PN sentence. Such a result would show that speakers have knowledge of the non-categorical constraint.

Returning to the ordering of the sentences on the scales illustrated in Figures 4.4 and 4.5, and only considering the AAPP and PPAA orders, there are two orderings that conform to the hypothesis:

(16) a. AN $<$ AV $<$ PN $<$ PV
b. PV $<$ PN $<$ AV $<$ AN

In these orders, the two marked cases (PV and AN) are at the ends. Six other orders (given in (17), contradict the hypothesis:

(17) a. AN $<$ AV $<$ PV $<$ PN
b. AV $<$ AN $<$ PN $<$ PV
c. AV $<$ AN $<$ PV $<$ PN
d. PV $<$ PN $<$ AN $<$ AV
e. PN $<$ PV $<$ AV $<$ AN
f. PN $<$ PV $<$ AN $<$ AV

However, as noted above, there were only 10 listener/scale pairs in which both A cases were rated above both P cases. Therefore, we will focus on the P $>$ A type, i.e., the orders (16a) and (17a-c).

The complete orderings in (16) assume that both copula presence and copula absence are socially significant, and thus should have social values sensitive to the following grammatical environment. However, it could be that only one variant (absence or presence) in fact bears social value for a given population. That is, copula absence could be socially and stylistically unmarked, available for use at any time, while copula presence is socially and stylistically marked (or vice versa).[31] If copula absence is unmarked, then the ratings for the copula absence sentences should reflect a baseline, while the ratings for the copula presence sentences reflect the additional effect of copula presence. In this case, one would only expect to see an effect of the following grammatical environment on the evaluation of copula presence. Of course, it is not possible to tell a priori which variant (or both) is socially meaningful, and different groups may indeed differ on this point. Therefore, the effect of the following grammatical environment will be considered separately for copula presence and copula absence.

[31] This is simplifying things a bit, as what is socially and stylistically marked surely varies with the situation and the interlocutor.

If we consider copula presence and copula absence separately, the hypothesis-confirming order (16a) breaks down into two orders, given in (18).

(18) a. AV/AN < PN < PV
b. AN < AV < PV/PN

These orders specify that both P sentences must be rated higher than both A sentences, but the relative ranking of the N and V sentences is only relevant for either P or A. In this form, the disconfirming orders would be:

(19) a. AV/AN < PV < PN
b. AV < AN < PV/PN

The null hypothesis is that the following grammatical environment has no effect, so the orderings of the V and N sentences should be equally distributed between (18) and (19). The alternative hypothesis is that the following grammatical environment systematically affects the ratings, and one ordering of the V and N sentences occurs significantly more than chance. In particular, the ordering in (18) should come up significantly more than chance.

The Exact Binomial Test can distinguish between the null and alternative hypotheses in a dataset like this one. The results of this test are given in Table 4.23. The first column of this table gives the group. The second column gives the total number of observations for the group. For example, there were 11 listeners in Group I and 7 scales, giving 77 observations for Group I. The third column gives the number of observations in which the listener rated both of the P sentences higher than both of the A sentences on the scale. Since this part of the hypothesis concerns the effect of the following grammatical environment on the social value of the variable, it was only tested within those cases where the listener found the scale relevant to that social value. (For Group I, there were 31 such cases.)

The fourth column gives the proportion of observations that matched the predicted order of AV and AN. The denominator in the fourth column is less than the number in the third column, as any cases where AV averaged the same as AN were discarded.[32]

[32]That is, they were considered to be cases where the instrument was not sensitive enough to tell which way the order went.

Group	Total N	P* > A*	AV > AN	p	PV > PN	p
I	77	31	[a]12/28	0.2888	*18/23*	*0.0053*
II	35	26	*23/25*	*0.0000*	16/25	0.1148
III	42	20	[a]6/17	0.1662	9/18	0.5927
IV	42	22	[a]4/14	0.0898	8/14	0.3953
V	49	16	5/9	0.5000	[b]3/10	0.1719
I&II	112	57	*35/53*	*0.0135*	*34/48*	*0.0028*
III-V	133	58	[a]15/40	0.0769	[b]20/42	0.4388

[a]AN > AV more frequent.
[b]PN > PV more frequent.

Table 4.23: Effect of following grammatical environment, all groups

The next column gives the p values produced by comparing this proportion to 1/2 (chance distribution) with the Exact Binomial Test. The only significant effect of the following grammatical environment on copula absence was for Group II, and it goes in the direction predicted. For Groups I and III-V, the distribution of the two orders of AN and AV is not distinguishable from chance. The last two columns give the results for the effect of the following grammatical environment on copula presence. In this case, the only significant effect was for Group I, again in the predicted direction. Finally, in order to make sure that the lack of significant results for Groups III-V was not due to the small sample size, the last row of the table gives the results for the combined group III-V. Even for this larger group, no significant result emerges.

These results show that, for Groups I and II only, there was significant effect of the grammatical environment on the social evaluation of forms of the copula. When listeners in Group I rated copula presence higher (on some scale) than copula absence, they also rated copula presence before a verb higher than copula presence before a noun significantly more often than chance. This means that the social value of copula presence for these listeners is intensified in the marked environment. There are two possible ways that this could come about. The first is that these listeners know that a following verb is a marked environment for copula presence, and judge that a speaker would only use copula presence in that environment if s/he were particularly emphatic about expressing the social value of copula presence. The second is that the social value of the copula presence is encoded in the grammar as being dependent on the part of speech of the predicative phrase. In this case, the fact that copula presence is rarer before verbs than before nouns would follow from speakers avoiding copula presence before verbs except when they really wanted to express that more intense social value. Either way, the listeners have some direct, if tacit, knowledge of the non-categorical constraint. Similar remarks hold for Group II with respect to copula absence.

Why should there be a difference between Groups I and II? The result that Group I listeners only showed an interaction with the grammatical environment in copula presence would follow if copula absence is socially and stylistically unmarked for Group I, while copula presence is socially meaningful. Similarly, the results for Group II would follow if it is copula presence that is unmarked and copula absence that is marked for this group. Recall that Group I listeners self-identified as AAVE speakers, while Group II listeners said they were familiar with AAVE but did not self-identify as AAVE speakers, and indeed distanced themselves from AAVE. It would seem quite plausible for the latter group to find copula presence unmarked and copula absence as meaningful (and indeed, as indicative of various negative personality traits). Copula absence would be less remarkable for the AAVE speakers (Group I). It is possible, although less predictable, that they associate social value only with copula presence. Another possibility, raised in more detail in Chapter 6, is that the fine-grained distinctions in social value according to the grammatical environment depend on the variant in question being relatively infrequent. Copula absence is obviously infrequent for non-AAVE speakers (Group II), and copula presence is more infrequent for AAVE speakers than for non-AAVE speakers. Whether it is infrequent enough, and indeed what contexts that frequency is calculated over, remains to be seen.

Although there are important differences between the results for Groups I and II, both groups' responses are sensitive to the same non-categorical constraint. One might ask why the listeners in Group II, as non-AAVE speakers, were aware of this pattern while the listeners in Group III, who also claimed to be familiar with AAVE, were not. The answer is most likely that the African Americans (Group II) had much more extensive and involved experience with AAVE as listeners than did the listeners in Group III who were not African American. The contrast between Groups I and II, on the one hand, and Groups III-V on the other, is an important result: only speakers who are sufficiently familiar with AAVE were showed knowledge of the non-categorical constraint, although speakers in other groups did find copula absence/presence relevant to their evaluations of the talkers. This shows that the non-categorical constraint is specific to AAVE, and not something latent in the structure of English in general.

4.6 Conclusion

This experiment remains a pilot study in many respects. The results need to be confirmed for a larger population, for other non-categorical constraints and for other variables. Nonetheless, the results are suggestive, and raise the interesting possibility that speakers have knowledge of non-categorical constraints on sociolinguistic variation and that these constraints are further relevant to the social evaluation of any given instance of a variable. Chapter 5 evaluates existing theories of non-categorical constraints on sociolinguistic variation against the results of this experiment and other aspects of socially meaningful variation.