1. {Forthcoming, in John Baugh (ed). *Linguistic Profiling in Interdisciplinary Perspective: Volume 2, Social Science and Legal Dimensions*; Final Report submitted to the Ford Foundation. Washington University in St. Louis: Program in Linguistics and African and African American Studies. Accepted 11/2009)}

Title: Statistical Profiles of Race of Interviewer Perceptibility in National Surveys

David C. Wilson, University of Delaware

Abstract

Making use of linguistic profiling and race of interviewer research, this study examines the extent to which respondents in national surveys can accurately guess their interviewer's race when asked by the interviewer. The concept of interest in this study is perceptibility: whether or not respondents guess accurately. In the analyses, I statistically control for interviewer and respondent race, and experimentally control for effects related to the survey setting (e.g., topic and length of conversation). The results indicate that race of interviewer perceptibility significantly varies across the interviewer's race, whereas black (and racial minority interviewers) are more likely to be guessed incorrectly. Features of the survey setting had no significant impact, and most noteworthy, perceptibility was no better at the end of the survey than at the beginning. The findings highlight the need to continue to examine racial perceptibility, as well as, incorporate methodological improvements in the study of linguistic profiling and race of interviewer effects.

Introduction

This research is about how well people can gauge race over the phone, and what factors might be related to this perceptibility. Is accurate racial identification widespread in telephone based social surveys? And are there systematic factors that might influence racial perceptibility.

In this chapter I argue that racial perceptibility over the phone should not be assumed. Differences in settings, interviewer related factors, and individual background may account for any differences in perceptibility. Yet, there are very few studies examining perceptibility rates or what may be influencing them. Thus, we are in serious need for accurate statistical profiles of perceptibility.

Linguistic Profiling and Race of Interviewer Research

There are two separate fields of research that deal with racial perceptibility over the phone.

Linguistic profiling (LP) is a research area that focusing on how differential treatment can occur through language and other verbal factors. The typical methodology for examining LP is to vary the race of a caller looking for, say, residential property, and then examine whether or not property owners or their representatives respond differently across caller race. The data from these studies provide ample evidence that racial minorities are likely to be discriminated against, even when there is no face-to-face contact.

There is a similar scenario in social survey research. This area of research is centered on what are known as race of interviewer (ROI) effects. Studies of ROI effects examine the extent to which respondents edit their answers to survey questions depending on the interviewer's race. Originating in face-to-face survey settings, ROI effects have also been found in telephone surveys.

The ROI and LP problems share many features. For instance, in both ROI and LP studies, callers (e.g., interviewers or potential residents) are the likely "victims" and call recipients (e.g., potential respondents, or property owners) are the likely perpetrators of the targeting behavior (i.e., response editing or discrimination). In addition, in both ROI and LP research areas there is a necessary social exchange -- valued resources that one party supplies and the other party demands -- that must take place. In surveys, interviewers (and survey organizations) desire valid and reliable responses supplied by respondents, and in residential searches consumers desire housing opportunities supplied by home owners and property management organizations. The most important feature that both ROI and LP studies share is the assumption that respondents can quickly and consistently identify race over the phone, particularly when one end of the phone line contains a racial minority (Davis, 1997a, Davis, 1997b, Purnell, Isardi, and Baugh, 1999; Thomas and Reaser, 2004).

Most research suggests that race is easily perceived in telephone conversations (Purnell, Isardi, and Baugh, 1999). Linguistic research shows that even in the absence of diagnostic morphosyntactic and lexical variables, or other prototypical African American vocal characteristics, racial identification, especially black identification, is viable (Thomas and Reaser, 2004). This is because listeners develop short-cuts through experiences and social settings that may cue race. However, many studies indicating strong racial perceptibility in typical telephone conversation settings incorporate obvious racial vernaculars and other vocal characteristics that make race of interviewer more obvious (see Purnell, Isardi, and Baugh, 1999).

In social surveys there are few published reports of perceptibility, and those that exist originate from studies examining either minority respondents or interviewers (Conover, Hatchett, and Jackson, 1985; Davis, 1997; Wolford et al., 1995), or they provide no breakdown across respondent or interviewer race. It is likely that either respondent or interviewer race (or their interaction) has a moderating effect on perceptibility (e.g., blacks are more perceptible than whites).[1]

Another area that has received no attention in the examination of ROI perceptibility is the effect of the surveys context on perceptibility. There are two components of the survey setting that may affect ROI perceptibility. First, the length of the conversation (i.e., survey) may help respondents to identify interviewer race. Longer conversations allow respondents to pay closer attention to the interviewer's vocal cues, whereas shorter surveys may provide less information about the interviewer's race. Second, the survey's topic may induce respondents to consider the ROI. Surveys asking questions about race may either open or constrain response depending on who people believe they are talking to over the phone. Alternatively, surveys asking about generally benign questions (e.g., life satisfaction) may not provoke any considerations of race. Thus, both the length and the topic of survey conversations may influence ROI perceptibility.

The points in the above mentioned paragraphs point to a need to examine the extent of racial perceptibility in national survey settings controlling for respondent and interviewer race, as well as, examine whether there are any factors related to the social conversational settings, such as length of conversation and topic of conversation which may impact ROI perceptibility.

Background Research

Perceptions of Interviewer Race

In early telephone related ROI effects correct ROI perceptibility was assumed. Cotter, Cohen, and Coulter (1981) were among the first to publish reports of ROI effects in telephone interviews. They assumed that since statistically significant differences existed across interviewer race, then respondents must have been able to discern race either through name association or verbal cues.

[1] For instance, as Alim (2003) has noted, black Americans are deft at style shifting: changing their vocal characteristics to accommodate to interlocutors.

Following up on the research by Cotter and colleagues, Gurin, Hatchett, and Jackson (1985) examined the extent to which respondents actually knew their interviewer's race. Using their National Black Election Study (NBES) data they asked respondents the following question at the end of the survey:

> "We'd like to find out what kinds of things people can tell just from listening to a person's voice over the telephone. During the interview did you think I was white, black, or someone of another race (Gurin, Hatchett, and Jackson, 1985, p. 276)?"

In their study, they did not report the error rate, however, an examination of the data indicates that interviewer race was guessed correctly 79% of the time (21% error rate). The Gurin, Hatchett, and Jackson (1985) analysis highlighted the importance of studying both perceived (PROI) and actual interviewer race (AROI). More research on perceptibility followed. Wolford et al., (1995) conducted similar research and found a 73% perceptibility rate (27% error rate), and in one of the few studies examining black and white interviewers, as well as black and white respondents, Wilson and Olesen (2002) found a perceptibility rate of 82% (18% error rate). The error rates from the aforementioned, and additional, studies are listed in Table 1.

On average, respondents can accurately guess their interviewer's race about 75% of the time. This represents an overwhelming majority of respondents, but is far from ideal. The question arises, what are the factors that are influencing perceptibility? One such factor is likely interviewer race.

[Table 1 about Here]

Differences in Perceptibility across Interviewer Race

One feature of literature is the consistent pattern of black interviewers being guess incorrectly more often than white interviewers. The middle column of Table 1 shows the error rates for black interviewers. In each case where there is comparable data on white and black interviewer error rates, black interviewers are guessed incorrectly more often. This is somewhat surprising given the linguistic research indicating that African-Americans are easily identified over the phone (see Thomas and Reaser, 2004) and thus likely to receive differential treatment (see Purnell, Isardi, and Baugh, 1999; Davis, 1997a). This examination of error rates in the literature and data imply that black interviewer are not as easily identified as white interviewer, at least in professional settings such as social survey interviews.

It's likely that although black Americans have distinct verbal tendencies, they also incorporate various behaviors, including "style shifting," and speech shifting, that allow for ambiguity and protection of self (Majors and Mancini-Billson, 1993; Thomas and Reaser, 2004). What's more, sociolinguistic researchers have proposed that black American's use of standard English makes identification even more difficult because it's

very close to the forms of English spoken by other races (Thomas and Reaser, 2004). Since black survey interviewers are more likely to be educated and adopt norms of professionalism, they may also be more likely to be guessed incorrectly because their vocal characteristics are less similar to stereotypical African American English vernacular (AAEV). Black interviewers are also likely to find ways to avoid having to meet and deal with potentially hostile racial treatment (see Lind et al, 2001). One way they might do this is to maintain a more standard and neutral tone throughout the survey interview, making it harder for respondents to identify them by race. This might be particularly true in studies that ask respondents to guess the interviewer's race. To the extent that black interviewers can control their vocal characteristics, it is an empowering skill to be able to deflect potential racial hostility.

Testing this belief in a post-hoc fashion, I gathered six studies of the U.S. adult populations. Each study consisted of random digit dial (RDD) telephone interviews conducted by the Gallup Organization. In each of the studies the topics dealt with issues of race, including perceptions of race relations and racial discrimination in the workplace. The topical nature of the survey is important because it should heighten considerations of race in the social conversation between the respondent and interviewer, thus inducing greater chances of perceptibility.

Table 2 provides detail of the analysis. The studies are listed out in rows at the far left side of the table. The sample sizes for each study, the error rates, and the percentage of interviewers by race are located in the columns. The bottom (last) row of the table presents the overall figures from the study.

[Table 2 about Here]

The data presented in Table 2 are extremely robust. Each of the studies was conducted independently of the other. Five of the six studies contain over 2,000 respondents, and the sixth contains just over 1,200. More over, each of the studies has a over-sample of black respondents, meaning that the study purposefully gathered more black, and in some cases Hispanic, respondent, than is proportionally represented in the population. Finally, the large majority of the studies contain a relatively large number of interviews conduct by minority interviewers, with the sole exception being the 2005 Employee Discrimination in the Workplace Study. This can be seen in the second to last column ("AROI"). For instance, in the 1998 Black-White Study, white, black, and other interviewers conducted 60%, 33%, and 7% of interviews respectively. These data represent the first multi-study analysis of ROI perceptibility, and set the stage for an experimental examination of ROI perceptibility.

The analyses confirm that black, and other minority, interviewers are more likely to be guess incorrectly. The overall error rate -- percentage of incorrect guesses -- for all six studies is 33%, and as expected the error rates vary across race. While the error rate for whites is 24%, the error rates for black and "other" [non-black and non-white] interviewers are respectively 43% and 73%. The same results exist for each of the

individual studies, racial minority interviewers are more likely than white interviewers to be guessed incorrectly.

The data from this preliminary analysis shows that there is much more to examine about the racial perceptibility in telephone interviews than what is currently assumed. Most studies of ROI effects assume that respondents know the interviewers race. Additionally, both the "name recognition" and "linguistic" phonetics explanations assume that respondents are attentive to the extent that they recall their interviewer's name, and can understand differences in vernaculars. However, the data show that respondents are not always certain about race over the phone, and thus it is likely that there are other factors in the interview that may lead to accurate perceptibility or conversely perceptibility problems. These factors can include the topical content of the survey and the length of the survey interviewer. Some respondents may be more perceptive of the nuances of auditory cues than others and thus have greater confidence, which may also influence perceptibility. Finally, all three -- survey topic, interview length, and confidence in perceptibility -- may work in conjunction with the interviewer's voice to influence perceptibility.

In the next section I briefly expound upon the survey related factors I consider important. While there are a number of additional aspects of the survey interview that may influence perceptibility such as region surveyed or how often someone is interviewed in surveys, the ones discussed below are my main interest.

Survey Design Related Factors

Stage of Interview. The point at which respondents are most aware of the interviewer's race is the point where ROI effects may be most likely to occur. But, there is no survey research regarding where perceptibility is most accurate. While respondents may have less contact or experience with people of different social groups or regions and thus are less likely to recognize different dialects, over the course of an interview the probability of correct recognition should improve. For instance, as interviewers engage in dialog, respondents may recall experiences with audio-visual media that may help to associate particular speech patterns with particular racial-ethnic groups. Therefore, the chances of respondents guessing correctly should be highest at the end, since they've had adequate time to listen to the interviewer's voice and thus make an educated guess about their interviewer's race.

Question Content. Another factor that may induce racial considerations is the type of question being asked, or what Orne (1969) has called the "demand characteristics of the situation." Accordingly, even questions that are not necessarily racial, but may have a racial association with public events or beliefs can bring about a sensitive to the interviewer's race (Davis and Silver, 2003; Schaeffer, 1980).

Survey topical content has been suggested as an explanation for perceptibility, but never examined. Finkel, Gutterbock, and Borg (1991, p. 317-18) suggest that interviewer's may use "verbal cues, intonations, or inflections that elicit certain responses to question

of particular important or salience to the interviewer." This implies that interviewer verbal cues may interact with question context to influence perceptibility, with certain questions increasing the probability of accurate perceptibility.

Confidence. A third factor that must be examined in any analysis of perceptibility is confidence. Respondents who are (i.e., report) more confidence in their perceptibility of interviewer race should be those individuals who are more certain that they have the experiences and social expertise to accurately judge auditory racial differences. For example, people who are older, more educated, and have greater interracial contact should have higher confidence. In addition, self-reported confidence may be associated with survey interview components such as certain stages of the interview or question content.

To my knowledge, there are no published studies examining both respondent accuracy and confidence, and their interaction with ROI and respondent. Presumably, accuracy and confidence should be highest at the end of the survey when respondents have had an opportunity to listen to the entire survey. Also, accuracy and confidence should be higher after being primed by racial questions, since the race should be made salient by topic. Finally, respondents should have lower perceptibility and confidence with non-white interviewers regardless of stage and content.

Interviewer Race. Finally, as suggested by Wilson and Olesen (2002) and the analysis presented above (see Table 2), it is likely that accuracy in guessing is mainly due to the interviewer's race. Perhaps white interviewers give off verbal cues that are easier to identify than non-white interviewers. Or, maybe black interviewer's are conscious of the possibility of racial hostility over the phone and adjust their communication styles -- don their "masks" -- to appear more accommodating and less discomforting to interviewees (Davis, 1997a). Davis (1997a) describes this process as one shaped by a context of psychological discomfort. Given the negative perceptions or stereotypes that whites (or blacks) may have of blacks, blacks will "adjust their behavior to the host of expectations, rewards, and penalties that surround them (p. 311)." While Davis is talking about respondents adjusting to social interview contexts, there is no reason to believe that interviewers would not also make a similar adjustment. In either case, the result is that accurate perceptibility would be lower for blacks than for whites. And the masking behavior might be present if results show equal or less confidence associated with black interviewers, but higher error rates.

The Experimental Design:

This research is driven by a national split-ballot survey experiment of the U.S. population. For each respondent, the interviewer's race was randomly introduced into the interview at one of three points: the beginning, the middle, or the end of the questionnaire. By inserting the interviewer's race into the survey conversation, I hoped to measure the differences in response for those who are explicitly aware that race is a

part of the conversation, versus those who are not made aware.[2] The introduction of the interviewer's race takes the following form.

> First ask: *"Up to this point, you may not have thought about this, but I'd like to ask you to guess my race. Would you guess that I am White, Black, or some other race?"*
>
> Next ask: *"On a scale from one to five, with one being "completely unsure" and five being the "completely confident", how confident are you in your guess?"*
>
> Finally state: *"Thank you for considering the question, it was only for research and training purposes, and just so you know, I am (a/an) (INTERVIEWER RACE) (INTERVIEWER GENDER)."*

This treatment allows for an analysis of not only the ability for respondents to guess correctly or incorrectly, but also how confident they are in their guess. The final statement in the treatment makes the interviewer's race explicit to the respondent. By systematically varying the point of insertion I can also gauge the ability of respondents to guess correctly depending on the point of the interview.

To control for instrument reactivity (i.e., question content), respondents were asked either non-racial questions (General/Political-GP), or race-related questions (Racial-R) first. Depending on the assigned questionnaqire, the GP module consisted of seven to thirteen items covering general presidential approval and life satisfaction, political and attitudes, cultural values, and social accommodation and politeness. Alternatively, the RR module had twenty-one to twenty-two items focusing on perceptions of race relations, evaluations of racial groups, racial attitudes toward candidates, and percepts of fairness.

Given the two modules of items and the three points at which the interviewer's race is inserted, the experiment has a 2 x 3 -- question content ordering by stage of the interview -- design.

The design plan was to interview 1,200 respondents, with 200 respondents --100 black and 100 white-- for each cell in the 2 x 3 design. A stratified sample design was based on race. Gallup would select only blacks and whites based on pre-collected demographic information, and then randomly assign one of the questionnaires once respondents agreed to take part in the interview.

The Data

Respondents for the study come from the Gallup Poll Panel (GPP), which as of April 2005, included over 26,000 recruited members representing broad and diverse segments of the U.S. population. GPP members are recruited by Random Digit Dialing (RDD), and are asked to complete 2-3 surveys per month, although this rarely happens. The GPP

[2] The question of how knowing the interviewer's race impacts response distributions is beyond the scope of this paper, but it is an important one that will be addressed in future research.

has been shown to provide similar results to national RDD studies.[3] The GPP provides a unique opportunity to study interviewer-respondent interactions and survey response primarily because of the ability to get roughly equal strata within race and other demographics.

The survey methodology consisted of a RDD telephone survey of GPP members. The call design consisted of five call attempts. The average interview time was fifteen minutes, and the response, completion, and incidence rates were respectively 56%, 99%, and 96%. The margin of error for the results is ± 3%.

The experimental design and final cell counts can be seen in Table 3. The Appendix section contains more extensive descriptions of the respondents and interviewers.

[Table 3 about here]

Results: Perceptibility and Confidence

Interviewer Race. The first analysis examines to what extent respondents inaccurately guess the interviewer's race, as well as their reported levels of confidence in their guesses. The results are calculated for each interviewer racial grouping -- black, white, and "other" race -- and respondent racial and gender categories. Table 4 shows the results.

First, as can be seen in the first row of the table, black interviewers are the only group where they are guessed incorrectly a majority of the time.

[Table 4 about here]

Black interviewers are significantly more likely to be guessed incorrectly than white interviewers (difference (d) =66%, $p<.01$).[4] In 84% of the interviews completely by black interviewers, respondents guessed they were either white or another race. Compare this to an error rate of 18% (interviews=779) for white interviewers. While error rates were different across race, respondents reported no differences in their confidence that they were making a certain guess. The mean confidence levels for black and white interviewers were both 3.30 (d=0, F=.38, n.s.).

Further analysis (not shown) also confirmed perceptibility differences across the interviewer race and gender. Black male (error rate=85%) and female (56%) interviewers had higher error rates than white male (26%) and female (15%) interviewers. Due to the small number of black female interviews, only the differences

[3] Out of 44 comparisons of GPP surveys to Gallup Poll national RDD samples between 2004 and 2005, none showed significant differences to Gallup presidential approval questions.

[4] Other interviewers were also guessed incorrectly at a relatively high rate (43%); however, there was only one interviewer in the "other" category (a female). Unless stated otherwise, the remainder of the findings will primarily focus on black and white interviewers.

across white males and females was significant (d=11%, z=3.11, p<.05). However, there were no significant differences in confidence across interviewer race and sex (F=.89, n.s.).

Respondent Race. As for respondents, blacks (34%) and whites (29%) showed no differences (d=5%; z=1.66, n.s.) in ROI perceptibility, but did differ in their confidence levels. The average confidence level for blacks was 3.70, compared to 3.00 for whites (Mean difference=.70, t=-8.96, p<.05). Thus, although blacks had greater confidence, they fared no better than whites in their ROI perceptibility.

Respondent Background and Perceptibility. After controlling for respondent background characteristics -- race, sex, college education, age, and racial population context -- the only significant predictor of perceptibility error was interviewer race. The first three columns ("Error") of Table 5 show the results from a logistic regression analyses predicting incorrect perceptibility (i.e., error). Model 1 includes only the respondent demographic predictors, and tests the extent to which respondent background determines perceptibility. The results show that both racial context variables (% of blacks and % of whites) are the only significant predictors. The positive coefficients for both black and white population variables indicate that the more whites or blacks one lives around, the less likely one will guess interviewer race correctly. This is partially consistent with the findings of Wolford et al. (1995) who found significant effects for organization and social based racial interaction variables (e.g., urbanicity, occupation, and employment status). Both the Wolford et al. (1995) analysis and the Model 1 analysis presented here both show that background demographics account for a small amount of variance (R^2 = .03, or 3%).

[Table 5 about here]

However, this paper argues that interviewer race itself plays a strong role in perceptibility because interviewer may seek to avoid racial detection. In Model 2 (of the "Error" columns), the inclusion of interviewer race not only removes the significant effect of racial context, but also significantly improves model fit from .03 to .33 (33%). Moreover, in Model 3, when the interaction between respondent race and racial context is considered, there is still no population size effect, yet the effect of interviewer race is still significant. Together these results show that when controlling interviewer race personal background characteristics do not add significantly to perceptibility accuracy.

The last three columns ("Confidence") of Table 5 present the results from ordinary least squares regression (OLS) predicting respondent confidence in their guess. The three models tested use the same variables used in predicting perceptibility. However, since there are no previous studies examining confidence, this analysis is mostly exploratory in nature.

When only the personal demographics are examined, respondent race, sex, and age are significantly related to confidence. The column labeled Model 1 shows that whites reported less confidence than blacks, males reported more confidence than females, and

as age increased respondents reported more confidence. However, racial population size and college education were not statistically related to confidence. When the interviewer race was considered (Model 2 under "Confidence), the effects of race, sex, and age stayed the same, as did the explanatory fit of the model ($\Delta R^2=0$, n.s.). Thus, confidence levels were no different across interviewer race. In the third model (Model 3) when interracial interaction was examined, not only were those variables not statistically related to confidence, but the effect of race became non-significant. This result not only shows the existence of racial segregation among the sampled population, it also shows the small effect of segregation (e.g., blacks living among more blacks) on confidence in perceptibility.[5] Thus, the results show that blacks and whites are no different in terms of their perceptibility or confidence.

Experimental Differences in Perceptibility. Are there interview related features that might assist in the ability to perceive interviewer race? Based on my analysis of the data, the answer is a strong "no." Neither the stage in the interview in which respondents were asked to guess interviewer race (Z=.58, ns.), the types of questions (Z=.67, n.s.), nor the interaction between the two (Z=.48, n.s.), had a significant effect on respondent perceptibility. Table 6 provides the results from the comparison of percentages and mean levels of confidence. Counter to expectations, there are no differences in perceptibility or confidence across stage of the interviewer. After having listened to all questions, and having ample time to hear the interviewer's voice, one might expect that respondents are most perceptible of their interviewer's race at the end of surveys. Yet, respondents are no more accurate in their perceptibility at the end, than at the beginning, or middle stage of the interview. The exact same results are found with confidence. There are no significant differences in the mean level of reported confidence.

Confidence and Perceptibility. Thus far the data indicate that survey related factors such as interview length, question order, and their interaction have no significant effect on ROI perceptibility. The final step is a multivariate analysis of perceptibility regression perceptibility on respondent background characteristics, survey related factors, or interviewer variables and their interactions. For this analysis I conducted another logistic regression using perceptibility (error=1) as the dependent variable. Predictor variables were entered into the model at different stages to observe the relative contribution of each towards perceptibility errors.

The multivariate analysis reveals that errors in perceptibility are primarily driven interviewer race. Table 7 shows the unstandardized regression coefficients, standard errors, and fit results of each model. Model 1 only considers the effects of confidence in guess. The negative sign indicates that greater confidence is associated with more positive guesses. However, in Model 5 where an interaction term between confidence and interviewer race is included (AROI=Black interviewer), the interaction term (i.e., confidence with black interviewers) has a positive sign, while the main effect of

[5] T-tests examining racial % differences across respondent race show that blacks are more likely to live around blacks and whites are more likely to live around whites. The mean black population percentage was 39% for Blacks, and 6% for whites (F=500, p<.001), and the mean white population percentage was 48% for blacks, and 82% for whites (F=490, p<.001).

confidence (i.e., confidence with white and other interviewers) has a negative sign. Thus, confidence with black interviewer is associated with less accuracy, and confidence with white and other interviewers is associated with more accuracy. This is confirmed by examining the main effect of AROI. While increasing the explained variance in Model 2 by 32%, AROI is no longer significant in Model 5. Thus, the data show that errors in guessing happen more often for people who have more reported confidence when they talk with black interviewers.

Discussion and Conclusion

To what extent can respondents in national telephone surveys detect the racial background of their interviewer? First, there were no effects due to the survey setting. Interviewer race was no more obvious at the end of the survey than it was at the beginning. What's more, the topics being discussed prior to asking respondents to guess their interviewer's race also had no effect on perceptibility. Basically, either respondents seemingly knew at "hello" (Purnell, Idsardi, and Baugh, 1999), or they were taking one-sided guesses (i.e., satisficing). There is no way to parse who accidentally guessed correctly, or knew race for sure, but the results show that perceptibility is associated more with interviewer and respondent characteristics, and not necessarily the survey topic or survey length.

Approximately 70% of respondents can guess interviewer race accurately, yet only 45% of them reported high (i.e., 4 or 5 on a 5-point scale) confidence. There were no racial or gender differences in perceptibility, but blacks were more confident than whites, even when controlling for gender. These results show that while the public is fairly good at guessing interviewer race, in their minds they are not completely confident, or possibly they are less willing to report their confidence. The implication is that racial perceptibility over the phone should not be assumed, or more reflective, should not be assumed to be equal across race.

Perceptibility varies by interviewer race, with black interviewers being guess another race more often (84%) than white interviewers (18%). The implication for the ROI literature is that the interviewer effect for black interviewers might be confounded. It's evident that in professional settings black interviewers can and will present themselves in a manner that makes their race ambiguous to the point where respondents confidently think they are talking to someone of another race. This "masking" behavior (Davis, 1997a) or "strategic style" (Majors and Mancini-Billson, 1993) should be considered a deliberate action on the part of interviewers to protect their self-esteem and avoid racial conflict (see Lind et al., 2001).

The implication for the LP literature is that there may be methodological limits to consider when selecting callers for fair housing testing (Purnell, Idsardi, and Baugh, 1999). Selecting individuals with distinctly African-American Vernacular English (AAVE) or other vernaculars making race and ethnicity obvious is acceptable for testing whether discrimination occurs in more apparent situations, but may be less rigorous for answering questions regarding perceptibility. In the research presented here, the black

interviewers each had verbal characteristics that were more standard, than they were AAVE, and thus, race was evidently more ambiguous. In addition, there needs to be more data collected on whether or not the person being called can actually identify the caller's race, similar to what's been done in Gallup and other national studies. This would be an ambitious task, but would benefit the understanding of LP in general population settings where both caller and called are clear about racial makeup (i.e., similar to a face-to-face setting).

The results also indicate that ambiguous interviewer vocal characteristics might affect black respondent's perceptibility more than whites. When controlling for confidence whites are less likely than blacks to makes guessing errors (see Table 7). While previous results in Table 5 showed no significant race of respondent effect on perceptibility, it appears that reported confidence is the difference maker.[6] Using Model 5 in Table 7 to predict the probabilities of an incorrect guess, Table 8 shows that there are substantial black-white differences in the predicted probability of error when controlling for confidence (Univariate ANOVA, eta=.164, $p<.01$). At each confidence level, blacks are more likely than whites to guess incorrectly, especially when confidence is lower ("completely unsure"). Thus, although whites and black overall do not differ in their errors in guessing, they do show significant differences when controlling for confidence. As such, when considering perceptibility confidence levels interviewer ambiguity may occur more often with black interviewers, *and* black respondents.

An alternative explanation might for the relatively high error rates with black interviewers may be that respondents are likely to "satisfice" when it comes to guessing interviewer race (Krosnick and Alwin, 1987). Satisficing is a psychological theory of cognition and response. Respondents who satisfice avoid deep consideration and analysis of tasks such as difficult survey questions, but consider their options well enough to provide plausible responses. When it comes to telephone interviewers, it's likely that the general public believes that they are white, or at least that they are more likely to be white. Thus, they will tend to guess white more often across the board to avoid having to incorrect guesses, recall all components of a conversation that might cue interviewer race, or more simply to just hurry up and get done with the survey.

In this experiment respondents guessed they were talking to a white interviewer 76% of the time. Respondents guessed they were talking to a black interviewer 7% of the time, and to an interviewer of another race 17% of the time. Somewhat surprisingly, blacks were guessed white 66% of the time, guessed some "other" race 17%, and guessed to be black only 16% of the time. There is no way to tell if, on average, respondents perceive that interviewers should or are more likely to be white, but the data seem to show such a pattern.

Yet, regardless of whether respondents are satisficing or interviewers are presenting themselves in a more ambiguous voice, the results show that racial identification over the phone varies by interviewer race, and can vary by respondent race when confidence is considered. Both satisficing and speech assimilation point to a need to go beyond the

[6] Error Rate for black respondents (34%) was not significantly different from whites (29%) (t=1.63, n.s.).

assumptions of past ROI and LP racial effects and further research the actual practices and behaviors related to race in telephone (or other non-face-to-face) settings.

References

Alim, Sammy H. (2003) "You Know my Steep: An Ethnographic and Sociolinguistic Study of Styleshifting in a Black American Speech Community." Unpublished PhD dissertation. Stanford, CA: Stanford University.

Anderson, Barbara A., Brian D. Silver, and Paul R. Abramson. 1988. The effects of race of the interviewer on measures of electoral participation by blacks in SRC National Election Studies. Public Opinion Quarterly 52 (1):53-83.

Baugh, John. 1983. Black street speech. Austin, TX: University of Texas Press.

Callegaro, Mario , Femke De Keulenaer, Jon A. Krosnick, and Robert P. Daves (2005) "Interviewer Effects in a RDD Telephone Pre-election Poll in Minneapolis 2001 An Analysis of the Effects of Interviewer Race and Gender." 2005 AAPOR Conference Proceedings.

Cotter, Patrick R., Jeffrey Cohan, and Philip B. Coulter. 1982. Race-of-interviewer effects in telephone surveys. Publc Opinon Quarterly 46 (2):278-284.

Davis, Darren W. 1997a. The direction of race of interviewer effects among African-Americans: donning the black Mask. American Journal of Political Science 41 (1):309-322.

Davis, Darren W. 1997b. Nonrandom measurement error and race of the interviewer effects among African Americans. Public Opinion Quarterly 61 (1):183-207.

Davis, Darren W., and Brian D. Silver. 2003. Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Poliical Science*. 47 (1):33-45.

Finkel, Steven E., Thomas M. Guterbock, and Marian J. Borg. 1991. Race-of-interviewer effects in preelection poll: Virginia 1989. *Public Opinion Quarterly* 55 (3):313-330.

Friedman, Linda, and Hershey H. Friedman. 1978. Does the perceived race of a telephone in-terviewer affect the responses of white subjects to an attitudes towards negroes scale. In Proceedings of the Section on survey research methods, edited by American Statistical Association. Washington D.C.: American Statistical Association.

Gurin, Patricia, Shirley Hatchett, and James S. Jackson (1985) *Hope and Independence: Blacks' Response to Electoral and Party Politics*. NY: Russell Sage Foundation.

Hatchett, Shirley , and Howard Schuman. 1975. White respondents and race-of-interviewer ef-fects. Publc Opinon Quarterly 39 (4):523-528.

Higgins, Edward T., Gilliam A. King, and Gregory H. Mavin (1982) "Individual construct accessibility and subjective impressions and recall." *Journal of Personality and Social Psychology*, 43:35-47.

Hyman, Herbert H., William J. Coob, Jacob J. Feldman, Clyde W. Hart, and Charles Herbert Stember. 1954. Interviewing in social research. Chicago: University of Chicago Press.

Krosnick, John A and Duane Alwsin (1987) "An evaluation of a Cognitive Theory of Responde-Order Effects in Survey Measurement." *Public Opinion Quarterly*. 51:201-219.

Krysan, Maria, and Mick P. Couper. 2003. Race in the live and the virtual interview: Racial def-erence, social desirability, and activation effects in attitude surveys. Social Psychology Quarterly 66 (4):364-383.

Lind, Katherine, Robert Oldendick, Carol Hall, Judi Rose, and Ron Shealy (2001) "The Confederate Flag Debacle: An Examination of Interviewer Race on Respondent Answers to Racial Sensitive Subject." Proceedings of the Survey Research Methods Section, American Statistical Association (August). 1-4.

Majors, Richard and Jane Mancini-Billson (1993) *Cool Pose: Delimmas of Black Manhood in America*. NY: Touchstone.

Meislin, Richard. 1987. Racial divisions seen in poll on Howard Beach attack. New York Times, January 8, B2.

Purnell, Thomas, William Isardi, and John Baugh. 1999. Perceptual and phonetic experiments on American English dialect identification. Journal of Language and Social Psychology 18 (1):10-30.

Schaeffer, Nora Cate. 1980. Evaluating race-of-interviewer effects in a national survey. Socio-logical Methods and Research 8 (4):400-19.

Schuman, Howard, and Jean M. Converse. 1971. The Effects of Black and White Interviewers on Black Responses in 1968. Public Opinon Quarterly 35 (1).

Sudman, Seymour, and Norman A. Bradburn. 1974. Response effects in surveys. A review and synthesis. Chicago: Aldine.

Tate, Katherine, et al. The 1984 National Black Election Study Sourcebook. Ann Arbor, MI: Institute for Social Research, 1988.

Tate, Katherine., Ronald E. Brown, Shirley J. Hatchett, and James S. Jackson (1988). The 1984 National Black Election Study: A Sourcebook. Ann Arbor, Institute for Social Research.

Tate, Katherine. 1998. National black election study 1996 [Codebook]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Thomas, Erik R. and Jeffrey Reaser (2004) "Delimiting Perceptual Cues Used for the ethnic Labeling of African American and European American Voices." *Journal of Sociolinguistics*. 8:54-87.

Walton, Julie H., and Robert F. Orlikoff. 1994. Speaker race identification from acoustic cues in the vocal signal. Journal of Speech and Hearing Research 37:738-745.

Williams, Allen J., Jr. 1964. Interviewer-respondent interaction: A study of bias in the information interview. Sociometry 27 (3):338-352.

Williams, Allen J., Jr. 1968. Interviewer role performance: A further note on bias in the informa-tion interview. Public Opinon Quarterly 32 (2):287-94.

Wilson, David C., and Eric P. Olesen. 2002. Perceived race of interviewer effects in telephone interviews. Paper read at 57th AAPOR/WAPOR conference, at St Pete Beach, FL.

Wolford, Monica L., Ronald E. Brown, Anne Marsden, James S. Jackson, and Chase Harrison. 1995. Bias in telephone surveys of African Americans. The impact of perceived race of interviewer on responses. Proceedings of the Survey Research Methods Section, American Statistical Association. (799-804).

Worcester, Robert M., and Kully Kuar-Ballagan. 2002. Who's asking? Answers may depend on it. Public Perspective (May/June):42-3.

Appendix: Study Information

Final Sample Description. Each cell in Table 3 contains the planned and actual numbers of respondents. The final sample size for the data is 1,026. The sample is characterized as having higher education and incomes, older, mostly female and White, located in the western part of the US. Prior to this study, members of the sample have been asked to participate in anywhere between 3 and 19 studies, with an average of 9, and they completed anywhere between 0 and 18 of them, with an average of 7. These final numbers indicate that the sample is fairly familiar with Gallup, as well as it's interviewing process.

Education and Income. The final respondents are highly educated and earn higher incomes. The majority of respondents are college graduates, with 30% having completed a bachelors program and 23% having post-graduate work. Twenty-six percent of the respondents have some college, while only 16.8% are high school graduates or less. This indicates that the sample over-represents education. With regard to income, average household income is between $50K - $70K, with 16.8% making under $25K, and 17% making over $100K. In all, 54% of the sample has household salaries over $50K. This is a clear over-representation of income.

Other Demographics. The majority of respondents are married, tend to be older, and tend to be located in the western region of the US. Fifty-five percent of respondents are married, 20.7% are single, 15.5% are divorced or separated, and 8.4% are widowed. The data undercount younger people and over-represent seniors. The average [mean] age in the sample is 52.7 years (median=52), with slightly more than 75% of the sample being over the age of 40. With regard to region, the fewest respondents (14.3%) are located in the northeast, and the most are located in the west (33%). The midwest and south regions have 24.5% and 28.2% of the respondents respectively.

Interviewer Location. All interviews were conducted from Gallup interviewing centers (IC) in Lincoln and Omaha, Nebraska. The majority, 644 (63%), came from Gallup's 68th Street IC in Lincoln, second was Gallup's Omaha Farnam IC with 242 (24%), third was Omaha's Greentree IC with 115 (11%), and the fewest interviews were conducted at Gallup's Downtown Lincoln IC with 25 (2%).

Interviewer Demographics. The interviewers were mostly white and female. Seventy-nine percent (38) of the interviewers were white, 19% (9) were black, and 2% (1) were self-reported as some other race. The gender mix was slightly more balanced with females comprising 58% (28) and males 42% (20) of the interviewers. The race by gender mix contained 23 (48%) white females, 15 (31%) white males, 4 (8%) black females, 5 (10%) black males, and 2 (2%) females in the other category.

The interviewers tended to be more mature in terms of age. While the age of the interviewers ranged from 16 to 62 years, the mean interviewer age was 29 (Median=21) years old, with a standard deviation of 14 years. The average age for both white and black interviewers was 28, while the single interview in identified as other was 38. Also, both male and female interviewers had an average age of 28 years old.

Interviewer Assignment. In the study, each interviewer was randomly assigned to a respondent and then randomly assigned a questionnaire (form) for the interviewer. There was no attempt to match respondent to interviewer. There were a total of 48 interviewers who participated in the study with at least one completed interview.

The study contained only black and white respondents, and most of the respondents were interviewed by white interviewers. Seventy-eight percent (470) of white respondents were interviewed by white interviewers, 16% (98) by black interviewers, and 6% (35) by

the other race interviewer. Black respondents had similar proportions. Seventy-three percent (309) of black respondents were interviewed by white interviewers, 20% (85) by black interviewers, and 7% (29) by the other race interviewer.

Interviewer Productivity. There were extremes of both high and low productivity among the 48 interviewers who participated in the study. On average each interviewer completed 21 interviews, although the median number of completes was 5 interviews. There were 7 interviewers with 64 or more completes, and 2 of those seven (both white females) completed at least 125 interviews. Moreover, 70% of the interviews were conducted by only 10 (about 20%) interviewers (1 white male, 2 black males, 6 white females, and 1 other female), each completing a minimum of 30 interviews. Low productivity was also common. There were 24 interviewers who completed fewer than 5 interviews and 30 who completed 10 or less.

Crosstabulation for Actual and Perceived Race of Interviewer

	PROI			
AROI	White	Black	Other race	Total
White	82% (617)	5% (36)	13% (100)	100% (753)
Black	66% (110)	16% (27)	17% (29)	100% (166)
Other Race	32% (20)	11% (7)	57% (36)	100% (63)
Total	76% (747)	7% (70)	17% (165)	100% (982)

Notes. 2005 Race of Interviewer Experiment, Gallup.

Table 1. Perceptibility Error Rates found in the Race of Interviewer literature

Study	Error Rate		
	Overall	**Blk Int**	**White Ints**
1984 National Black Election Study; Tate (1984) - Black Resps only	27%	34% [b]	9% [b]
1986 Howard Beach Survey Meislin (1987)	25%	Not reported	Not reported
1993 National Black Politics Study; Wolford et al. (1995) - Blk Resps only	27%	27%	NAP
1996 National Black Election Study; Tate (1998) - Blk Resps only [a]		37%	29%
1999 Gallup Race Relations Social Audit; Wilson and Olesen (2002)	18%	26%	14%

Notes. [a] Adapted from Callegaro et al. (2005). [b] Error rate does not include guesses for "Other" [than black or white] interviewers.

Table 2. Examination of Previous Gallup National Studies Assessing Race of Interviewer Perceptibility.

		Error Rate				PCT W/B/O	
Gallup Study	**N**	**Overall**	**White Ints**	**Black Ints**	**Other Ints**	**AROI**	**PROI**
1998 Black-White Study	2,004	35%	28%	43%	54%	60/33/7	57/26/17
1999 Black-White Study	2,006	34%	26%	35%	79%	58/34/8	54/32/14
2001 Black-White Study	2,004	28%	17%	36%	86%	60/33/7	61/26/13
2004 Minority Rights and Relations Study	2,250	33%	23%	51%	69%	68/25/7	65/18/17
2005 Minority Rights and Relations Study	2,264	38%	26%	48%	76%	49/48/3	55/29/16
2005 Employee Discrimination in the Workplace Study	1,252	30%	27%	50%	73%	92/6/2	70/7/23
Totals Across 6 Studies	11,780	33%	24%	43%	73%	66/27/7	60/24/16

Notes. Each study asked "guess interviewer race" at the end of the survey; "PCT W/B/O" is the percentage of interviews conducted by Whites, Blacks, and Others; AROI = actual race of the interviewer, PROI = perceived race of the interviewer; Error Rate is the % guessed incorrectly.

Table 3. Experimental Design: 2 x 3 Design - Respondent Counts

	Introduce Race			
Order of Item Set	*Beginning of Interview*	*Middle of Interview*	*End of Interview*	**Totals**
Genera/Political Items First	(Prop N =200) (Act N = 184)	(Prop N=200) (Act N =170)	(Prop N=200) (Act N = 171)	Prop N=600 Act N = 525
Race Items First	(Prop N=200) (Act N = 168)	(Prop N=200) (Act N = 168)	(Prop N=200) (Act N = 165)	Prop N =600 Act N = 501
Totals	Prop N = 400 Act N = 352	Prop N =400 Act N = 338	Prop N =400 Act N =336	Prop N =1,200 Act N = 1,026

Notes: Prop N = Proposed cell size; Act N = Actual cell size. χ^2 = .285, df=2, sig.=.867

Table 4. Error Rates in Perceptibility and Confidence Across Interviewer and Respondent Characteristics

Breakout	Error Rate (Incorrect Guess)			
	Overall	For White Interviewers	For Black Interviewers	For Other Interviewers
Sample	31% (1,026)	18% (779)	84% (183)	43% (64)
Respondent Race				
White Respondents	29% (603)	18% (470)	82% (98)	38% (35)
Black Respondents	34% (423)	19% (309)	86% (85)	48% (29)
Respondent Gender				
Male	30% (412)	17% (313)	85% (68)	43% (31)
Female	31% (614)	19% (466)	83% (115)	42% (33)
Respondent Race x Gender				
White Male	28% (249)	17% (197)	85% (36)	33% (16)
White Female	29% (354)	18% (273)	80% (62)	42% (19)
Black Male	34% (163)	18% (116)	84% (32)	53% (15)
Black Female	33% (260)	19% (193)	87% (53)	43% (14)

Breakout	Mean Confidence			
	Overall	For White Interviewers	For Black Interviewers	For Other Interviewers
Sample	3.29 (1,026)	3.30 (779)	3.30 (183)	3.16 (64)
Respondent Race				
White Respondents	3.00 (764)	3.05 (458)	2.81 (89)	2.83 (35)
Black Respondents	3.70 (415)	3.68 (306)	3.84 (80)	3.55 (29)
Respondent Gender				
Male	3.38 (412)	3.41 (313)	3.29 (68)	3.26 (31)
Female	3.23 (614)	3.23 (466)	3.30 (115)	3.06 (33)
Respondent Race x Gender				
White Male	3.11 (238)	3.19 (189)	2.82 (33)	2.75 (16)
White Female	2.92 (344)	2.95 (269)	2.80 (56)	2.90 (19)
Black Male	3.78 (162)	3.78 (115)	3.78 (32)	3.80 (15)
Black Female	3.65 (253)	3.62 (191)	3.88 (48)	3.29 (14)

Notes. None.

Table 5. Logistic and OLS Regression Estimates for Race, Perceptibility and Confidence.

	Error			Confidence		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Race (Whites)	-.262	-.267	.047	-.641**	-.643**	-.757
Sex (Males)	-.015	.007	.006	.167*	.167*	.167*
Education (College)	-.065	-.013	-.017	-.117	-.119	-.118
Age	.006	.004	.004	.014**	.014**	.014**
% Black	2.19**	1.008	1.191	.164	.199	.134
% White	2.24**	1.262	1.568	-.129	-.097	-.176
AROI (Black Interviewer)		3.0**	3.0**		-.077	-.078
% Black x Race			.209			.119
% White x Race			-.445			.137
Constant	-2.89**	-2.52**	-2.75**	2.99**	2.98**	3.04**
R^2	.03**	.33**	.33	.11**	.11	.11
N	954			969		

Notes. ** p<..01; * p<.05;

Table 6. Correct Guesses and Confidence by Experimental Main Effects and Interaction

	Introduce Race [a]			
Order of Item Set [b]	*Beginning of Interview*	*Middle of Interview*	*End of Interview*	**Totals**
Genera/Political Items First	PCT=33% M=3.30	PCT=28% M=3.29	PCT=34% M=3.24	PCT=32% M=3.28
Race Items First	PCT=30% M=2.28	PCT=29% M=3.20	PCT=30% M=3.44	PCT=30% M=3.31
Totals	PCT=32% M=3.29	PCT=29% M=3.25	PCT=32% M=3.34	PCT=31% M=3.29

Note. Percentages (M) represent correct guesses and Mean (M) values represent the average level of confidence in the guess. [a] Main effect of PCT (Z=.58, n.s.) and M (F=.48, n.s.); [b] Main effect of PCT (Z=.67, n.s.) and M (F=.13, n.s.); Overall Interaction of [a] and [b] PCT (Z=.48, n.s.) and M (F=.65, n.s.).

Table 7. Logistic Regression Estimates of Perceptibility of Interviewer Controlling for Survey Interview Factors and Confidence

	Error				
	Model 1	Model 2	Model 3	Model 4	Model 5
Confidence	-.267**	-.374**	-.454**	-.459**	-.440**
AROI (Black Interviewer)		3.17**	3.17**	3.190**	.999
Race (Whites)			-.482**	-.499**	-.474**
Sex (Males)			.078	.072	.081
Education (College)			-.069	-.070	-.095
Age			.011	.012*	.012*
Stage (Beginning)				-.227	.185
Stage (Middle)				-.134	.706
Order of Items (Race First)				-.150	-.183
Stage x Order [a]				-	-
Stage (Beginning) x Confidence					-.149
Stage (Middle) x Confidence					-.285
AROI (Black Interviewer) x Confidence					.657**
Constant	.048	-.221	-.261	-.050	-.172
R^2	.03**	.35**	.36*	.37*	.39*
N	971				

Notes. ** p<..01; * p<.05; [a] The "Stage x Order" interaction was not significant in Models 4 (Wald=.029, n.s.) and 5 (Wald=.041, n.s.). Since the 2 x 3 interaction results in 5 dummy coded variables for the analysis, the coefficients are removed from the table due to space considerations.

Table 8. Predicted Probability of Inaccurate Guess of Interviewer Race, Controlling for Confidence and Respondent Race.

	Respondent Race		
Confidence	Black	White	Total
Completely unsure	60% (33)	47% (94)	50% (127)
2	46% (24)	34% (91)	37% (115)
3	33% (109)	27% (191)	29% (300)
4	28% (109)	19% (125)	23% (234)
Completely confident	30% (137)	21% (76)	26% (213)
Total	34% (412)	29% (577)	31% (989)

Note. Estimates are the predicted probabilities from logistic regression analysis based on Model 5 in Table 7.