Language Variation and Change, 3 (1991), 33–74. Printed in the U.S.A.

Near-mergers and the suspension of phonemic contrast

WILLIAM LABOV, MARK KAREN, AND COREY MILLER
University of Pennsylvania

ABSTRACT

In 1972, Labov, Yaeger, and Steiner reported a series of "near-mergers" that have since proved to be difficult to assimilate to the standard conception of the phoneme and that challenged our current understanding of how language production is related to perception and learning (Labov, Yaeger, & Steiner, 1972). In these situations, speakers consistently reported that two classes of sounds were "the same," yet consistently differentiated them in production. Labov (1975a) suggested that this phenomenon was the explanation for two "falsely reported mergers" in the history of English, where word classes were said to have merged and afterward separated. It appears that sound change may bring two phonemes into such close approximation that semantic contrast between them is suspended for native speakers of the dialect, without necessarily leading to merger. This article reports on further observations of near-mergers, which confirm their implications for both synchronic and diachronic issues, and presents the results of experiments that show how phonemic contrast is suspended for an entire community.

THE CATEGORICAL VIEW AND ITS CONCEPTUAL BASIS

One of the most striking facts about the data on near-mergers is the difficulty that linguists have had in recognizing them. Over many decades, the facts that show this asymmetry of production and perception have been rejected in favor of theoretical assumptions that are inconsistent with them. To understand why this should be so, and to appreciate the significance of the facts, it may be helpful to outline those assumptions and see how they clash with the observed facts.

That linguistic categories are discretely separated into mutually exclusive nonoverlapping sets is perhaps the most fundamental concept of linguistics; without it, linguistic analysis as we practice it would not be possible. I refer to this position as *the categorical view*. It is a view that is shared by both the structuralist and generative traditions, which differ from each other in so many other notions of phonology.

The research on which this article is based has been supported by grants from the National Science Foundation for "A Quantitative Study of Sound Change in Progress," "Linguistic Change and Variation in Philadelphia," and "A Study of Cross-Dialectal Comprehension."

Though the categorical view privileges discreteness, linguists have not ignored cases that are difficult to categorize discretely. The history of phonology is in fact a long series of struggles with marginal phenomena. But in each case, the resolution of the struggle is the creation of new discrete categories, new levels of organization, or an abandonment of old restrictions, in order to preserve the notion of a discrete binary contrast between categories. Phonologists have wrestled with:

the categorization of segments that were in complementary distribution except for a few lexical contrasts. The structuralist approach was to merge this situation with all other phonemic contrasts under the slogan, "Once a phoneme, always a phoneme." The generative solution to this problem is the opposite one: to derive the contrasting forms from a single underlying form by a rule at the lexical level.

the problem of segments that were in contrast only across grammatical boundaries. Moulton (1962) considered the case of German *ich-laut* and *ach-laut*, which can be reduced to a single velar fricative only when the grammatical status of the suffix *-chen* is admitted. The solution of the structuralist position is to ignore this boundary at the phonemic level and to recognize a single unit only as a morphophoneme; the generative position is to disregard the distinction between phoneme and morphophoneme (Halle, 1962).

the status of phonetic elements that occur primarily in loanwords of various degrees of integration into the native vocabulary. The usual solution here is to recognize a distinct "co-existent system" (Fries & Pike, 1949).

the categorization of segments that are phonetically intermediate between two phonemes in environments where their contrast is neutralized, as in the much discussed case of English stops after /s/. One resolution of this issue is to create a new type of segment, the *archiphoneme*, at a different level of organization, but it is more common to select one or the other phoneme as underlying on the basis of phonetic similarity or pattern pressure.

Though these issues are stated in terms of the status of *segments*, the same problems necessarily arise when one moves to the level of features aligned in separate tiers. The question as to which elements are to be recognized as distinctive or underlying elements continues to arise whenever one set of categories shows a more limited distribution than normal or is largely but not completely predictable. To deal with such issues, the binary distinction between *phoneme* and *allophone* or *underlying* and *derived* element is preserved through the creation of separate levels of organization. Thus, the simplicity of the theory in regard to the fundamental character of categories is preserved at the cost of creating a larger number of abstract modes of representation.

The marginal phenomena of near-mergers have not been a part of the traditional debates on categorization and marginality just discussed. One reason is that the fine-grained phonetic observations or measurements that would show such approximation were not represented in the literature. That

is a result of the strength of the categorical view, which puts pressure on phoneticians to phonemicize their impressionistic transcriptions. But first, we have to explore more deeply the theoretical principles that are so deeply ingrained as to mold the behavior of linguists so forcefully.

The concept of contrast

The phonology of structural linguistics placed a great deal of emphasis on the concept of contrast, and a great deal of attention was given to deciding whether a particular phonetic difference was contrastive or not. Contrast was, of course, defined at a level of phonetic representation divorced from all grammatical information. Though higher level units were also considered, an equivalent amount of attention was not given to deciding whether two morphophonemes were in contrast. Contrast was defined in a number of different ways; an overview of the literature shows that most linguists considered contrast as a set of co-occurring properties that hold for structural linguists concerned with an autonomous phonology, for Prague School linguists concerned with more integrative systems, or for generative phonologists trying to establish a set of underlying forms. Two segments or features are in contrast when:

1. There is at least one environment where the difference between them is the only difference between two utterances that differ in meaning.
2. The distribution of the two forms is not predictable by any general rule.
3. Native speakers are sensitive to the difference between the two categories, at some level of behavior, but not to differences between realizations of the same category.

In recent decades, much less attention has been given to the issue of contrast in formal linguistics, because it is not as clearly demonstrated in syntax, and in phonology it is subordinated to considerations of the functional economy of the system as a whole. Nevertheless, the three properties outlined appear to hold.

The concept of distinctiveness

Membership in a contrastive category is determined by the presence or absence of one or more distinctive features. Any features that are not distinctive are redundant and have no relevance to category membership.

The discreteness of boundaries

The boundaries between categories are infinitely sharp, in the sense that there are no intermediate forms. If the distinctive features that define membership

in a category are all present in a given token, it is a member of that category, and if they are not all present, it is not.[1]

The irrelevance of phonetics

The basic notion of phonology is that it represents a distributional, behavioral, or mental organization of phonetic material. Everyone would then agree that one could not understand a phonology fully without some information on the phonetic material that had been organized. But once the organization is established, the phonetic structure within the category, and the phonetic relations between categories, is no longer relevant. At one period of structural relations, this led to the suppression of phonetic information in linguistic description that is no longer acceptable. But the amount of phonetic information provided in modern phonologies is severely limited, usually not much more detailed than the abstract set of features that are taken as distinctive. Thus, vowels are described as high, low, or mid, and the finer differences in fronting, backing, raising, lowering, rounding are ignored. It is even less common to find information on the formant trajectory, amplitude contour, or energy distribution.

Most importantly, the phonetic distances between realizations of categories are not considered relevant to their categorical status, unless it amounts to full overlap. This position is the result of cross-linguistic experience that native speakers easily recognize and respond to phonetic differences that are almost imperceptible to the outside investigator at first. But with familiarization and practice, the differences become much more apparent, and eventually the descriptive linguist cannot recapture the original impression that the categories are phonetically close. Such experiences led Bloomfield (1926) to formulate the strong statement: "Such a thing as a "small difference of sound" does not exist in language."

The symmetry of production and perception

It is a general assumption of linguistic analysis, and often formally acknowledged, that grammars or phonologies should be neutral to the perspective of speaker and hearer. It is argued that we are all listeners as we talk and speakers as we listen. This means that a map of instrumental measurements of phoneme production should show the same structure as the output of a categorization or discrimination experiment. The results of the Peterson–Barney experiment (1952) give considerable support to this principle. Ten vowels were pronounced in the identical /h_____d/ framework by 150 different speakers, including men, women, and children, and presented randomly to judges. The measured distributions showed comparatively little overlap, and what overlap exists can be eliminated by a normalization algorithm that is psychologically plausible (Hindle, 1978), whereas the judges' confusion between categories was concentrated largely in the area where merger is known to be taking place in many dialects.

The reliability of intuitions

The working principle that is most important in supporting the categorical view is that native speakers know what they say and know what they cannot say, and that they have free access to this knowledge. As a working principle, it has been quite effective in advancing the study of language, by quickly gathering a great deal of data through the rapid accumulation of introspective judgments. Of course, it is only a working assumption. The more thoughtful theorists have recognized that there is no a priori guarantee that we have access to our intuitions (Chomsky, 1964, p. 939), but in practice, it governs almost everything that linguists do. The principle is not limited to the introspections of the theorist but applies equally well to the elicitations of data from speakers of other languages. Voegelin and Harris (1951) made it clear that the data-gathering procedures of Boas, Sapir, and most other descriptive linguists were simply to "Ask the informant," and the reliability and validity of this method depended on the doubtful assumption that the informant had free mental access to his or her language.

There is ample evidence from experimental phonetics and sociolinguistic fieldwork that native speakers do think categorically about their language. If we believe that the discrete categories of the language exhaust its structure, then the phenomenon of categorical perception (Harnad, 1987; Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni, 1971) will reinforce our confidence in the reliability of native speakers' intuitions. But a profound circularity underlies this confidence. The tendency of the categorical structure to influence perception also makes it difficult for the native speaker to recognize any intermediate forms that are not governed by that structure.

The categorical view of the minimal pair test

The combination of these four concepts in the categorical view leads to a clear prediction about the results of minimal pair tests. When minimal pair tests are done properly, the end result of the procedure is two independent pieces of information: one on how the speaker pronounces the words, and the other on how the subject evaluates any difference between them. These two observations then determine the entry in one of the cells of the four-cell table (1).

(1)

		SPOKEN: same	SPOKEN: different
JUDGED	same	a	b
	different	c	d

It is normally assumed that entries will occur in cells (a) and (d), where speech production and judgments of sameness are matched. This expectation is the direct correlate of the investment of the categorical view in the reliabil-

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

ity of intuitions. It would not be difficult to understand that some entries would occur in cell (c). Subjects are often misled by differences in spelling to think that two words sound different when they pronounce them the same (like *latter* and *ladder* or *which* and *witch*). But it was always assumed that cell (b) must be empty. How could it be otherwise? Speakers could hardly learn to make a difference without perceiving it. The assumption that contrasts were discrete and binary, that there was no such thing as a small difference in sound, that production and perception were symmetrical, and that introspections were reliable, all militated against such an idea. It was therefore very difficult to gain a hearing for any research that reported observations in cell (b).

NEAR-MERGERS

Source *and* sauce *in New York City*

The first indication that this was not the case came from an instrumental study of the productions of /oh/ and /ohr/ in New York City (Labov, Yaeger, & Steiner, 1972: Ch. 6). In the vernacular pronunciation of words like *source*, the /r/ is vocalized to an ingliding shwa. Because the raised tense /oh/ in *sauce* is also ingliding, it was expected that *source* and *sauce* would be indistinguishable. For many speakers, pairs like *god* and *guard* are distinguished by vowel quality, but no subject and no investigator had ever suggested that *source* and *sauce* might be opposed by vowel color.[2] Instrumental studies of the nuclei of /ohr/ and /oh/ words showed a statistically significant pattern, where the nucleus of /ohr/ was further back and/or higher than the nucleus of /oh/ (Labov, Yaeger, & Steiner, 1972: Figure 6-1). What makes the data all the more surprising is that this distribution is the same as that which we find in 100% *r*-pronouncing regions, even though the New York City vernacular has been consistently *r*-less for almost two centuries.

Fool *and* full *in Albuquerque*

Exploratory studies in the southwest United States in the late 1960s showed a widespread merger of /uw/ and /u/ before /l/. It was first observed in Salt Lake City (Labov, Yaeger, & Steiner, 1972:236) and more generally as it spread eastward from a center in the southwest. Considerable variation was found in a study in an Albuquerque high school. One of the subjects whose minimal pair test showed a clear merger was Dan Jones, 16, who consistently judged *pool* and *pull*, *fool* and *full* as "the same." The speech productions of the minimal pair test showed no distinction (Figure 1a). However, an extended interview at his home later in the day produced the pattern of spontaneous speech of Figure 1b. Though there is only a single token of /uw/, in *school*, it is well back of *pull*, *bully*, and *pulled*. A clear distinction also appeared in the productions that Dan Jones made for a commutation test

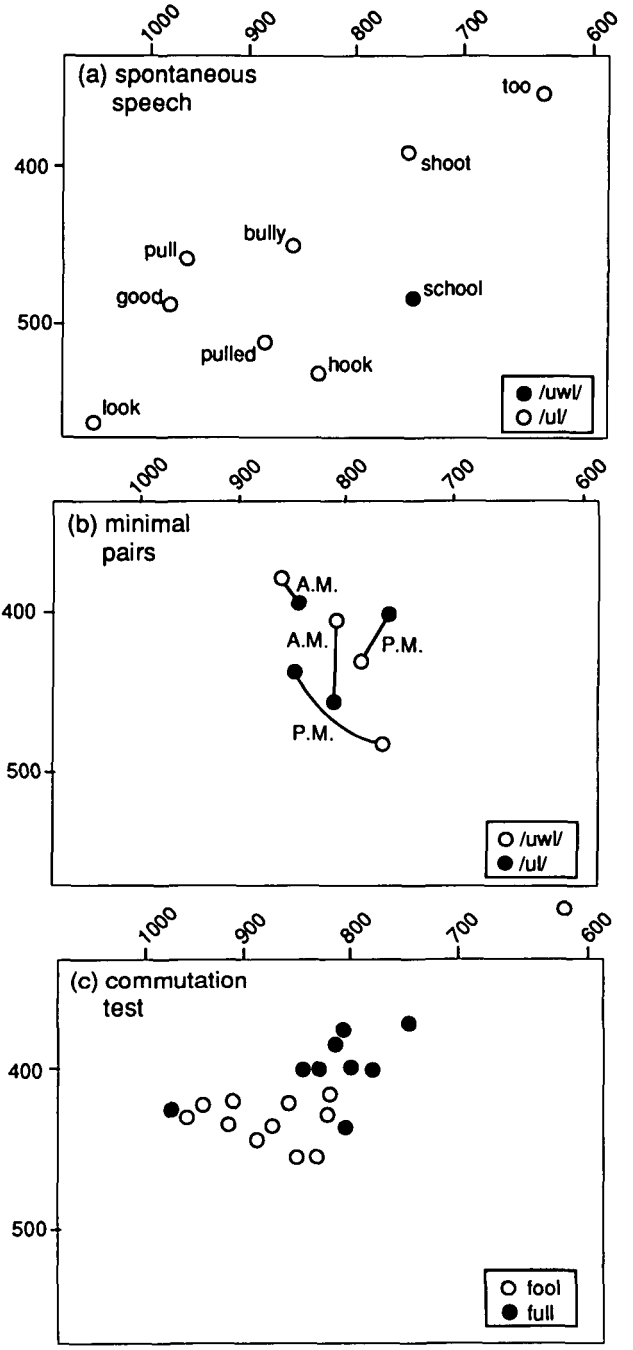


FIGURE 1. Near-merger of /uwl/ and /ul/ in the speech of Dan Jones, 16, Albuquerque, 1969. (a) /uw/ and /u/ in spontaneous speech. (b) Minimal pairs *fool/full* and *pool/pull* in the morning and afternoon interviews. (c) Tokens of *fool* and *full* in the commutation test.

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

TABLE 1. *Categorization and pronunciation of* /uwl/ *in Salt Lake City by age and sex (% responses other than* /uw/ *for* cool, school, pool*)*

	Generation 3 (youngest)		Generation 2		Generation 1 (oldest)	
	Male	Female	Male	Female	Male	Female
Categorization	20	27	17	17	8	27
Pronunciation	20	47	0	8	0	0
N	15	15	12	12	12	15

Source: Di Paolo, 1988: Table 4.

(Figure 1c), where, with one exception, *fool* and *full* show nonoverlapping distributions. His girlfriend and her brother were the judges in the commutation test. They had a great deal of trouble in deciding which of Dan's words were "double-O" and which were "double-L," but they were ultimately correct in 83% of their judgments.[3] Normally, 100% correct is the criterion we use for a passing grade on a commutation test, so these results must be considered marginal.

A more extensive investigation of this merger has since been carried out on Salt Lake City by Di Paolo (1988). High school students and their families were recruited as informants, so that her subjects can be categorized into three generations: the students (Generation 3), their parents (Generation 2), and their grandparents (Generation 1).[4]

In the Vowel Categorization Experiment, subjects first read a table with 10 categories in separate cells, each containing three exemplars of a vowel phoneme. The /uw/ cell contained *moot*, *hoot*, and *food*; the /u/ cell contained *could*, *book*, and *hook*. Subjects then read a list of 39 words, including *cool*, *full*, *pool*, and *pull*, and were asked to locate the cell with the vowel that matched that word, and write the word in that cell. The subjects showed a general laxing tendency for all vowels before /l/, but Table 1 shows results of this experiment that pertain to the /uw/–/u/ distinction.

The figures in Table 1 show a relation between categorization and pronunciation that is important for the understanding of near-mergers. In the youngest generation, pronunciation has moved ahead rapidly, particularly among the females.[5] But in the parents' and grandparents' generations, an earlier stage is visible where a change in categorization is in advance of pronunciation. For these two generations, cell (b) in the four-cell table (1) is filled.

Too *and* toe *in Norwich*

Since Trudgill's (1974) comprehensive report on the speech community of Norwich, its vowel system and those of surrounding communities in Norfolk

have been the site of a number of important discoveries concerning sound change and dialect contact. The fronting of the word class of Middle English (M.E.) o: in *too*, *root*, and so on, follows the path of many other Southern English dialects: both the nucleus and glide are front rounded vowels. In addition, the vowel of M.E. ɔ: in *so*, *stone*, *toe*, and so on, is raised to [u:] in the conservative dialect. Labov, Yaeger, and Steiner (1972) reported that this nucleus was also fronted, though the glide kept its high back target. In an interview with two 14-year-old boys, one showed a near-merger of these two classes. For one (David), it became evident that there was a close approximation of the nuclei of the *too* and the *toe* classes, although the glides went in opposite directions. The other boy (Keith) could not hear the distinction at all in David's speech, though the glide distinction was also preserved in his own speech.

Beer *and* bear *in Norwich*

In evaluating the evidence for the near-merger of *too* and *toe*, Trudgill noted that another case can be found in the apparent merger of the vowels of the set of *beer* and *bear*. All except the oldest informants regard these two word classes as identical (Trudgill, 1974), and dialect poetry rhymes the two:

> Ah, more'n once I'a stopped there jus' to hear
> Their lovely songs that fill the evenin' air.

However, the actual productions of the /ihr/ and /ehr/ classes are quite different for all but the youngest generation of speakers. In this case, the distinction is characteristic of casual speech and tends to disappear in minimal pair tests, as Trudgill's Figure 4.2 shows.

Line *and* loin *in Essex*

Throughout the 18th century, it was consistently reported in Southern England that the word classes of *line* and *loin* had fallen together, yet these two word classes, /ay/ and /oy/, are now separate in most British and all American dialects. The only explanation that had been advanced for this was that the two phonemes separated under the influence of spelling (Jespersen, 1949: 330), a dubious explanation at best in light of the fact that spelling is notoriously ineffective in the face of many other mergers. Labov (1975a) suggested that this may not have been a true merger but a near-merger of the type reported in Labov, Yaeger, and Steiner (1972). Nunberg (1980) supported this suggestion through a study of all the reports of merger for all the words from 1569 to 1799. It appeared that the reports of merger are concentrated in just those allophones that we would expect to be in close approximation as M.E. Ī passed /oy/ in descending to /ay/.

Another approach to investigating this possibility is through a study of current dialects where the 18th-century situation is maintained. One area of

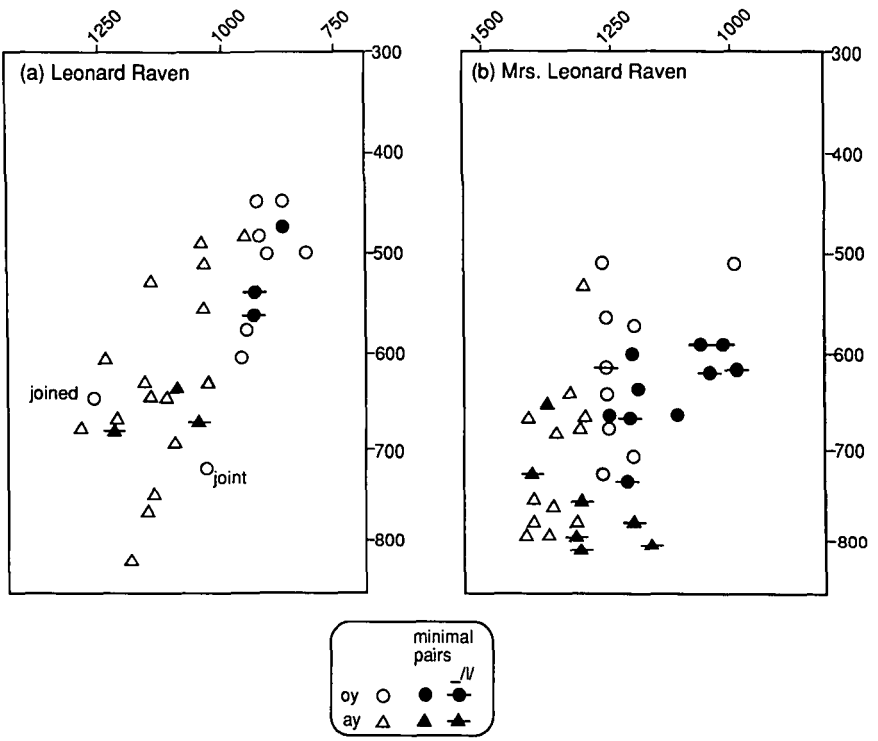


FIGURE 2. Near-merger of /ay/ and /oy/ in the speech of (a) Leonard Raven, 70, and (b) Mrs. Leonard Raven, 69, Tillingham, Essex, 1971–72.

England still shows the merger of /ay/ and /oy/ in the records of the Survey of English Dialects (Orton & Dieth, 1962–67). In the summer of 1971, Labov interviewed three Essex speakers in the village of Tillingham: Jack Cant and Mr. and Mrs. Leonard Raven. All three showed near-mergers. Figure 2 shows the /ay/ and /oy/ tokens for Mr. and Mrs. Raven. Though Mrs. Raven hears *line* and *loin* as "the same" and he hears the pair as "different," they show the same distributions in spontaneous speech and minimal pairs. For both husband and wife, /ay/ is lower and/or less peripheral than /oy/. Words ending in /l/ are more peripheral for both speakers, but the same relations hold between /ayl/ and /oyl/ as with other allophones. Though some of Mrs. Raven's /ay/ tokens are as high as /oy/, these are clearly less peripheral; and though some /ay/ is as peripheral as /oy/, these are clearly lower. There is no difficulty in drawing a boundary between the two sets.

The same relationships hold for Mr. Raven in Figure 2b. One /oy/ word appears to have crossed over into the /ay/ class—*joined*. Otherwise, we can draw a boundary between the two classes: basically, a separation of peripheral from nonperipheral. Again, we note that minimal pairs are closer together than the forms used in connected speech.

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

In the summer of 1972, Labov returned to Tillingham and met again with the three older speakers he had interviewed the year before. He brought with him a commutation test prepared from the spontaneous speech of Jack Cant. The first 10 items were a random alternation of two of his pronunciations of *line* and *loin* in spontaneous, unreflecting speech. Though they were selected as the most closely approximated tokens in spontaneous speech, they are clearly different in height. The second 10 items alternated tokens of *loin* and *line* from speech where Jack Cant was consciously reflecting on whether these words were "the same" or "different." None of the three subjects was able to pass the commutation test.

Meat *and* mate *in Belfast*

A second case of reported merger and separation that has troubled students of the history of English concerns the relations of the word classes of *mate*, *meat*, and *meet*. The first two were widely reported to be merged in the 16th century, but the second two from the 17th century onward. Labov (1975a) suggested that this reported merger was also a case of near-merger. Milroy and Harris investigated this possibility by examining the current phonology of older speakers in Belfast (Harris, 1985; Milroy & Harris, 1980).

The history of Belfast English is better known than that of many other English dialects. Patterson (1880) provided a list of 100 words that contain the reflexes of the *ēa* or **meat** class, including many words spelled with *ea*, but also items such as *Jesus* and *decent*, often spelled *Jay-sus* and *day-cent* in Irish dialect literature. The vowel used in these words is apparently considered identical to the reflexes of the *ā* or **mate** class; *meat*, *please*, *weak* are written as *mate*, *plays*, *wake*. Other scholars who have treated Hiberno-English have generally considered the **meat** and **mate** class to be "the same" (Harris, 1985:241).[6]

The minimal pair and commutation tests that we have used in various investigations are not suitable in the Belfast context, where the vernacular is highly stigmatized. All speakers have access to the more standard system where the **meat** class is merged with **meet**, and in such formal contexts, most speakers give the standard pronunciation. The data were therefore drawn from only 8 of the 50 Belfast inner-city speakers who had made the greatest use of the vernacular **meat** alternants—all men. A total of 60 tokens of the **meat** class were obtained, and 99 of the **mate** class.

The only information we have on the speakers' perceptions of the contrast is therefore limited to observations made by the interviewers in discussions after the more spontaneous part of the interview was completed. Harris wrote, "When BV speakers' attention is drawn to the nonstandard variant, they generally agree that it is identical to the vowel in the **mate** class" (Harris, 1985:241).

To investigate the relations of these word classes in production, four impressionistic levels of height were established for the nuclei:

TABLE 2. *Distribution of meat and mate by vowel height and inglide in the Belfast vernacular*

	Meat		Mate	
Nucleus	Glide	No Glide	Glide	No Glide
1 [ɪ]	0	0	33	0
2 [e]	18	2	54	6
3 [ę]	18	20	4	2
4 [ɛ]	0	2	0	0
Total	36	24	91	8

Source: Harris, 1985: Table 4.2.

(2) 1 [ɪ]
2 [e]
3 [ę]
4 [ɛ]

The highest three nuclei were frequently accompanied by an inglide; the lower mid vowel [ɛ] did not show this, though there were in all only two tokens of this variant.

Differences between the two vowel classes were found in both the distributions by height and the frequencies of inglide, as Table 2 shows. The most striking difference in Table 2 is that more than a third of the tokens of the **mate** class occur with a high nucleus, whereas there are no **meat** tokens at that level. The modal height for **mate** is [e]; for **meat**, it is [ę]. The situation here resembles that in New York City, rather than in Essex, as the amount of physical overlap appears to be considerable. The pronunciation [e^{ə}] is on the whole most common; two-thirds of the **mate** class and one-third of the **meat** class occur with this vowel and are indistinguishable in this impressionistic transcription. On the other hand, the distributions are very different (chi-square yields $p < .01$). One-third of the **mate** class occurs with high vowels, which were not heard with the **meat** class, and two-thirds of the **meat** class are heard with lower mid nuclei, which are rarely associated with the **mate** class.

Another way of summing up the situation is to note that in roughly half of the utterances that include one of these words, listeners would not be able to guess which word was which from the sound alone; in the other half of the cases, they would be able to do so.

It must be pointed out, however, that the two classes are distinct as classes. For any word in the **meat** class, the probability of a lower mid nucleus is about .67; for any word in the **mate** class, the probability of a lower mid nucleus is about .33. The overlap has not prevented the distinction between

the two classes from being maintained for almost 300 years. The first report of merger in Hiberno-English was as early as 1700, and by Garde's (1961) well-established principle of the irreversibility of mergers, the distinction we see here could not have been maintained if complete merger had occurred at that time or any time after. It follows that speakers are capable of tracing the frequency of occurrence of the two classes, as shown in Table 2, and this differential distribution is a part of their fundamental knowledge of the language, one that maintains the two sets of underlying form.

The Belfast situation bears directly on the 16th-century situation in London. It is not suggested that the Belfast vernacular mirrors the London vernacular of the time. In fact, Harris (1985) showed that throughout Ireland, Scotland, and England, there are a great variety of relations among the **meat, mate**, and **meet** classes, any one of which might be closer related to the London of the 17th century than Belfast is.

The general characteristics of the cases just discussed may be summed up as follows:

1. The opposing phonemes are differentiated by a smaller phonetic distance than the normal phonemic difference.
2. This difference is most often an F2 difference, instead of a combination of F1 and F2.
3. There is considerable individual variation within the community. Some individuals show a near-merger, others show a complete merger, and others a distinction.
4. Speakers who make a consistent difference in spontaneous speech often reduce this difference in more monitored styles.
5. Speakers judge the sounds to be the same in minimal pair tests and fail commutation tests.
6. Phoneticians from other areas are better able to hear the difference than the native speakers.

RESISTANCE TO THE CONCEPT OF NEAR-MERGER

The cases of near-merger that have been considered so far are drawn from experiments carried out in the course of fieldwork in the speech community. The largest body of data cited stems from the New York City study, where 20 speakers were examined, but the most clear-cut instances are drawn from the behavior of one individual in each community (Dan Jones, Albuquerque; David Branson, Norwich). Supporting evidence was contributed by the community studies of Di Paolo in Salt Lake City and Trudgill in Norwich, but these did not provide for each individual the clear disjunction between production and perception that defined the near-mergers for these three individuals. Although it is true that every community that shows near-mergers also shows wide individual variation, the individuals who display the phenomenon are not difficult to find.

There is no doubt that the first observations on near-merger were considered paradoxical by those who heard them. At a number of oral presentations of the data, linguists and phoneticians rose to express their disbelief in this phenomenon, despite the wide range of data and the steady accumulation of evidence for it. How can we account for the fact that near-mergers have not been reported before and that linguists find it difficult to accept their existence? First, it is evident that the existence of a near-merger is not consistent with the categorical view outlined at the beginning of this report. In this view, any given element or set of elements is a member of a category or not a member of a category, and any two elements are either members of the same category or not members of the same category. In these cases of near-merger, it is not possible to maintain this binary opposition, and we are forced to recognize an intermediate class of membership.

Second, the existence of near-mergers puts into question the symmetry of production and perception. From the productive viewpoint, there are two categories; from a perceptual one, only one. This is not merely a problem of description or nomenclature. How do speakers learn to articulate each member of one category in one way, and each member of the other category in another way, if they cannot recognize the difference between the categories? This is a substantive issue of some weight. Nevertheless, the fact that our theory does not explain these results should not lead us to reject them.

Before proceeding, it may be helpful to get a clearer view of how the rejection of facts actually operates in the field of linguistics. On June 23, 1977, Labov received a letter from David de Camp, two years before his death in October of 1979. It begins with a reference to "On the Use of the Present to Explain the Past" (Labov, 1975a). De Camp said that he had wrestled with the problem of *meat* and *mate* as a graduate student and had come up with no solution, even though the key to the solution was "literally under my nose: right on top of my desk, as a matter of fact." The main body of the letter explained what he meant by this.

> This spring I had occasion to reread your paper, this time more thoroughly and really digested your pp. 840–849 (Current re-evaluation of speaker reports) which hold the key to your solution, and suddenly I was carried back to June of 1953 and my desperate attempts to defend a dissertation involving precisely this issue. In my study of San Francisco phonology, my sampling techniques (like those of everyone in 1951) were hardly sophisticated and no portable tape recorders were available, but at least I was confident of my phonetics. I had just spent two years doing a lot of raw transcription, checking it with that of other transcribers and against my subsequent retranscription of phonograph records. That is, I was sure until the whole linguistic establishment of Berkeley landed on me in the defense examination.
>
> The trouble was that I followed up my interviews with a perception pair test, and all four mathematical possibilities showed up: people who could hear and produce a contrast, people who could neither hear nor produce, people who could hear but not produce, and even people who could produce but not hear.

My examining committee, supplemented by several other linguists who came along for the fun, grudgingly allowed that the third possibility just might, though probably wouldn't occur, but insisted that the fourth was flat impossible. They insisted that I had just mistranscribed the data. For a fleeting moment I considered bringing one of my San Francisco speakers into the lab at Berkeley where Y. R. Chao had an early-model Kay Sonograph, but I abandoned the idea, thinking that the committee members wouldn't be convinced anyway—and I probably was right. Well, I got the dissertation accepted by hedging the point with lots of might's and it seem's, and perhaps's, and in the portion I published, I toned this point down and even marked it with a self-defensively apologetic exclamation point. I don't think I've ever given you a copy, but I now think you might be interested in it, just as a chapter in history, especially the treatment of this problem.

For all practical purposes, however, your claim still holds that it had "not been reported before . . . that . . . speakers can report two sounds as 'the same' even though they regularly make the distinction in their own natural language," for I was indeed intimidated by such adamant resistance from the entire linguistic establishment, began to doubt the accuracy of my own ears, and certainly did not see the relevance to the problem of *mate/meat* which was also worrying me at just that same time. And so years later, when tape recorders and Sonographs had become common and when rigid structuralist theories (both Tragerian and Hallean) had relaxed so that such theoretically unacceptable facts could at least be reported, the curious San Francisco data were deeply repressed and locked away until even the first quick reading of your paper a few years ago didn't recall them to memory. Obviously there is no moral for our profession in all this, for modern theorists, when confronted by a discrepancy between their theory and the data, would all unhesitatingly question the theory rather than the data, wouldn't they? Wouldn't they? (de Camp, personal communication)

Some unanswered questions about near-mergers

So far, we have introduced considerable evidence of the reality of near-mergers and made it increasingly difficult to accept the categorical view without serious modification. If the testimony of David de Camp has not been enough to establish this point, the data from Essex and Belfast have made it even more necessary to recognize the asymmetry of production that characterizes near-mergers. However, it must be admitted that each of the cases of near-merger introduced so far concerns the speech of a few individuals, who are admittedly not typical of all members of the community. The Essex data concern three persons, who varied in their perception of the *line/loin* contrast among themselves. Even the eight conservative Belfast speakers who gave evidence on the *meat/mate* situation were selected as a special subset of older, conservative males. In every case we have studied, there is wide individual variation in reaction to the marginal phonemic contrast that assumes the form of a near-merger for a certain subset of the population. Does the existence of a near-merger for Speaker A affect Speaker B, who has a complete merger, or Speaker C, who makes a clear distinction? We do not yet

have a clear view of how the phenomenon of near-merger effects an entire community, and how it is located in the speech economy of that community.

The balance of this article presents the results of a series of experiments that illuminate this question and at the same time enrich the data on near-mergers considerably. A much larger number of subjects are drawn into the arena, and we get a much more complete view of the range of perceptual processes connected with the phenomenon of near-merger. The subjects are not selected by any other criterion than their membership in the speech community, so that each batch of subjects spans the range of contrastive relationships for the phonemic distinction in question. We thus get a clearer view of the relation of the near-merger phenomenon to the community as a whole.

The first experiments to be reported here were not my own but were designed by Tore Janson as a study of dialect differences in the categorization of Swedish vowels. He encountered a series of surprising results that he immediately realized were related to the near-merger phenomena reported in Labov, Yaeger, and Steiner (1972). He then formulated a characterization of the underlying process that appears here as the title of this article. Some earlier unpublished and inconclusive experiments of Labov had pointed in the same direction, and the main body of this article is concerned with two series of recently concluded experiments that seem to provide definite answers to the various questions raised earlier.

THE CATEGORIZATION OF SWEDISH VOWELS

Janson and Schulman (1983) reported a set of experiments concerning the categorization of a series of 23 synthetic vowels by groups of Swedish subjects. Their article begins with a direct attack on the discrete binary view of phonological structure that was discussed above as the *categorical view*.

> If a certain phonetic distinction is regularly used to convey distinctions in meaning, it is phonologically distinctive. Thus one of the presuppositions behind the concept of distinctiveness is that a phonetic difference is either distinctive or not distinctive. This can sound innocuous enough: cases in between do not regularly come to mind. However, there is at least one situation which may cause theoretical problems, and that is when phonetic differences are linked with differences in meaning in speech production, but the distinction is not utilized by the perceiver to distinguish meanings. (Janson & Schulman, 1983:321)

The initial condition that triggered this experiment is variation across Swedish dialects in the number of short front vowels. In the Stockholm dialect, there are three long vowels /iː eː ɛː/, but only /i e/ among the short vowels. The present-day /e/ is the result of a relatively recent merger between /e/ and /ɛ/, a merger that has not taken place in areas of Northern Sweden. Though the Stockholm pronunciation has influenced many northern coastal

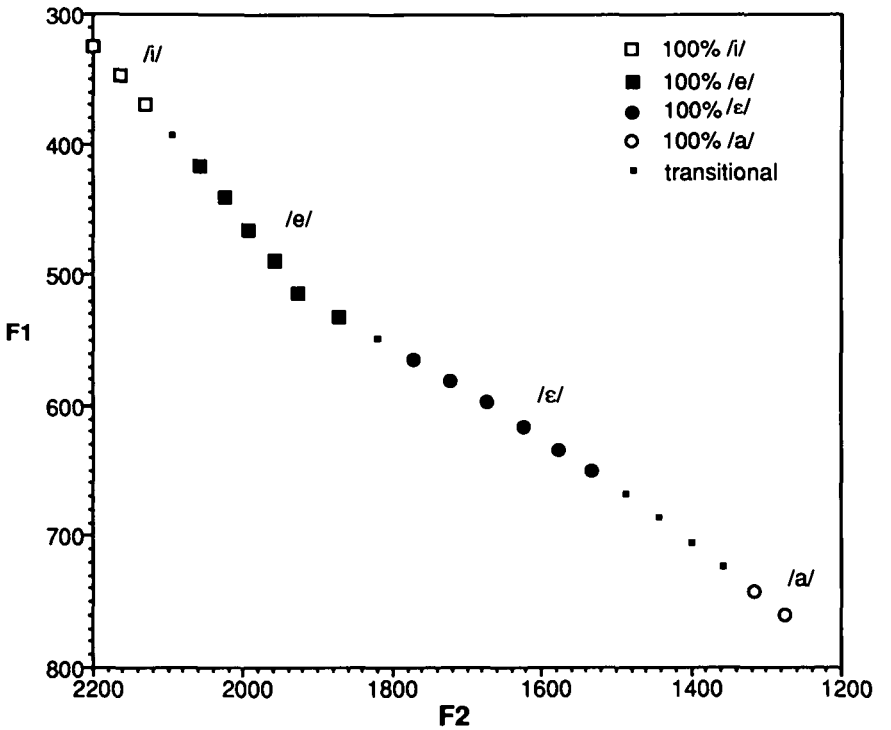


FIGURE 3. Synthetic vowel series for the categorization of Swedish front vowels (from Janson & Schulman, 1983).

areas, inland areas are generally unaffected and maintain a system of three front short vowels /i e ɛ/. Typical of such inland Northern dialects is the municipality of Lycksele, which was used as a site for experiments that would contrast the perception of Swedish vowels.

A series of 23 vowels was prepared, starting with the natural pronunciation of [seːt] by a Stockholm speaker. The original vowel was removed, and a series of 23 synthetic vowels inserted, using the OVE3 synthesizer to synthesize the vowels and adjust formant transitions for maximum naturalness. The F1 and F2 measurements of the 23 vowels are shown in Figure 3. Subjects were first asked to read a list of 18 monosyllabic Swedish words, including two tokens of *sett* and two of *sätt*. The formant values of the recorded *sett* and *sätt* vowels were analyzed to obtain measures of production. Subjects were then asked to categorize a randomized series of 5 each of the 23 stimuli, as one of the Swedish words *sitt*, *sett*, *sätt*, *satt*. In Figure 3, the results are indicated for a Northern Swedish speaker with a system of four vowels—not a subject of the main experimental series. In this example, there are four clear areas of 100% categorization as /i/, /e/, /ɛ/, and /a/, sepa-

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

TABLE 3. *Width of regions of uncertainty for perception of Swedish vowels*

	Boundary		
	/i-e/	/e-ɛ/	/ɛ-a/
Stockholm (n = 15)	2.9	7.1	1.5
Lycksele (n = 43)	3.7	6.9	2.5

Source: Janson and Schulman, 1983:327.

rated by two brief transition zones of one item, and one more extended transition of three items, between the third and fourth vowels. This clear categorization into four vowels is the pattern that was expected from Lycksele subjects, if there were any correspondence between the production and perception of these vowel distinctions.

A first experiment was conducted with 15 Stockholm residents, born and raised in the city and suburbs, and 43 students at a Lycksele high school. As far as speech production was concerned, there were no surprises. The Stockholm speakers showed no significant difference between *sett* and *sätt*. A random sample of 10 of the 43 Lycksele students showed significant F1 and F2 differences for 9, and a significant F2 difference for the tenth. Thus, the Stockholmers plainly show a three-vowel system, and the Lycksele subjects a four-vowel system.

Results of the perception test

The results of the Janson and Schulman perception tests are given in terms of the width of the region of uncertainty at each boundary. The region of uncertainty is the number of contiguous stimuli vowels between the regions where the subject shows no uncertainty, or 100%. Thus in Figure 3, the uncertainty of the /i-e/ boundary is 1, /e-ɛ/ is 2, and /ɛ-a/ is 3. Table 3 shows the mean uncertainty for the two subject groups in their first experiment.

We have every reason to expect that Table 3 would show sizeable differences between the two groups of subjects at the /e-ɛ/ boundary, but instead they are almost identical. Furthermore, the width of uncertainty is so large that it occupies the entire region of the two intermediate vowels. Figure 4 shows the results in more detail for a typical subject in the Lycksele group. It is obvious that there is no distinction perceptually between the two mid vowels. In other words, the perceptual reactions of the Lycksele students in these experiments are identical with the Stockholm subjects, even though their productive system is radically different.

Though these synthetic stimuli were systematic and carefully controlled, subjects may have missed cues from natural speech or have been uncon-

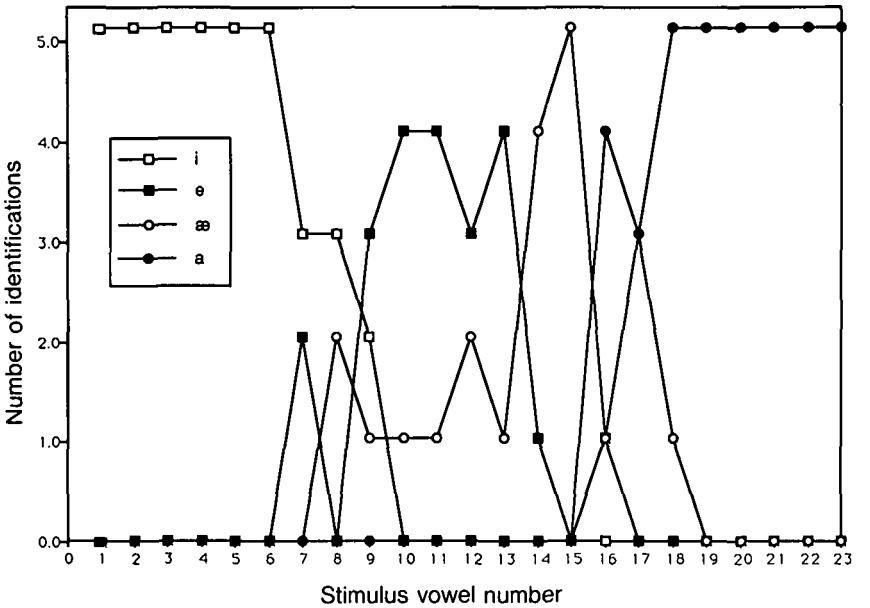


FIGURE 4. Categorization of the /s_____t/ continuum by a typical subject in the Lycksele group (from Janson & Schulman, 1983).

sciously persuaded that the speaker was from Stockholm. A second experiment addressed these problems by using as stimuli the natural pronunciations of *sitt*, *sett*, *sätt*, and *satt* from six Lycksele speakers and making it explicit that the speaker was from Lycksele. The subjects were 34 Lycksele high school students and 34 Stockholm natives. The results reflected the same general picture. For the identification of *sitt* and *satt*, the mean error rate in Lycksele was 1.1%; in Stockholm, 1.0%. For the identification of *sett* and *sätt*, the error rate in Lycksele was 21%; in Stockholm, 36%. One of the most important parallels with our other studies of asymmetry of production and perception is in the heterogeneous character of the subjects' behavior. Table 4 shows the distribution of error rates for the two mid vowels for both cities. As there were two copies of two pronunciations of each word by six speakers in the test tape, the total number of possible errors was 48 for each subject.

In Lycksele, seven subjects produced the error-free results that match their productive pattern, and there were four subjects who showed the high rate of error, approaching chance. On the other hand, many Stockholm speakers were at the chance level, and none performed perfectly. Thus, the Lycksele community is not equivalent to Stockholm in its perception of the mid vowels but shows a variety of different reactions to the task, as we have found in other communities.

TABLE 4. *Error distributions for perception of natural Swedish front short vowels*

	Lycksele	Stockholm
Group 1 (0–3 errors)	7	0
Group 2 (4–18 errors)	23	17
Group 3 (19+ errors)	4	17

Source: Janson and Schulman, 1983:330–331.

Perhaps the most remarkable aspect of this work concerns a third experiment. Another group of Stockholm subjects was presented with the test tape of 23 synthesized vowels and asked to identify the English words *sit*, *set*, *sat*, and *sot*.[7] The region of uncertainty for *set* and *sat* was only 3.8 vowels, almost 50% smaller than the 7.1 of the first experiment. Thus, the difficulty in distinguishing between the mid vowels cannot be that the task is physiologically or acoustically difficult. The interfering factor appears to be that Stockholm and Lycksele subjects define [e] and [ɛ] as "the same," and before they can respond to any categorization task, this linguistic definition intervenes. It is evident that linguistic norms, which may be socially loaded or socially neutral, can affect tasks in ways that are independent of the productive system.

In evaluating this situation, Janson and Schulman made an immediate connection with the data on near-mergers of Labov, Yaeger, and Steiner (1972). They concluded that their subjects "maintain a distinction in their speech production which is not used for semantic differentiation between words in their own or their listeners' perception" (Janson & Schulman, 1983:333). They then proceeded to inquire more generally into the function of nondistinctive features and argued that the productive distinction between [e] and [ɛ] in Lycksele may function as a marker of local identity without serving to distinguish words. How then does it come about that these speakers do not distinguish [e] and [ɛ] in perception? Janson and Schulman gave the following characterization of the situation:[8]

> Language users who employ an ordinary phonological distinction are exposed to speech in another dialect and/or another style in which this distinction is not upheld. Their response to this is to cease using the distinction for semantic differentiation in perception, which makes sense, since it has become inefficient for that purpose. (Janson & Schulman, 1983:335)

The Janson and Schulman experiments have given us a much broader view of the place of near-mergers in the speech community than we had before. Nevertheless, their inference about the suspension of semantic contrast is made on the basis of perception experiments that are not directly related

to the process of semantic interpretation. There is some evidence that in the formal experimental situation, the social norms of the Stockholm dialect are interfering with the Lycksele speakers' use of their own dialect. Is it not possible that speakers of the language use the nondistinctive feature in the unreflecting interpretation of speech in everyday life to make distinctions they do not make in formal situations? A series of experiments in Philadelphia yields a much more direct answer to this question and gives us a more detailed picture of the range of behaviors that characterize near-mergers in the speech community.

THE IMPAIRMENT OF SEMANTIC CONTRAST IN PHILADELPHIA

In pursuing the study of near-mergers, we were fortunate enough to find a case in the Philadelphia speech community. The Project on Linguistic Change and Variation in Philadelphia encountered a near-merger in the opposition of short /e/ and/ʌ/ before intervocalic /r/.[9]

There are many well-known variations in the number of contrasting vowels in this position, illustrated in (3).

(3)

I	II	III	IV	V
Mary	Mary / Merry			Mary
merry		Mary / merry / marry	Mary / merry / marry / Murray	merry / Murray
marry	marry			marry
Murray	Murray	Murray		

The geographic distribution of these systems has never been fully mapped. System I, with four different vowels, is found in New York City, the South, and areas of New England. System II, with a merger of *Mary* and *merry*, is characteristic of northern New Jersey and Maine. System III is quite widespread throughout the Midland and the West. System IV is found in a few areas of the West with strong /r/ constriction. System V is Philadelphia, where the short /e/ is centralized, quite distinct from both *Mary* and *marry* but merged or in close approximation to *Murray*.

In our study of sound change in Philadelphia, we found subjects with complete merger of the /er/–/ʌr/ contrast, with a clear distinction, and with a near-merger. The majority of Philadelphians have only a small difference in F2 between vowels in the /er/ class and the /ʌr/ class. In spontaneous speech, we hear this most frequently as a mid-central vowel in the common

TABLE 5. *Distribution of* /er/ ~ /ʌr/ *contrasts: Speech production as shown in commutation tests*

		Philadelphia		Phila. suburbs		Non-Phila.	
		Same	Different	Same	Different	Same	Different
Minimal pair Judgments	Same	4	6	1	0	0	1
	Different	0	11	0	3	0	13

words *very* and *terrible*, and in the name of the suburb *Merion*. When we conduct minimal pair tests with *ferry/furry* and *merry/Murray*, we find three of the four cells filled regularly. Minimal pair tests give us only one or two items to judge "same" and "different" from the point of view of speech production, and it is often hard to say whether these tokens are physically "the same" or "different." To get a more exact estimation of the proportions of merger, near-merger, and distinction, we can use the production information from commutation tests, which gives us enough information to perform *t* tests on the significance of the difference between means, combining this information with the judgments given in minimal pair tests. A recent experimental series produced the results shown in Table 5, combining commutation test information with the judgments given in minimal pair tests. Within the city, we find almost one-third of the speakers show the phenomenon of near-merger. The suburbs show patterns much more similar to the dialects outside Philadelphia, though as we will see, there are some influences of membership in the Philadelphia speech community. The one subject from outside Philadelphia who said that *ferry* and *furry* were "the same" must have been temporarily confused, since in the commutation test itself, judging his own productions, he scored 100% like all others from outside the city.

The near-merger of *ferry* and *furry*, like many others, shows a close approximation along the F2 dimension and no significant difference on the F1 dimension. Short /e/ before intervocalic /r/ shifts towards the center, where it rests next to /ʌ/ in a nonoverlapping or partially overlapping distribution. Figure 5 shows the *ferry/furry* tokens produced in commutation tests by four speakers. In each diagram, the position of the /e/ and /ʌ/ individual tokens are shown as squares and diamonds, respectively, and the mean is plotted for each. The tokens that were labeled correctly by the subject are open figures; those that were incorrect are solid. The shaded squares are tokens in the list that were read but were not used in the commutation test; they are used in calculating the mean. In this table, the mean values for each word class are given for F1 and F2, with the differences, and a *t* test for the significance of difference in the means. It should be remembered that for non-Philadelphians, these are not neighboring word classes like /i/ and /e/. Their means are quite widely separated, and they show no vowels in close approximation.

In Figure 5a, a clear distinction is made by Caroline F., with a mean F2

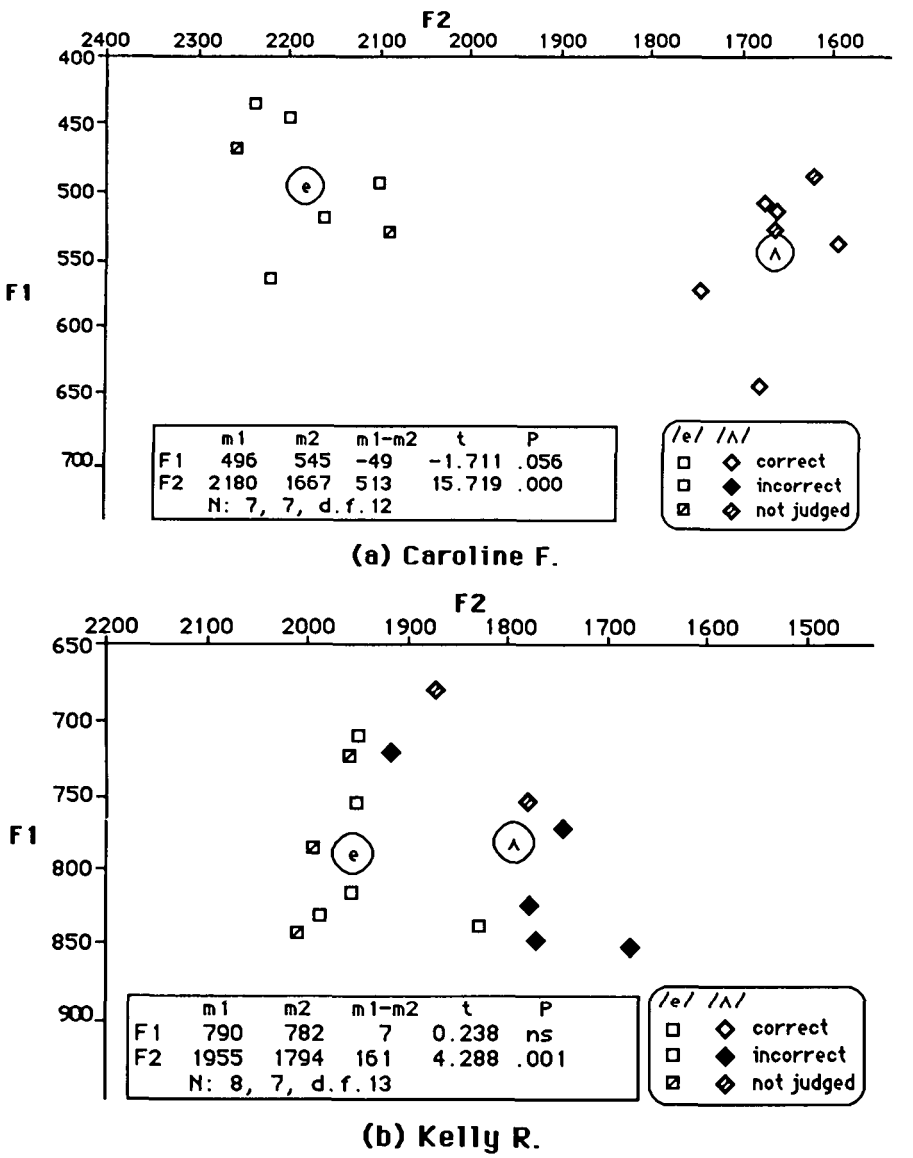
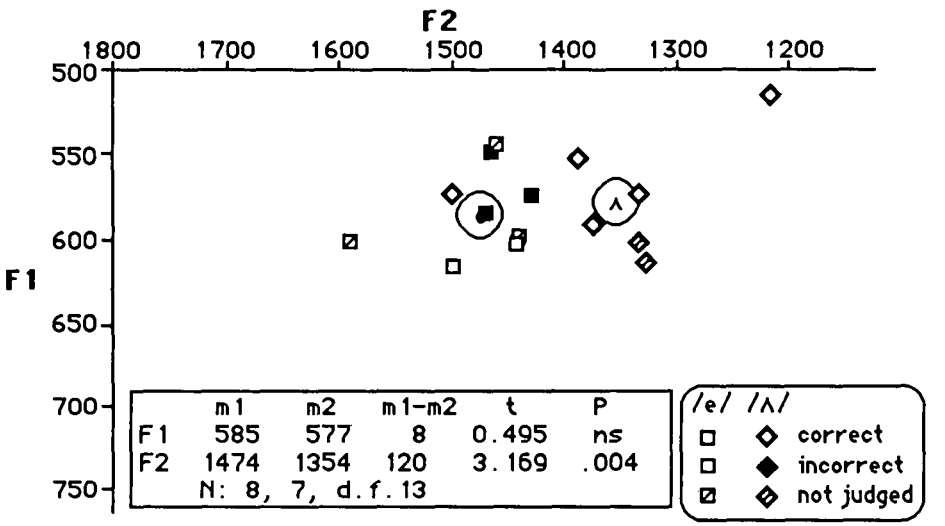


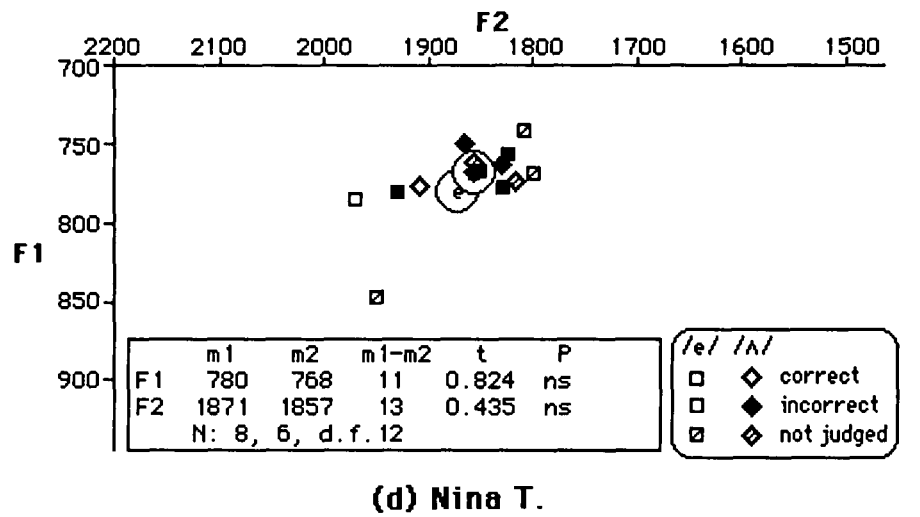
FIGURE 5. Acoustic locations and means of nuclei of *ferry* and *furry* produced by four Philadelphia subjects for the commutation test. (a) Caroline F., clear distinction. (b) Kelly R., nonoverlapping near-merger. (*figure continues on page 56*)

difference of 513 Hz between the two classes, and no /e/ token closer than 300 Hz to the nearest /ʌ/ token. There are no solid squares or diamonds, since she was one of the two Philadelphians who scored 100% in the commutation test.

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press



(c) Matt W.



(d) Nina T.

FIGURE 5. (*continued*) Acoustic locations and means of nuclei of *ferry* and *furry* produced by four Philadelphia subjects for the commutation test. (c) Matt W., overlapping near-merger. (d) Nina T., complete merger.

In Figure 5b, a typical near-merger pattern appears in the commutation test tokens of Kelly R. There is a nonoverlapping distribution of /e/ and /ʌ/, and a difference between the F2 means of 161 Hz that is significant at the level of $p < .001$. Though this difference is greater than the just noticeable difference from a psycho-acoustic point of view, it is considerably less than

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

the difference needed for a reliable linguistic distinction. We observe that several tokens of opposing word classes are in close approximation and obviously could not be distinguished one from the other. However, this does not account for Kelly R.'s own performance. She guessed "ferry" for all ten words.

A second type of near-merger is shown in Figure 5c, for Matt W. Again, there is a significant difference in the F2 means at the .004 level, but it is reduced now to 120 Hz, and there are several overlapping tokens. This pattern is similar to that of the New York City *source/sauce* pattern. Though Matt W. scored only 7 out of 10, it can be observed that he labeled correctly the extreme tokens of /e/ and /ʌ/. In this respect, his ability to categorize his own productions was superior to that of Kelly R.

Finally Figure 5d shows a total merger in the commutation test pattern for Nina T. There are no significant differences between the means, and /e/ tokens are just as likely to be found to the back of /ʌ/ as to the front. They are impossible for the speaker to label correctly.

These diagrams show the physical correlates of the near-merger situation. Many of those who reported a distinction in the minimal pair test cannot pass a commutation test so that of the 11 subjects listed in Table 5 who reported a distinction, only 5 showed the type of clear and reliable distinction characteristic of non-Philadelphians and exemplified by Figure 5a.

THE LABELING FUNCTION VERSUS SPEECH PERCEPTION

The data just given continue to confirm the finding that speakers are capable of maintaining a distinction in speech which is so narrow that it qualifies as "a small difference in sound." It remains to be seen what kind of a role this difference plays in the linguistic system. It is a common characteristic of the near-mergers discussed so far that the critical evidence is not found in spontaneous speech but in highly formal situations where the maximum attention is focused on speech. In the investigation of chain shifts, on the other hand, we relied entirely on data from spontaneous speech. Evidence from controlled styles is much less reliable, since many of the important allophonic differences are wiped out, and, depending on the particular sociolinguistic configuration, the mean values may shift radically backwards towards an older, corrected value, or radically forwards towards the apparent target of the change. Though we can give a sociolinguistically sophisticated account of subjects' behavior in these formal situations, a description based on data of these kind does not have the inherent validity of the study of spontaneous speech.

We then return to the problem of the near-mergers with the question as to whether the asymmetry of production and perception is an artifact of commutation tests and minimal pairs. These formal situations ask the speakers to perform a metalinguistic act, the *labeling* of linguistic categories. Subjects

are asked to associate a spoken form with a category by applying a label such as "animal," "double-L," "Merry Christmas," or *sett*. This is quite distinct from the *perception* of features or contrasts, which is a largely unconscious act that enters into the *interpretation* of utterances.[10]

The behavior that we study in minimal pair and commutation tests cannot be considered irrelevant to linguistic structure. Throughout this article, we see that those who have physically distinct and well-separated phonemic categories have no difficulty in applying labels to them, and they pass commutation tests with the expected 100% success. The impairment of the labeling function is therefore associated with the marginality of the phonemic distinction, in a combination of three factors: (1) physical approximation of the targets, (2) widespread individual variation in the community, and (3) loss of the labeling function itself. But as long as our data are confined to the labeling function, we cannot know whether or not the normal phonemic function—the use of a phonemic distinction to distinguish words in unreflecting interpretation—is also impaired.

The Coach Test

Because it is not likely that we will find the distinction between *ferry* and *furry*, *merry* and *Murray*, *Kerry* and *curry*, or any other minimal pair in spontaneous speech, it was necessary to devise an experiment that would test subjects' ability to use the /er/–/ʌr/ distinction in an act of unreflecting semantic interpretation. To do this, it would be important that the subject not be in the least conscious of this distinction, and be reflecting about nothing else than the actual matters being presented by language, that is, the content and not the form.

The general format for such an experiment has been developed in a number of previous experiments, in which the subjects use their linguistic competence without reflection or introspection, applying the rules of grammar to the interpretation of sentences in a natural context. The type of experiment may be called Semantic Disambiguation. It involves the following structure:

1. Subjects are asked to listen to a narrative and give their own judgment as to what was the right or wrong thing to do, or whether the act performed was right or wrong.[11]
2. In one sentence near the end of the narrative, a sentence is constructed that involves the crucial distinction being studied, so that if one form or the other is used, the interpretation of the actual events being described will be radically different. Subjects are randomly given versions of the narrative with one or the other of the forms being tested.
3. The narrative is continued to the end with sentences that are all ambiguous in regard to the critical event and are necessarily interpreted by the listener in a way consistent with the interpretation made in (2).

4. Subjects are asked for their opinions, and the discussion is continued until it is clear which interpretation has been made in (2).

The semantic interpretations being studied in these experiments are often very subtle. When people are asked to listen to a single sentence, and immediately given an interpretation out of context, the results are often unreliable and unstable. But in the Semantic Disambiguation experiments, the interpretations are fixed in a context and serve as the basis for further interpretations, so that when the experimenter intervenes, there is usually no doubt about what meaning the subject has extracted from the text.

Labov (1975b) reported the Jaywalking Experiment, which was designed to examine subjects' interpretation of the *get* versus the *be* passive, with or without disambiguating purpose clauses. Labov (1988) described the results of the Legal Advice Experiment, designed to test the negative effect of the quantifier *any* versus the determiner *the*, in a study of the objectivity of a legal notice being sent to clients in a class action suit. The Coach Test used the same basic strategies to test subjects' ability to use the /er/–/ʌr/ distinction for unreflecting semantic interpretation. To achieve that goal, the test had to be somewhat more elaborate than the others just mentioned.

The design involved a minimal pair that would not appear to subjects as a minimal pair. Instead of opposing *merry* to *Murray*, the opposition made use of the name of the Philadelphia suburb *Merion* and the name *Murray* combined with *in*, that is, *Murray in*. To do so, a story was written that would motivate the nickname *Merion* applied to a girl and a moral choice involving the selection of a boy or a girl. The story concerns the coach of a local school team, who were known as the "Also Rans" because they regularly finished second. One of the players was *Murray*, an earnest but hopeless fielder who dropped every ball that came to him. The coach made him happy by naming him his first utility outfielder. The Board of Education then declared that girls could try out for the team, and a girl who had just come to the school tried out. She couldn't hit, but she could catch the ball fairly well. Her mother was a very aggressive type from Upper Merion, who kept needling the coach for not playing her daughter. The coach was exasperated and began calling the girl *Merion*. He finally placated her mother by naming her the first utility outfielder, sure that he would never need one. But at the crucial game, when the team might finally finish first, his center fielder tripped, fell, and was carried off. The coach was then faced with a difficult decision.

"I just don't know," he says. "If I put Murray in there and they hit one ball to the outfield, we come in second, again. The Council will nail my hide to the wall. But if I put a girl in there it will break the kid's heart. But if they hit something out there, that Merion might just catch it."

So he thought about it a minute. "That's it.

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

A I gotta play Merion there."
B I gotta play Murray in there."

So that's what he does. So our pitcher gets hot, right. And none of them hit the ball out of the infield. And we were really ahead. But Coach wasn't happy. He was sittin' there all worried. He says to me, "Do you think I did the right thing?"

Anyway, we won. And everybody was cheering away, but I didn't see Coach. He just disappeared. But I often use to wonder about that. You know: Did he do the right thing?

Subjects alternately hear a version with sentence A, indicating that the coach chose the girl, or sentence B, indicating that he chose the boy. The subject is asked whether the coach did the right thing or not, and in the course of the discussion, the interviewer finds out whether the subject understood that the coach selected a boy or a girl.

The Coach Test is introduced with the following instructions:

I'd like to play a story for you that was told to us by a Philadelphian. It's about a problem that came up a few years ago that has to do with the right thing to do in a difficult situation, and we'd like to get your opinion about it.

The story is then played, and at the end, the interviewer asks: "What do you think? Was it the right thing to do?" After the subject gives his or her opinion, the interviewer says:

Let me play the Coach's argument again, so you can get an idea of what his thinking was. Do you agree with him?

The interviewer then plays again the section beginning with "I just don't know" to the end of the crucial sentences, "I gotta play Merion/Murray in there." In this replay, the stimulus word is switched, so that *Merion* is replaced by *Murray in*, and vice versa. For those non-Philadelphians who have a clear distinction between /er/ and /ʌr/, it is normal for subjects to simply change their interpretations, assuming that they had not heard right the first time. Occasionally, subjects detect an ambiguity, but they are never aware that the key terms have been switched. The reversal of the tokens of /er/ and /ʌr/ provides the most strict test of the phonemic functioning of the contrast. If /e/ and /ʌ/ are perceived as different phonemes in this context, then even in the most unfavorable situation where the opposite interpretation has already been made, the presence of one or the other will lead a listener to reverse that interpretation.

The experimenter does not usually accept the subject's first remark but asks probing questions to continue the discussion until the underlying semantic interpretation is clearly evident. For example, the following response to

a version where the stimulus is *Merion* is easier to rate as "correct" after the second remark.

Experimenter:	What do you think? Did he do the right thing?
Subject:	I would've put the other guy in because I gave him my word first!
Experimenter:	How do you think the girl's mother reacted?
Subject:	She probably liked it.

At times, the experimenter pursues a possible ambiguity until he encounters a reflection on the critical utterance. The following response to a version where the stimulus is *Murray in* shows an unusual sensitivity to the linguistic event.

Experimenter:	What do you think? Did he do the right thing?
Subject:	Do I think he did the right thing by putting Murray in? I'm not sure.
Experimenter:	How do you think the girl's mother reacted when he made that choice?
Subject:	She probably wasn't too happy about it. I took it as Murray was the guy they put in.
Experimenter:	Why are you saying that?
Subject:	Because of the way he said it.
Experimenter:	How's that?
Subject:	The way that the coach said he was going to put "Murray in" instead of "Merion."

The Coach story was told by David De Pue, a graduate student at the University of Pennsylvania who was a native Philadelphian.[12] We prepared one version with his natural pronunciation of *Merion* and another with his pronunciation of *Murray in*. These were closely approximated, like that of most Philadelphians. I also asked him to imitate my own much more distinct pronunciations, which he did in an accurate and natural manner. We prepared versions with these distinct forms, separated by a distance of 250 Hz F2. We decided to begin the experiment with the more distinct forms and use the natural forms only if local subjects distinguished the first set accurately. As this did not happen, all of the Coach Test data were gathered with the more distinct forms of the test stimuli.

When the Coach Test was first designed in 1976, it was applied to 15 Philadelphia subjects from the Neighborhood Study. In all of these tests, the stimulus was the A form, where the coach selected the girl. Responses showed close to 50% interpretation as male and 50% as female, and we came to the provisional conclusion that Philadelphians did not use this distinction in their unreflecting semantic interpretations. In 1988, the Project on Cross-Dialectal Comprehension began a controlled study of the /er/–/ʌr/ distinction

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

TABLE 6. *Results of commutation test of* ferry/furry

	Pass	Fail
Philadelphians	2	19
Non-Philadelphians	15	0

across dialectal boundaries, contrasting Philadelphians with non-Philadelphians. In this use of the Coach Test, the A and B conditions alternated, and the Coach Test was combined with minimal pair and commutation tests to give a full view of production and perception.

EXPERIMENT 1: THE CONTRAST OF /er/ AND /ʌr/ IN PHILADELPHIA

Experiment 1 included the following procedures:

1. a demographic inquiry into subject's geographic and linguistic background;
2. the Coach Test;
3. minimal pair tests including *ferry/furry*; and
4. a commutation test. Subjects recorded a randomized list of seven tokens of *ferry* and seven of *furry*. The investigator then began at a point on the list unknown to the subject and asked the subject to identify 10 words (five *ferry* and five *furry*) by associating the word "animal" with *furry* and the word "boat" with *ferry*.

The formant positions of the nuclei of all *ferry* and *furry* productions were measured, and each subject classified according to the degree of overlap of /er/ with /ʌr/.[13]

Thirty-six subjects, all students at the University of Pennsylvania or Drexel University, were exposed to this battery of tests: 21 from the Philadelphia speech community and 15 from outside the region. In addition, 22 subjects were tested as a group at Drexel University with written responses; 12 were from the Philadelphia area.

The results illuminated the relationships among production, self-evaluation, performance in commutation tests, and perception of the distinction in discourse.

The commutation tests

We first see a dramatic difference in the commutation test. As Table 6 shows, the difference between Philadelphians and non-Philadelphians is almost categorical. Here "pass" means 10 out of 10 correct. This is a reasonable crite-

TABLE 7. *Relation of minimal pair response to success on commutation test*

	Mean % Correct
Non-Philadelphians ($N = 15$)	100
Philadelphians who respond "different" in minimal pairs ($N = 10$)	78
Philadelphians who respond "same" in minimal pairs ($N = 11$)	55
Philadelphians with nonoverlapping vowel patterns ($N = 11$)	79
Philadelphians with overlapping vowel patterns ($N = 10$)	51

rion, in view of the fact that non-Philadelphians pass it with ease. When a normal phonemic distinction is involved, with normal margins of security, speakers have no difficulty in showing perfect performance on this test. On the other hand, only two Philadelphians managed to do so. Their inability is not a product of the fact that they all show merger, however. Philadelphians show a wide variety in both categorization (the minimal pair test) and production (distribution of tokens in the commutation test). Of the 21 Philadelphians, 10 said that *ferry* was different from *furry*, and this difference was reflected in percent correct on the commutation test, as shown in the upper half of Table 7. Their mean scores are significantly better than those who reported *ferry* and *furry* as "the same" ($t = 2.38$, $p = .014$), even though they are considerably worse than the non-Philadelphians ($t = 3.90$, $p < .001$).

This difference in performance on the commutation test can be directly related to the task that the subjects had to perform. Instrumental measurements of the tokens produced by subjects in the commutation test showed that half of them were like Matt W. or Nina T. in Figure 5c,d: they showed overlap in /e/ and/ʌ/ before /r/ that made it impossible to obtain 100% correct scores.[14] The other half showed no overlap and resembled either Caroline or Matt W. in Figure 5a,b. Without taking account of the more subtle point of the size of the margin of security (a vs. b in Figure 5) or the degree of overlap (c vs. d), we can ask whether the simple fact of overlap can predict success on the commutation test. The lower half of Table 7 shows that it does. The 10 Philadelphians who produced nonoverlapping distributions were able to categorize them more successfully than those who produced overlapping distributions ($t = 3.34$, $p = .0017$). But again, they were significantly worse than the non-Philadelphians ($t = 3.85$, $p < .001$). Indeed it is only natural that we should get the same results for both comparisons, since the two sets of subjects are the same, with the exception of one subject.[15]

The fact that the token distribution predicts success in identification indicates that we are not necessarily dealing with any difference between the perceptual system of Philadelphians and non-Philadelphians. It may be that given the Philadelphia tokens, the non-Philadelphians would do no better on a commutation test. Nor is there any evidence that Philadelphians would not be able to attend just as well to the difference between *ferry* and *furry* if they

TABLE 8. *Results of the Coach Test for Philadelphians and non-Philadelphians*

	Correct	Incorrect
Philadelphians	14	7
Non-Philadelphians	14	1

$\chi^2 = 3/6$; $p = .06$.

had produced tokens as distinct as those of the non-Philadelphians. So far, then, the results confirm the existence of widespread variation in Philadelphia but do not indicate any fundamental alteration in the linguistic system. To examine this question, we must compare responses to the Coach Test.

Results of the Coach Test

The Coach Test presents both Philadelphians and non-Philadelphians with the same stimuli. The results should answer two questions:

1. Does the impairment of Philadelphians ability to distinguish /er/ and /ʌr/ hold for unreflecting semantic interpretation?
2. Is there any difference between Philadelphians and non-Philadelphians in this ability?

Table 8 gives positive answers to both questions. Philadelphians are worse than non-Philadelphians in their ability to interpret /er/ and /ʌr/ as indicating a girl and a boy, respectively.[16] The non-Philadelphians did extraordinarily well on the Coach Test, with only 1 wrong out of 15. The Philadelphians were worse, but their performance was well above chance, and the difference is not quite at the .05 level of significance.[17] Unlike the commutation test, there was no observable correlation between results on the Coach Test and self-evaluation or vowel production. Table 9 shows the distribution of Coach Test answers when broken down by subjects' productions and self-evaluation on the minimal pair test.

EXPERIMENT 2: A STANDARDIZED COMMUTATION TEST

While the results of Experiment 1 clearly show the suspension of phonemic contrast in Philadelphia, the strong results of the commutation tests do not allow us to project a difference in the perceptual systems of local speakers. Success in categorizing items on the commutation test was a direct product of the tokens that the individual produced. We now consider some results of Experiment 2, which confirm and extend these observations by allowing us

TABLE 9. *Coach Test results for Philadelphians by self-report, production, and commutation score*

	Correct	Incorrect
Minimal pairs		
reporting merger	8	3
reporting distinction	6	4
Vowel production		
with overlap	7	3
without overlap	7	4
Commutation scores		
high	4	2
medium	1	2
low	9	3

to compare the abilities of Philadelphians and non-Philadelphians in categorizing the same series of vowel productions.[18]

Two speakers were chosen for a standard comparison: one that posed a very difficult problem for categorization, and one that was relatively easy, within the characteristic Philadelphia range. Figure 6 shows the extended commutation test results for Laura M., who read an extended list of 42 tokens. There is a great deal of overlap, yet we cannot call this system a merger. The difference in the F2 means is less than any we have seen, only 72 Hz, yet it is also significant, with a chi-square of 2.476 ($p = .009$). Half of the /er/ tokens are more front than any /ʌr/token, though the differences involved here are not much more than 100 Hz. A commutation test was prepared from a random alternation of five each of the two tokens indicated by surrounding squares on Figure 6.

Figure 7 shows the *ferry* and *furry* tokens produced by Jed F., a suburban Philadelphian who shows a clear distinction, with a mean F2 difference of 681 Hz. While this difference is not as great as that of most non-Philadelphians, it is representative of the type of distinction made in the city. No tokens are close to the neighboring word class, and the separation between the two distributions is sizeable. The commutation test was prepared from the two tokens marked with squares on Figure 7. The /ʌr/ token was the only one that was wrongly labeled by Jed F. himself. However, he corrected himself after a short hesitation. This is the one that is closest to the /er/ distribution, but it is still 200 Hz from the /er/ mean and not at all difficult for non-Philadelphians to recognize as *furry*.

Experiment 2 recruited 57 subjects from a wider range than the subjects of Experiment 1: 17 subjects were from outside of the Philadelphia area, 5 from the suburbs, and the rest from Philadelphia. Most of the subjects were recruited from the University of Pennsylvania community, but a group of 8 were included from working class areas in North Philadelphia.[19] Fourteen

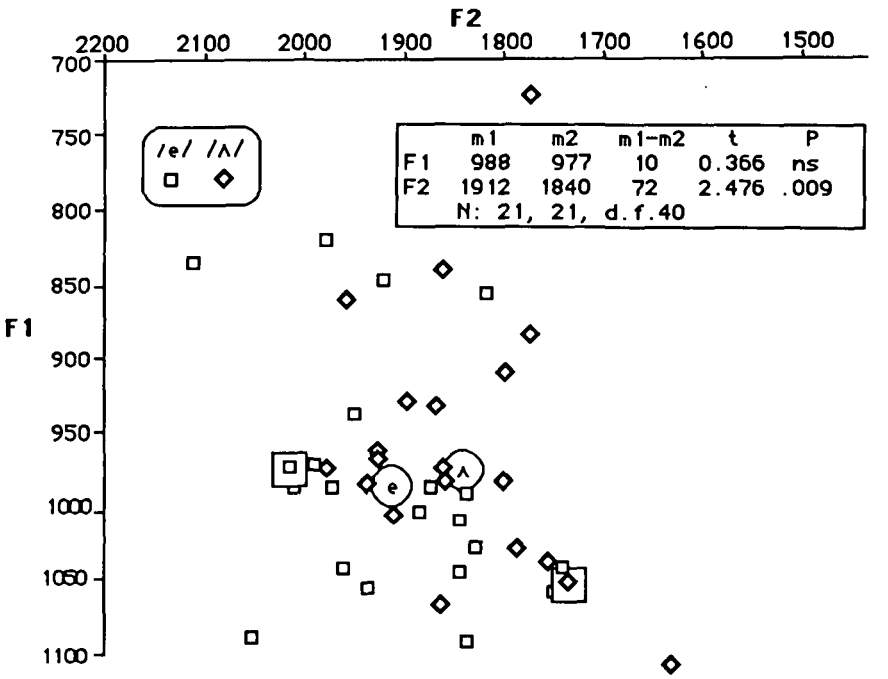


FIGURE 6. Distribution of *ferry* and *furry* tokens produced by Laura M. in the long commutation test. (Symbols framed by squares are tokens selected for commutation test stimuli in Experiment 2.)

black subjects were included, 11 from the Philadelphia area, and 3 from other areas. As in other studies, the black subjects proved to be members of a distinctly different speech community. None showed any tendency towards a merger of /er/ and /ʌr/. Five subjects were drawn from the Philadelphia suburbs. Since all but one had a clear distinction between /er/ and /ʌr/, it seems likely that they differ from the city population as a whole. This analysis will therefore focus upon the white subjects in Experiment 2.[20] We will therefore be examining the results from 20 white Philadelphians and 14 white subjects from outside of Philadelphia.[21]

Table 10 shows results for the self-commutation test in Experiment 2. The responses are strikingly similar to Table 6 for Experiment 1. The performance of non-Philadelphians is perfect, while only one-fifth of the Philadelphians completed the test without error.

Table 11 shows the results of the Coach Test, which are again strikingly similar to the results of Experiment 1 (Table 8). Figure 8 compares the results of Experiments 1 and 2 for both the commutation test and the Coach Test. The replicability of the responses indicates that the two modes of behavior tap invariant structural components of the speech community.

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

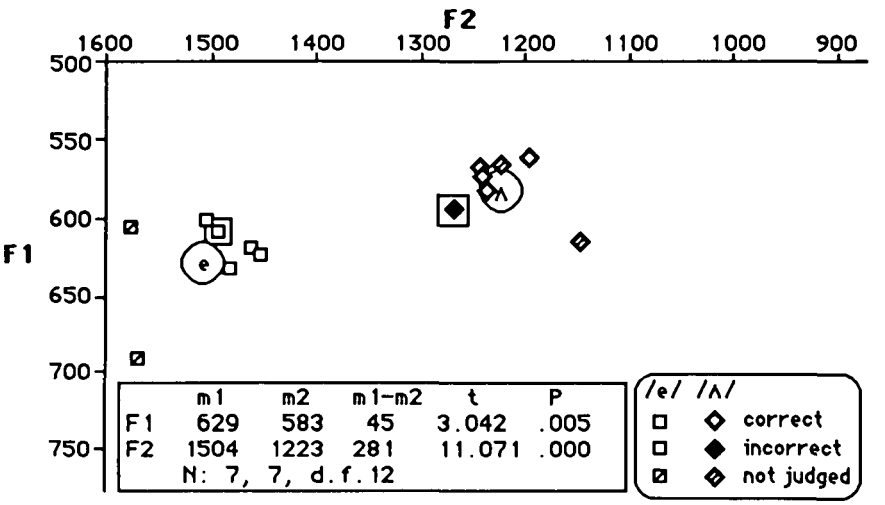


FIGURE 7. Distribution of *ferry* and *furry* tokens produced by Jed F. in the commutation test of Experiment 1. (Symbols framed by squares are tokens selected for commutation test stimuli in Experiment 2.)

TABLE 10. *Commutation test of* ferry/furry *in Experiment 2*

	Pass	Fail
Non-Philadelphians	14	0
Philadelphians	4	16

It seems clear from the evidence of the Coach Test that Philadelphians as a whole do not use the difference between /er/ and /ʌr/ to distinguish words. The results of the Coach Test are above chance, yet on the surface of it, performance on the Coach Test is worse than performance on the commutation test. If we assume that the Coach Test measures the same ability as the commutation test, then the commutation test is 10 repetitions of the Coach Test. But if we apply the rate for Experiment 2 ten times, we have $.737^{10}$ or .047. This would predict a success rate on the commutation test of less than 5%, instead of the 20% obtained (and 10% in Experiment 1).

This line of reasoning assumes that all Philadelphians use the distinction to the same extent. But evidence of the commutation test scores indicates instead that *some* Philadelphians are capable of using the difference between /er/ and /ʌr/ to differentiate words. Let us assume that the small number

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

TABLE 11. *Coach Test results in Experiment 2*

	Correct	Incorrect
Non-Philadelphians	13	1
Philadelphians	14	5

$\chi^2 = 4{,}00$; $p = .05$.

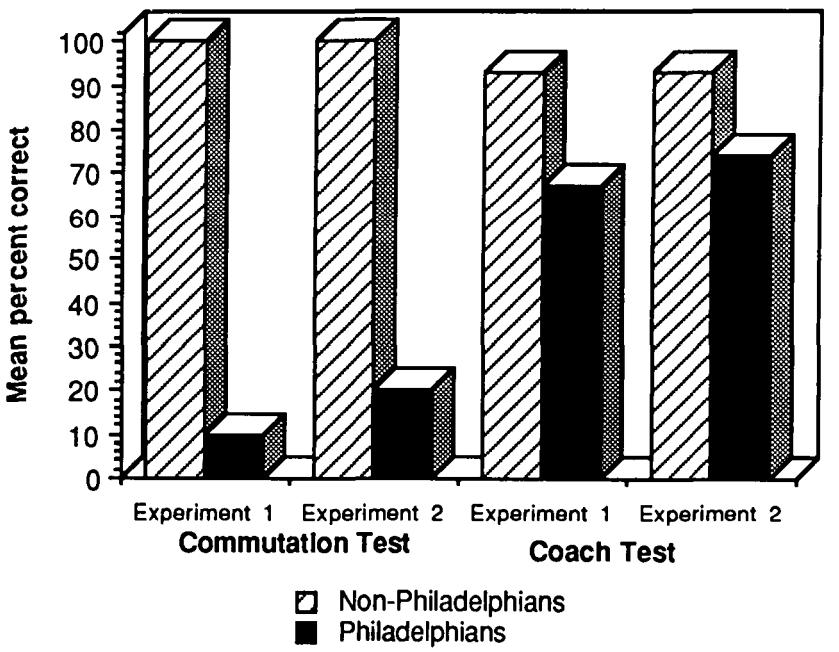


FIGURE 8. Comparison of results of commutation and Coach tests in Experiments 1 and 2.

of Philadelphia subjects who achieved 100% in the self-commutation test are also capable of matching the performance of non-Philadelphians on the Coach Test. This means that 4 of the 14 who gave correct answers on the Coach Test in Experiment 2 also passed the commutation test. This turns out to be true. If we now set those respondents aside, we can call the rest "normal Philadelphians." The proportion of normal Philadelphians who gave the correct answer on the Coach Test is not 14/19, but 10/15 or .67. (A chance performance would be .50 or 7.5/15.) If we assume that the chance of success on any individual commutation trial is the same as that for the Coach

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

TABLE 12. *Results of commutation tests for Self, LM, and JF*

	Self		Laura M.		Jed F.	
	Pass	Fail	Pass	Fail	Pass	Fail
Non-Philadelphians	14	0	2	7	7	2
Philadelphians	4	16	0	18	3	18

Test (.67), then the chance of passing the commutation test for these 15 subjects is $.67^{10}$ or .018. Assuming that the .67 response on the Coach Test is the average success rate for a normal Philadelphian, we would not expect any of the 15 to pass the commutation test, and this of course completes the argument. To put it in reverse, a success rate of 74% on the Coach Test is not inconsistent with the 20% success rate on the commutation test. We are looking at the same form of behavior.

Let us now extend our view of the commutation test to consider the responses to the tokens produced by the near-merger type, Laura M., and the Philadelphia distinction type, Jed F. Table 12 makes this comparison with the pass/fail criterion of 100% success versus less than 100%.

The Laura M. test proved to be a very difficult one. No Philadelphians passed it, not even the four who passed the self-commutation test. Yet two non-Philadelphians did get 100%. If subjects were guessing randomly on the Laura M. data, the proportion of success on the commutation test would be $.5^{10}$ or just about 1 out of 1000. This means that there is enough information in the Laura M. tokens for a sensitive person to detect the pattern consistently.

The case of Jed F. is quite different. For the non-Philadelphians, this seems to be more difficult than judging their own productions, though the difference is not significant. It is not surprising to find that Philadelphians are significantly worse than non-Philadelphians. But it is surprising to find that Philadelphians have no more success in judging Jed F.'s tokens than their own. When we consider that over half the Philadelphia subjects have mergers or near-mergers, this means that they do not benefit from the clear (though moderate) distinction that Jed F. makes in production. The situation is somewhat clearer when we add the four suburban subjects to the city subjects. Of seven subjects who have a clear distinction with no tokens of /er/ and /ʌr/ less than 200 Hz apart, four achieved 100% success on the self-commutation test. But of these, only one got 100% in judging Jed F. It is evident that Philadelphians are not able to use the minimal separation of 200 Hz to consistently distinguish /er/ and /ʌr/.

The striking difference in the difficulty of categorizing the Laura M. and Jed F. series is puzzling if we consider only the F1 and F2 differences in-

TABLE 13. *Categorization of* ferry/furry *in Experiment 2 by type of production in self categorization test (% correct)*

	Self	Jed F.	Laura M.
Non-Philadelphians			
clear distinction	100 (14)**	96.7 (9)**	78.9 (9)
Philadelphians			
clear distinction	91.7 (9)*	82.5 (8)	73.8 (8)
nonoverlapping with close approximation	74 (5)	76 (5)	66 (5)
overlapping with significant difference in F2 means	68 (5)	78 (4)	56 (4)
merged	40 (4)	77 (3)	56 (3)
All clear distinction			76.5 (17)***
All merger and near-merger			57.5 (12)

Note: Numbers of subjects shown in parentheses.
*Significant difference at $p \leq .05$, with group immediately below.
**Significant difference at $p < .01$, with group immediately below.
***Significant difference at $p < .001$, with group immediately below.

volved. There was no difference in the crucial F2 dimension: the two Laura M. tokens differed by 228 Hz and Jed F. tokens by 221 Hz. As far as F1 is concerned, the Laura M. *furry* was 81 Hz higher, and Jed F. only 13 Hz lower. However, if we examine the two distributions in Figures 5 and 6, it is evident that there is a great difference in the pattern. The Jed F. tokens are clustered into two sets which appear to cross a psychologically real boundary. The Laura M. tokens form a continuous distribution, at a considerably higher F1 and F2 range. The higher F2 range for Laura M. is not merely a product of the female vocal tract: all of her tokens sound front to most listeners. In spite of the bias towards giving equal number of answers for two alternatives, only six subjects did so. Twenty-seven heard more *ferry* than *furry* (with five subjects hearing only *ferry*), and only seven hearing more *furry*. The mean responses for 40 subjects was 6.3 *ferry* and 3.7 *furry*. The major difficulty with the Laura M. tokens is that the wedge tokens had moved well into the /e/ range.

So far, we have been considering only the strict criterion of 100% success in Experiment 2. When we turn now to consider the actual rate of correct answers in the commutation test, we obtain a more sensitive indicator of subjects' ability to use the /er/ ~ /ʌr/ distinction. Table 13 correlates the percent correct in the categorization of *ferry* and *furry* in the three commutation tests with the type of distinction made by subjects in their self-commutation test. The first line shows the non-Philadelphians, who all make a clear distinction. The second line compares these with the nine Philadelphians who also make a clear distinction (including the four suburban subjects). Despite the high performance of the Philadelphians, their mean scores are significantly lower than non-Philadelphians ($p = .0147$). Results for Jed F. are very slightly

https://doi.org/10.1017/S0954394500000442 Published online by Cambridge University Press

lower for both groups: again Philadelphians rate significantly worse (p = .008). Results for Laura M. drop off sharply for both groups, and here there are no significant differences between them. This indicates that on the easier tasks, Philadelphians who make the distinction are impaired in their ability to categorize vowels, but on the more difficult tasks, their categorization (and discrimination) is no worse than others.

Turning to the differences among Philadelphians, there is a significant effect on the self-categorization test between those who make a clear distinction and those with close approximation (p = .056). Though the other groups are graded in the way that we would expect, the differences are not significant. In the case of Jed F., all Philadelphians are basically alike: the small difference between those with clear distinctions and others is not significant. When we come to Laura M., the situation is the converse. There are no significant differences among Philadelphians, but if we group the Philadelphians who show clear distinctions with the non-Philadelphians, the differences between them and all others are significant (p = .009).

We can sum up the results of this fine-grained examination of the commutation test with four findings that hold for the results of Experiments 1 and 2 generally.

Philadelphians with a full merger show the expected random response to self-commutation tests, and a severe reduction in the ability to categorize tokens that are clearly distinct.

Philadelphians with a near-merger, either overlapping or nonoverlapping, do not show a significant improvement over speakers with a merger in their categorizations.

Philadelphians with a clear distinction are significantly better than others in categorizing their own productions but are not better than other Philadelphians in categorizing a clear Philadelphia distinction. They are better than others in categorizing a near-merger.

All Philadelphians are worse than non-Philadelphians in judging their own productions or a standard clear distinction.

These results indicate that for Philadelphians in general, the semantic utility of the distinction is considerably less for non-Philadelphians. Even the Philadelphians who would be expected to do better on the Coach and commutation tests, those who claim a distinction and keep their vowels distinct in production, scored significantly worse on these tests than non-Philadelphians.

Thus the near-merger of /er/ and /ʌr/ is accompanied by a reduction of the semantic contrast of this opposition. What is most remarkable about this development is that the entire community is affected in the decreased semantic utilization of the distinction. This gives partial confirmation to the proposals of Janson and Schulman (1983) and Herold (1990) that speakers who make a distinction—but are in close contact with those who do not—find that the distinction is not a useful one and stop using it to differentiate words. The suspension of phonemic contrast in Philadelphia is not complete

but impaired to a degree that is correlated with the system of the speaker. We are therefore looking at a dynamic process in Philadelphia, where the mechanism of merger is exposed for our inspection. The real-time status of the process is not yet evident: we do not know how fast the merger is proceeding, if at all. It is abundantly clear, however, that the phenomenon of near-merger is solidly entrenched in the process. While our previous reports of near-merger concerned a few individuals in each community, the near-merger affects a solid plurality of Philadelphia speakers and appears to be an integral part of the linguistic economy.

The question then remains as to how Philadelphians can maintain non-overlapping distributions, yet show such severe limitations in their ability to categorize their output. This is a question about discrimination rather than categorization. An experimental series now being conducted as an extension of Experiment 2 is designed to provide answers to this question. Some indications of the direction of the answer can be foreseen in Table 13. The Philadelphians who make a clear distinction are the subjects who show most clearly the influence of the community-wide norm that *ferry* and *furry* can be, or are, "the same." The gap between their performance and that of non-Philadelphians is maximal in judging the tokens of Jed F., who makes a clear distinction, but the gap is no longer significant in the more difficult task of judging the tokens of Laura M. The influence of the Philadelphia norm is strongest when there are clear linguistic judgments to be made; as the task approaches the psycho-acoustic level of discrimination within categories, the difference between Philadelphians and others tends to disappear.

NOTES

1. In theories of underspecification, some distinctive features may not be present in underlying forms, but are derived by implicational relations.
2. The merger is actually assumed rather than reported in Hubbell (1962:83), Wetmore (1959), and Kurath and McDavid (1961).
3. The one aberrant token of *fool* in Figure 1c was consistently judged as *full* by all listeners.
4. Di Paolo and Faber (1990) is an investigation of phonation differences in the /uwl/ and /u/ tokens of the Salt Lake City subjects, which may help to explain some of the anomalies in the F1/F2 plots, such as the fact that the isolated token of *fool* on the left of the *full* distribution in Figure 1c is regularly heard as *fool*. Their Figures 9–18 show the F1/F2 positions of the vowel nuclei before /l/ for teenagers. They show a remarkable similarity to Dan Jones' system in Figure 1c, with a close approximation of the /uwl/ and /ul/ classes.
5. This is a pattern that is also found in the studies of the laxing of /uwl/ in Texas (Bailey & Ross, forthcoming).
6. A number of other sources indicate that the merger of the **meat** and **mate** classes took place by 1700 and is maintained to the present day for those who use the traditional pronunciation. However, the number of words in the **meat** class has been steadily shrinking, drained off into the ē or **meet** class. Of 100 words noted by Patterson in the **meat** class in the 19th century, only 35 survive in the modern Belfast class.
7. The series does not include the area of English [æ], because it follows a more centralized path in the lower mid region. When Labov listened to it, he was not sure that he would identify any word as prototypical *sat*, though some were closer to *sat* than any other word. This problem may not be at all relevant to Stockholm speakers, whose knowledge of English phonetics, no matter how excellent, is mediated by their Stockholm system.

8. They arrived at the same inference about the suspension of semantic contrast that Herold (1990) inferred from her study of the merger of *cot* and *caught* in eastern Pennsylvania.

9. The work reported here was carried out by Mark Karen and Corey Miller of the research group on Cross-Dialectal Comprehension at the University of Pennsylvania, supported by the National Science Foundation.

10. The distinction between labeling and perception was first called to our attention by Leigh Lisker, in reviewing the problem of this article. Herold (1990) made the further distinction between labeling and correct identification. It is quite possible for a subject to develop a consistent and reliable labeling strategy that does not match the actual use of the distinction by those who maintain it. This was the case in Herold's investigation of the ability of Toronto speakers to deal with the New York City /a/~/oh/ distinction. A number of them did show a consistent, 100% labeling function, but almost one-quarter of these reversed the labels on the words. What the speaker intended as *cot* was heard as *caught*, and vice versa. Thus, we can identify three levels of behavior: perception, labeling, and correct identification.

11. In this respect, the experiments are quite similar to the "moral dilemmas" posed by Kohlberg (1981) in the study of moral judgments.

12. We are much indebted to Dave De Pue, not only for the realistic and convincing performance, but for many additions of colloquial forms, some specific to Philadelphia, which made the Coach Test more interesting and convincing.

13. The GW 12-bit digitizer was used, and narrow band spectrograms were prepared by the MacSpeech Lab I program on a Macintosh SE. The central tendency of the formants were identified using the digital procedure outlined in Labov, Yaeger, and Steiner (1972: Ch. 2), which yields measurements with a maximal error range of one-quarter of a pitch period.

14. Always assuming, of course, that the F1 and F2 of the nucleus as plotted in Figure 5 is an indicator of the central tendency of the vowel trajectory as perceived. Of course, there are other features that differentiate the words. We have not, as yet, detected any other acoustic feature that systematically differentiates /er/ and /ʌr/, but such features may exist.

15. This one subject is of particular interest, since the two sets of vowels are opposed diagonally: /ʌ/ tokens have either lower F2 or lower F1. One token of *ferry* is within 50 Hz of the *furry* concentration on the F2 axis but is much lower (over a 100 Hz away) on the F1 axis. This implies that at least for this subject, F2 is a more important dimension than F1.

16. "Correct" here means correct on the first response. We did not consider that a tendency to persist in the same interpretation when the tokens were switched in the second part of the test gave unambiguous information on the subject's linguistic system, since there are nonlinguistic factors that could lead a person to ignore a change in the input data. The second part of the test did show non-Philadelphians' greater sensitivity to the distinction through their ability to change judgments with change of input data, but the effect is a smaller one.

17. Further applications of the Coach Test with groups produced a larger sample, and significant differences between Philadelphians and non-Philadelphians. It should be remembered that they were judging tokens of *Murray in* and *Merion* where the difference between /er/ and /ʌr/ was exaggerated by the Philadelphia speaker in the direction of the non-Philadelphians' production. The F2 difference was 250 Hz.

18. Experiment 2 was designed primarily to compare subjects' psycho-acoustic abilities to discriminate sounds as compared with their linguistic ability to categorize them.

19. These were members of the Guardian Angels, a volunteer anti-crime group.

20. Black Philadelphians showed higher performance on commutation tests of their own productions than whites but did worse in categorization of the standard white subjects and worse on discrimination tests based on stimuli drawn from the production of white speakers.

21. 6 from the Northern Cities, 2 from the Middle Atlantic States, 4 from the South Midland, 1 from east Texas, and 1 from the far west.

REFERENCES

Bailey, G., & Ross, G. (forthcoming). The evolution of a vernacular. To appear in O. Ihailainen (ed.), *Proceedings of the International Conference on the History of Linguistics 6*. London: Mouton.

Bloomfield, Leonard. (1926). A set of postulates for the science of language. *Language* 2:153–164.

Chomsky, Noam. (1964). The logical basis of linguistic theory. In Horace G. Lunt (ed.), *Proceedings of the Ninth International Congress of Linguists.* London: Mouton. 914–1008.
Di Paolo, Marianna. (1988). Pronunciation and categorization in sound change. In K. Ferrara et al. (eds.), *Linguistic change and contact: NWAV XVI.* Austin: University of Texas, Department of Linguistics. 84–92.
Di Paolo, Marianna, & Faber, Alice. (1990). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change* 2:155–204.
Fries, Charles C., & Pike, Kenneth. (1949). Co-existent phonemic systems. *Language* 25:29–50.
Garde, Paul. (1961). Réflexions sur les différences phonétiques entre les langues slaves. *Word* 17:34–62.
Halle, Morris. (1962). Phonology in generative grammar. *Word* 18:54–72.
Harnad, Stevan (ed.). (1987). *Categorical perception.* Cambridge: Cambridge University Press.
Harris, John. (1985). *Phonological variation and change: Studies in Hiberno-Irish.* Cambridge: Cambridge University Press.
Harris, Zellig. (1951). *Methods in structural linguistics.* Chicago: University of Chicago Press.
Herold, Ruth. (1990). *Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania.* Ph.D. dissertation, University of Pennsylvania.
Hindle, Donald. (1978). Approaches to vowel normalization in the study of natural speech. In D. Sankoff (ed.), *Linguistic variation: Models and methods.* New York: Academic. 161–172.
Hubbell, Allan F. (1962). *The pronunciation of English in New York City. Consonants and vowels.* New York: Columbia University Press.
Janson, Tore, & Schulman, Richard. (1983). Non-distinctive features and their use. *Journal of Linguistics* 19:321–336.
Jespersen, Otto. (1949). *A Modern English grammar on historical principles. Part I: Sounds and spellings.* London: George Allen & Unwin.
Kohlberg, Lawrence. (1981). *Essays on moral development.* San Francisco: Harper & Row.
Kurath, Hans, & McDavid, Raven I., Jr. (1961). *The pronunciation of English in the Atlantic States.* Ann Arbor: University of Michigan Press.
Labov, William. (1975a). On the use of the present to explain the past. In L. Heilmann (ed.), *Proceedings of the 11th International Congress of Linguists.* Bologna: Il Mulino. 825–851.
_____ (1975b). *What is a linguistic fact?* Lisse: Peter de Ridder and New York: Humanities.
_____ (1988). The judicial testing of linguistic theory. In D. Tannen (ed.), *Language in context: Connecting observation and understanding.* Norwood, NJ: Ablex. 159–182.
Labov, William, Yaeger, Malcah, & Steiner, Richard. (1972). *A quantitative study of sound change in progress.* Philadelphia: U.S. Regional Survey.
Liberman, Alvin, Harris, K., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phonemic boundaries. *Journal of Experimental Psychology* 54:358–368.
Milroy, James, & Harris, John. (1980). When is a merger not a merger? The MEAT/MATE problem in a present-day English vernacular. *English World-Wide* 1:199–210.
Moulton, William. (1962). *The sounds of English and German.* Chicago: University of Chicago Press.
Nunberg, Geoffrey. (1980). A falsely reported merger in eighteenth century English: A study of diachronic variation. In W. Labov (ed.), *Locating language in time and space.* New York: Academic. 221–250.
Orton, Harold, & Dieth, Eugen. (1962–67). *Survey of English dialects.* Leeds: E. J. Arnold & Son.
Patterson, W. H. (1880). *A glossary of words in use in the counties of Antrim and Down.* London: Trübner.
Peterson, Gordon E., & Barney, Harold L. (1952). Control methods used in a study of the vowels. *JASA* 24:175–184.
Pisoni, David. (1971). *On the nature of categorical perception of speech sounds.* (Status Report on Speech Research SR-27). New Haven: Haskins Laboratories.
Trudgill, Peter. (1974). *The social differentiation of English in Norwich.* Cambridge: Cambridge University Press.
Voegelin, C. F., & Harris, Zellig, S. (1951). Methods for determining intelligibility among dialects of natural languages. *Proceedings of the American Philosophical Society* 95:322–329.
Wetmore, Thomas. (1959). *The low-central and low-back vowels in the English of the Eastern United States* (Publication of the American Dialect Society 32). University: University of Alabama Press.