

Collecting and using race and ethnicity information in linguistic studies

Robert Squizzero^{*a}, Martin Horst^a, Alicia Beckford Wassink^a, Alex Panicacci^{bc}, Monica Jensen^a,
Anna Kristina Moroz^a, Kirby Conrod^a, Emily M. Bender^a

Department of Linguistics, University of Washington
Box 354340, Seattle, WA, 98195-4340
rsquizz@uw.edu

^a Department of Linguistics, University of Washington

^b Department of Linguistics, Queen Mary University of London

^c Department of Psychology, University of Washington

*Corresponding author

Abstract: Despite being regarded by some as the most humanistic of the social sciences, linguistics has been criticized for its undertheorized application of the notions of race and ethnicity. This white paper is written for practicing linguists. We provide definitions of these terms and develop attendant issues that contribute to their complexity, such as the multiplicity and fluidity of racial identification. A survey of methods texts reveals that limited attention is given to race and ethnicity in training researchers. To address this need, the bulk of the paper uses a “challenges and recommendations” format to work through common design concerns and suggest better practices. We consider issues pertinent to collecting information about self-identification in a range of study types, from quantitative, experimental, computational or intuitional approaches to qualitative and mixed methods designs. We consider the advantages and disadvantages of eliciting demographic data using multiple-choice, free-response and interview formats, and offer recommendations drawing on best practices from within linguistics and its sister fields. Ethical concerns are raised, including using locally constructed labels, respecting communities, analyst positionality, recognizing the potential for harm. Throughout, brief examples are provided where possible to speak concretely to linguists’ concerns. The final section presents a detailed case study of the decision-making process for a multi-phase research project in which ethnic identification was explicitly investigated. Our goal is to provide researchers with tools to reflect on their own study design, reflect on their own responsibility to participants and communities, and design study prompts that allow more nuanced representation of race or ethnicity.

Keywords: race, ethnicity, sociolinguistics, research methods

CONTENTS

[0. About the committee and this project](#)

[1. Introduction](#)

[1.1 Defining race and ethnicity](#)

[1.2 State of the field](#)

[1.3 Survey of Research Design Advice](#)

[1.3.1 Guidance regarding study design](#)

[1.3.2 Texts that “didn’t go there”](#)

[1.3.3 Texts that “went there”... kind of](#)

[2. Research question design](#)

[2.1 How do linguists go about asking questions? Inductive v. deductive reasoning in linguistics research design](#)

[2.2 What makes a good research question?](#)

[3. Methods of collecting data](#)

[3.1 Method: demographics collected through multiple-choice survey](#)

[3.1.1 Strengths of multiple-choice surveys](#)

[3.1.2 Weaknesses of multiple-choice surveys](#)

[3.1.3 Example of a study which uses a multiple-choice survey](#)

[3.1.4 Summary and Recommendations](#)

[3.2 Method: demographics collected through free-response written or online survey](#)

[3.2.1 Strengths of the free-response method](#)

[3.2.2 Weakness of the free-response method](#)

[3.2.3 Example of study which uses a free-response question](#)

[3.2.3 Summary and recommendations](#)

[3.3 Method: demographics collected through interactive interview](#)

[3.4 Method: perception experiments](#)

[3.4.1 Types of perception studies in sociolinguistics](#)

[3.4.2 Confounding factors and generalizability](#)

[3.4.3 Balancing risks and benefits in perception studies](#)

[3.5 Method: demographic info was not collected](#)

[3.5.1 Grammatical intuitions and the source of example sentences](#)

[3.5.2 Experimental studies](#)

[3.6 Method: Computational methods and large datasets](#)

[3.7 Collection of data that potentially endangers participants](#)

[4. Potential Issues and Our Recommendations Regarding How to Address Them](#)

[4.1 Analyst positionality](#)

[4.2 Issue type: over-reliance on category labels](#)

[4.2.1 People’s authentic self-identifications and intergroup variation](#)

[4.2.2 Outsider’s categorization of the self](#)

[4.2.3 Fluidity of racial identities](#)

[4.2.4 People with multiracial/racially ambiguous identities](#)

[4.2.5 Intergroup perceptions and dynamics](#)

[4.2.6 General good practices when using categories](#)

[4.3 Issue type: possible confounds in your data \(and how to discuss them\)](#)

[4.3.1 Intersectionality should inform your understanding of social categories](#)

[4.3.2 Certain quantitative methods can obscure insight](#)

[4.3.3 Sampling bias may affect your findings](#)

[4.3.4 Insufficient data collection may affect your findings](#)

[4.3.5 Main takeaways](#)

[4.4 Issue type: how to identify and talk about potentially harmful implications of one's work](#)

[4.4.1 Writing to prevent bad-faith citation](#)

[4.4.2 Handling bad-faith citation](#)

[5. A Case Study: The Pacific Northwest English Study \(PNWE\)](#)

[5.1 Context of case study](#)

[5.1.1 Introduction to the PNWE](#)

[5.1.2 PNWE design considerations](#)

[5.1.3 Handling unexpected situations in the field](#)

[6. Conclusion](#)

[Appendix 1: Author positionality statements including: linguistic subfields, race/ethnicity \[and gender\], and more](#)

[Appendix 2: Linguistics methods texts surveyed for guidance regarding treating race and ethnicity in study design](#)

0. About the committee and this project

The authors of this white paper are affiliated with the University of Washington Sociolinguistics Laboratory, whose biweekly meetings provide a space to discuss topics within sociolinguistics. We are graduate students, lecturers, or professors within the UW Department of Linguistics with differing gender, race, ethnicity, sexual orientation, ability, profession, and lived experiences informing our work in sociolinguistics. Members of our group hold memberships on the LSA Ethics Committee, the LSA Committee for LGBTQ+ [Z] Issues in Linguistics, and the executive committees of the American Dialect Society and the Linguistic Society of America. Our collective goal in writing this work is to elucidate the extent to which linguists have historically included racial and ethnic identity data in their research. We strive to highlight problems concerning the collection and use of these data, offer recommendations on how to better one's research, and connect to the larger goals of linguistic social justice (see e.g. Bucholz et al. 2014, Charity Hudley et al. 2020). Because we recognize that our individual identities shape our investment in this issue, positionality statements for each author appear in Appendix 1.

This white paper project is the outgrowth of a series of Sociolinguistics laboratory meetings from November 2019-June 2020 on the theme, "Race and Ethnicity in Studies of Language Variation and Change." This series focused on questions in three broad areas: 1) Racializing and essentializing practices in the field of linguistics, 2) methodological considerations (do I need to collect ethnicity-related information, and if so, how much, and what

kind?), 3) ways of increasing representation of underrepresented groups and incorporating more nuanced ethnic self-identification data in sociolinguistic research. Serendipitously, the Linguistic Society of America Statement on Race was newly published (in May 2019). Available to us, as well, was a newly revised slate of race and ethnicity questions in the 2020 American Community Survey (US Census). We were fortunate to be able to discuss both. We closely considered the complex layering of national, ethnic, cultural, and racial identities explored in Smith (2019). With the author as our guest for these meetings, we were able to ask questions about her own design and interview process, working with expatriate Senegalese multilingual speakers in the United States, France, and Italy. The outcome was a decision to coauthor a paper responding in part to our studies, and in part to the LSA's efforts to "foster dialogue and encourage linguists to critically reflect on the changing nature of academic, social, cultural and linguistic understandings of race" (Linguistic Society of America, 2019).

1. Introduction

In research on language production conducted in the tradition of language variation and change, linguists, especially in the United States, continue to directly and uncritically ascribe differences in language to race, often making use of predefined, widely used racial categories. This is evident, for example, by the use of racially essentializing terms for supposedly monolithic language varieties, such as African American English or Chicane English (cf. Charity Hudley, Mallinson, & Bucholtz 2020). These uncritical connections between language and the race of those who produce it contribute to and sustain socially imposed divisions between arbitrarily-defined racial groups. Such connections have often been created by White scholars who neglect to involve members of racially minoritized groups in their research practices, subscribing to and even perpetuating the deeply flawed ideology that the only possible objective viewpoint is that of an outsider. We argue that linguists in *all* subfields of linguistics should carefully consider whether differences in language production are truly race-based or if they are, in fact, ethnicity-based or practice-based (see Section 1.1).

Lacking a cohesive model for race and ethnicity (until the LSA's formal statement published May 2019), linguists have relied too heavily on racialized, taxonomy-like approaches for their research. Charity Hudley (2017) details how sister disciplines (sociology, archeology, anthropology, etc.) have even used linguistics to exemplify how *not* to model race within the field. Despite the overall underdevelopment of racial theory within linguistics, some linguists have begun to highlight research-based and institutionalized models of race impacting language policy (see Charity Hudley and Mallinson (2011); Garcia Sanchez (2014)). These calls (among others) to reform racial models employed throughout the field form the foundation for this white paper.

1.1 Defining race and ethnicity

Race and ethnicity are often grouped together or used interchangeably, both in everyday use and in academic writing. This section defines race and ethnicity following their theorization in the tradition of sister fields to linguistics, especially anthropology, psychology, and sociology, rather than in the tradition of linguistics itself. We acknowledge that there are shared aspects of race and ethnicity in sister fields to linguistics, and that there is disagreement within these fields as to the

precise definitions of these two terms. Still, this section argues that distinguishing the two concepts is important in linguistic research.

In this paper, *race* is defined as a classification of people into arbitrary and mutually exclusive groups based on physical features, especially skin color, facial features, eye shape, and hair texture (Bobo 2001). *Race essentialism* is the tendency to view race as biologically based, immutable, and informative (e.g., Haslam, Rothschild, & Ernst 2000; Prentice & Miller 2007) and has been linked to racial stereotyping and prejudice (Levy & Dweck 2003, Williams & Eberhardt 2008). Essentialist racial classification schemes that are based on biology or genetics have been shown to be unreliable at best (García 2020) and entirely arbitrary at worst (Omi & Winant 2015). In contrast to essentialist theories on race, *constructivist* theories emphasize that social and historical forces are what create a popular perception of differences between members of groups that tend to share one or more specific phenotypic features (Sen & Wasow 2016). In addition to perceived physical features, racial classification is also based on expectations, activated stereotypes, and social perception (Freeman, Penner, Saperstein, Scheutz & Ambady 2011), as well as cultural and social history (Charity Hudley 2017), and is often conflated with nationality and language (Rosa & Flores 2017). All these interacting aspects guide how a subject is categorized by different people, including linguists, in different contexts and at different times, following what Omi & Winant (2015) describe as *racial formation*, the “sociohistorical process by which racial identities are created, lived out, and destroyed” (p. 109), a process which can operate within and across individuals’ lifespans.

Ethnicity is defined as a classification of members of a group sharing one or more aspects of a common culture. Comaroff & Comaroff (2009) write that ethnicity is

“...best understood as a loose, labile repertoire of signs by means of which relations are constructed and communicated; through which a collective consciousness of cultural likeness is rendered sensible; with reference to which shared sentiment is made substantial.” (p. 38).

To expand, aspects of a common culture include shared, locally relevant knowledge and value systems and a sense of belonging and community. According to the American Anthropological Association and the Society for Anthropology in Community Colleges, material manifestations of shared aspects of culture that are used as cues for community members to navigate spaces include, but are not limited to, following patterns of dress, adhering to diets or eating particular foods, observing holidays, practicing religions, and speaking languages and language varieties (García 2020). While heredity is included in some definitions of ethnicity, we argue that ethnicity is mostly *practice-based*, rather than based on genetic or phenotypic qualities; individuals can choose to participate in practices that associate themselves with a particular ethnic group (Cheng 2003). This participation can be an important component of one’s social and personal identity, and the language variety used can be reflective of attachment towards one’s ethnic group, according to *ethnolinguistic identity theory* (Giles & Johnson 1987). In contrast to racial categories, which are generally arbitrary, socially imposed, and perceptual in nature, ethnicity is both self-recognized by a collectivity and recognized by outsiders (Fishman 1977), though whether a person’s ethnic affiliation is recognized depends both on social context and individual-level variation (Clément & Noels 1992; Gurin, Peng, Lopez & Nagda 1999; Hurtado, Gurin & Peng 1994; Noels, Leavitt & Clément 2010). The ethnic affiliation of members of socially subordinate ethnic groups may also go unnoticed, such as in cases of *intersectional invisibility*

(Purdie-Vaughns & Eibach 2008), where these individuals are also members of other socially subordinate groups, e.g. based on sexual orientation (see Section 4.3.1). Lastly, analogous to the way that race is shaped by sociohistorical forces, one's ethnic identification can change over time via acculturation processes (Liebkind et al. 2016, Phinney & Ong 2017) or in response to contextual factors, which may have crucial consequences for people's psychological well-being. (Yip, 2005).

Linguistics stands to benefit from the greater problematization of race that is taking place in neighboring fields (for an overview of this and its relevance to linguistics, see Charity Hudley et al. 2020). To provide just one of many possible examples, López, Vargas, Juarez, Cacari-Stone, & Bettez (2017) propose multidimensional measurement of race. In their study of Latinx Americans, respondents are asked to provide their race in three contexts: *street race*, i.e. how a respondent believes a stranger who saw them walking down the street would classify their race, *ascribed race*, i.e. how a respondent believes others usually classify their race in the United States, and *self-perceived race*, i.e. how a respondent usually self-classifies their race on questionnaires. A more nuanced view of race explicitly acknowledges that racial categorization is context dependent, and it can ultimately lead to greater descriptive accuracy in the relationship between language and race. Linguistics also stands to benefit from a view of ethnicity that takes into account its variable nature based on reference group, social context, and situational context (for a comprehensive overview of this variation as pertains to sociolinguistic research, see Noels 2014). For more on racial and ethnic categorization, see Section 4.2.

In most cases, linguistic research questions (regardless of subfield) tend to probe issues of language use and structure, which emerge by virtue of a speaker's or signer's setting of language acquisition and linguistic practice, supported by membership in a (or several) speech community(ies). These are practices, and they are not deterministically inherited. In most such cases, then, the linguist may be said to primarily be concerned with ethnicity. However, it is clear that research participants may themselves conflate race with ethnicity. The researcher must be sensitive to community-members' usages, so it may be necessary to use the term "race" in communicating the goals of the research project, where study of groupings of people is part of study design. It is also the case that racial categories are fluid, and that racial identity is fluid and can be ambiguous (see Section 4). In some cases, contextual or locally constructed notions of race can help the analyst understand practices by or oriented toward, speakers (see Dovchin, 2019a, b; Espinosa, Tikhonov, Ellman, et al., 2018; Rzepnikowska, 2018). But it is imperative for researchers to understand the basis of the grouping (and describe or clarify such states of affairs carefully in any published work). Outgroup perceptions of speakers can, for example, help to inform these speakers' linguistic choices. Again, this is not behavior based upon their biological heritage, but reflects the intersection between practice, social dynamics, and racial consciousness. When race is mentioned in the pages below, it is always with regard to social dynamics, social construction of power inequalities, and locally constructed notions of group salience, never biology (or any of its close cousins, such as phenotype).

Linguists should never talk about biological dispositions toward language use.

1.2 State of the field

To contextualize the pursuit of racial linguistic justice in this paper, the following section provides a historical cross section of differing models of race as found within the linguistics

literature. Paralleling social conceptions of racial identity, the theoretical models of race and racial classification defined by modern linguist study have undergone massive changes. Extending the work of Eckert (2012), Charity Hudley (2017) details how the three waves of sociolinguistic variation study (see Eckert (2005)) conceptualized race, while adding a fourth wave to the mix. The first wave analyzed linguistic features (most frequently grammar and phonology) correlating these to population-level factors, which included a racial classification taxonomy and “gross notions of community as stand-ins for more in-depth racial analysis” (2017:9). Labov et al. (1968) exemplifies this model; the authors test whether certain peer groups can determine the race of speakers from auditory stimuli alone. The second wave employed ethnographic techniques to question the social significance of placing individuals into different, racially based groupings, using both internal and external definitions from individuals and communities (see e.g. Eckert (1989) for an ethnography of White adolescent peer groups in Detroit). The third wave incorporates even more speech community stakeholders and ethnographic practices into its investigation. While previous models of race generally compared linguistic groups *interracially*, Mendoza-Denton (2008) studied linguistic differentiation within one socioethnic group and provided her participants the opportunity for self-identification. Charity Hudley (2017) notes that more recent models transcend altogether the taxonomic view of race:

The fourth wave emphasizes the differences between what you learn about a language and/or racial group by studying it and what you learn by living the experience. In this model, both linguistic and racial ideology are co-constructed and co-negotiated between researcher, individual, and community. As such, the emphasis is on what the individual, group, race, and/or culture value and see as crucial to the investigation of language, as well as linguistic social justice. (2017:8)

While these four waves illustrate the distance between linguistic models of race past and present, Bucholz (2020:242) concisely expresses the current reality: “[linguists] have by and large failed to demonstrate an equal commitment to the well-being of the communities where we conduct our research.” Evidence of this may be found in the enduring lack of linguists of color participating in peer-reviewed research (LSA 2018:28) as well as the dearth of leadership within the LSA itself (LSA 2019a). To that end, a number of scholars have sought to examine exactly that form of racial justice within the discipline. While the purpose of this white paper differs from analyzing what structural changes must happen for linguistic social justice to truly take place, we recommend Charity Hudley, Mallinson, and Bucholtz (2020), Baker-Bell (2020), DeGraff (2020), and Gaby and Woods (2020) for readers seeking more context and clarity on this subject.

1.3 Survey of Research Design Advice

It is commonplace for linguists-in-training, whether graduate students or linguists tooling up to conduct new original research, to look for written advice regarding research design. Students in linguistics programs frequently look for guides of their own accord, as part of a methods course, and/or at the behest of thesis supervisors. Textbooks in linguistics are an obvious place to start. But what do these texts recommend, and what advice are researchers-in-training likely to find when it comes to conceptualizing race or ethnicity, or to eliciting information about racial or

ethnic identification? As part of this project, we asked: are there linguistics methods texts that describe how race and ethnicity might be thoughtfully considered in study design?

To answer this question, a mini study was conducted to search for published guidance in both older and more recent methods textbooks. The approach was to look for texts that study designers might be exposed to in a basic search. Only English-language publishers were considered. In order to concentrate on the guides most likely to influence students (and other study designers), we attempted to sample texts most likely to be discovered using typical searching techniques:

- *Online web search*: 13 publishers' websites were browsed for books on research methods in linguistics, including: Blackwell, Bloomsbury, Cambridge, DeGruyter, Edinburgh, Elsevier, John Benjamins, John Wiley, MIT, Oxford, Palgrave, Routledge, and Springer
- *UW Libraries online catalog*: the keyword "linguistics," in combination with keywords "research methods," and "study design" were entered into the University of Washington Libraries online catalog.
- *LSA annual meeting handbook*: a recent copy of the LSA Meeting Handbook was perused for any titles not uncovered by the two previous methods, together with a set of flyers brought back from the conference's book display tables.

These procedures returned a sample of about 61 texts that appeared to contain relevant content. Appendix 2 contains a full list of these textbooks, ordered by Year, and listing Author/Editor(s), Title, and Publisher. Next, the table of contents for each of these was scanned. From this, notes were made regarding the subfield focus of the work (if one could be determined), types of methods covered, whether any of the chapters mentioned sampling considerations around race or ethnicity, or failing this, human subjects considerations. Not included were general introductory texts explicit in *not* presenting themselves as methods texts. For example, texts meant to introduce a subfield of linguistic inquiry might not cover study design as much as talk about the subfield's central defining issues, canon of core establishing research, and types of linguistic data. Despite the approach to search, which was to incorporate keywords meant to narrow the results to titles about or containing "study design" or "research methods," some of the titles turned out to talk primarily about data analysis (not methods). This means that some data analysis-themed texts "slipped through." Where this happened, however, these titles were kept in the sample because the goal, as stated above, was to see what first-time researchers would find on a similar search. It is certainly expected that this approach likely missed some relevant source materials. Future versions of this work will hope to incorporate more resources as these become known.

The books were then broadly classified into two groups. From the point of view of consideration of race and ethnicity, books either "went there" or "didn't go there." First, the next few paragraphs briefly describe what topics were common to the books surveyed. Following are brief summaries of the texts that didn't treat race and ethnicity in study design and those that did.

1.3.1 Guidance regarding study design

New researchers are most likely to find explicit guidance on how to conceptualize ethnicity or race in relation to research study design in texts published in critical sociolinguistics, applied sociolinguistics, language documentation, and language variation and change. Of the 61 texts

sampled, only two make explicit mention of conceptualizing race or ethnicity as part of study design.

1.3.2 Texts that “didn’t go there”

None of the texts mentioned in this section (i.e., the majority of those sampled) explicitly talk about conceptualizing race and/or ethnicity. Quite understandably, most linguistics methods texts surveyed focused on linguistic data. Čermák (2002:62) succinctly states that, “In a narrow sense, the goal of linguistic theory is an adequate understanding of linguistic data....” Similarly, articulated by Furbee (2010) is the idea that linguistic theories

propose testable inventories of universal categories, properties, relations, and interactions that may constitute a language; for a particular language, they define sub-inventories of these that are legitimate expressions of the general design” (2010: xxi).

Many texts focus on inductive methods in structural linguistics, deductive research techniques in generative approaches, classification of linguistic forms and features (Zellig, 1952; Campbell, 1997; Ender, Leemann & Wälchli, 2012), descent relationships and reconstruction (Hale, 2007); and linguistic issues pertaining to specific language families (Rauch & Carr, 2018). Beginning in the 1980s, more texts focus on quantitative methods, mathematical, and statistical methods (Woods, 1986; Švejc, 1986; Těšitelová, 1991; Manning, 2000; Paolillo, 2002; Bod, Hay and Jannedy, 2003; Bayley and Lucas, 2007; Baayen, 2008; Rasinger, 2008; Johnson, 2008; Gries, 2009; Hemforth, 2013; Xu, 2013; Holmes & Hazen 2013; Levshina, 2015; Jensen, 2017; Desagulier, 2017; deGroot & Hagoort, 2018; Winter, 2020). At the same time, we see an increase in texts providing guidance on applied or qualitative research techniques (Bell, 1981; Wetherell, Taylor and Yates, 2001; Wengraf, 2001; Li & Moyer, 2008; Franzosi, 2010; Zhu, 2016; Loerts, 2020; Rose, 2020). A small number of texts examine both quantitative and qualitative methods (e.g. Litosseliti, 2018; Milroy and Gordon, 2003).

The texts mentioned above, then, did not address issues concerning speakers or speech communities. A smaller set of texts center the speech community in study design. With the rise of experimental methods in linguistics many texts begin to follow a common format. Such texts begin with explication of how to define a research question, centering the researcher's interests and goals (Schilling, 2007; Gass, 2007; Podesva & Sharma, 2013). Determination of the appropriate population is discussed next, followed by general information regarding how to properly subset a population (based again upon the analysts' interests and goals), addressing such common questions as: how many speakers do I need? How much data do I need? Here, several texts in sociolinguistics are representative, including Milroy and Gordon (2003), Bayley and Lucas (2007), Tagliamonte (2009), and Podesva and Sharma (2013), Meyerhoff et al. (2015). Several chapter writers focus on the notion of the speech community, asking what counts as a community for the purposes of a given study in the first place? and at what scale (the city?, community of practice?, 2 people who share a given interaction?). Schilling (2007) discusses differing applications of the Labovian definition of speech community (which relies upon shared *language and evaluation of linguistic forms*). She carefully sets out how sociolinguistics, at the intersection of language and society, has long had to contend with the dilemma of whether to determine community on the basis of language or on the basis of “social matters.” Some

practitioners, such as Patrick (2002) treat community as a “socially-based unit of LINGUISTIC analysis,” while others such as Bucholtz (1999) treat it as a “language-based unit of SOCIAL analysis” (emphasis ours). She concludes that there is no one best answer; sociolinguistics benefits from studying a range of types of community.

Wagner (2007) begins with advice regarding the import of understanding the population of interest, but then moves to important considerations related to whether the study is intended to be representative, explicating the important distinction between probability vs nonprobability sampling (including guidelines for choosing between random, stratified random, systematic, and convenience sampling). Subsequent chapters typically cover assurance of validity, reliability, and replicability (Gass, 2007), selection of appropriate linguistic variables (Milroy & Gordon, 2003; Tagliamonte 2009), and choice of experimental method among a very wide range appropriate for linguistic research (Podesva & Sharma, 2013; Drager, 2018). Sometimes specific populations were considered, such as in texts discussing research with protected populations, such as when working with children (Hoff, 2012). Meyerhoff et al. (2015) review similar topics to those described above, but offer additional useful discussion, of the archiving and sharing of data, and issues of establishing who owns the data.

Thus, not all texts foregrounded entry into communities. Those that did presented general directions for selecting the level of analysis (centered on careful consideration of how the notion of speech community is defined). Gass (2007) frames the need that all researchers face of making choices given limited resources of time and money. Questions, she says, must be feasible in relation to the time and “budget of the problem.” More specific considerations regarding classification of people into social groupings were found in the texts covered in the next section.

1.3.3 Texts that “went there”... kind of

When it comes to texts that made specific mention of the need for researchers to attend carefully to community concerns or self-conceptions (including ethnic self-identification), a pattern may be discerned. These texts tended to come, as mentioned above, from just three broad subfields: documentation of endangered languages, applied linguistics (including critical sociolinguistics), and variationist sociolinguistics. Six texts had relevant content. Of these, only two (Milroy and Gordon 2003; Heller et al. 2018) explicitly mentioned issues related to the conceptualization of ethnicity. Thus, only two of 61 texts sampled actually “went there.” The texts in this section all tended to be books with an ethics focus. Some emphasized qualitative methodologies. Some were books about fieldwork. These tended to discuss labels to be used for speakers within the frame of ethical treatment of people. For example, Rice, in several publications (in Grenoble & Furbee, 2010; in Thieberger, 2012) provides guidance for documentary linguists in applying the protocols for responsible conduct of research with Human Subjects, as spelled out in the Belmont report (e.g., respect for persons, beneficence, and justice). It stands to reason that careful definitions of ethnic in-group status or self identification would be carefully attended to in studies of language endangerment and documentation. Many native and indigenous communities have suffered harm at the hands of outsider scientists. Rice urges that as part of linguists’ preparation, they must learn about not just the “obvious” issues of the structure of the language of study and its relatives. The well-trained linguist must also read about the culture in which they plan to work, talk to as many community members as possible, and seek to learn from fieldworkers who have already conducted research in the area. She emphasizes the need to avoid unintentional harm, to understand whether the analyst’s actions might produce good, and ensure that respect is not naïve.

In other words, the research needs to take seriously their relationship to the community of study, the individuals consulted, and local systems of knowledge (including identity construction).

Rice positions herself specifically in the context of documentary work in Canada, which is a “known” context for her. The emphasis on deep knowledge of communities runs through applied, critical (and anthropological) linguistics. Starfield (2007) sets forward guidelines specifying that ethnographers work with a community they “know.” She doesn’t say directly to use ethnic groupings that the community uses, but this could be implied from her emphasis on understanding locally salient social constructs.

Heller et al. (2018) in one of the opening chapters of their book *Critical Sociolinguistic Research Methods*, provide the very relevant recommendation that how the analyst names and defines both research sites and people is something that should be handled with care. They continue on to say that this care is not so much about representing groups “accurately,” but is more about “genealogy:” the analyst should know where the categories come from and why they make sense to the participants (2018:46). Most contemporary critiques caution against simplified and reified notions of group, place, community, and time. They consider a detailed example of who “counts” in Canada as First Nations, a term shot through with the baggage of colonial and neocolonial relations, power differentials and the ideologies that position researchers in relation to these differentials. One specific word of guidance they provide is this: that the researcher should find the ways that different distinctions are meaningful locally and seek to understand the ways that people make these meaningful. These are the distinctions that should be centered in the research.

Several scholars, including Eckert (in Podesva & Sharma, 2013) and Rice (2012) referenced models of ethics laid out by Cameron (1992). Rice discusses these as situated within sociolinguistic research. Three traditions of ethical research, according to Cameron, may be differentiated:

1. *research on subjects*: this model centers a concern to minimize damage and offset inconvenience to the *researched*. It may also include recognizing the contributions of the *researched*.
2. *advocacy research*: this model centers the *researcher’s* commitment not just to do research on subjects, but “on and for subjects.”
3. *empowering research*: this model centers research “on, for and with” the *researched*.

Rice and Eckert recommend that analysts clarify for themselves which of these traditions they are, and desire to be, positioned within. One useful point is that Cameron’s formulation clarifies that there is a *researched* as well as a *researcher* in any study using human subjects, a concept that disrupts the traditional focus of linguistic research on the analyst only (Charity Hudley 2017, and evident in the texts surveyed above). This model builds in an additive way from research “on” to “on and for” to “on, for, and with,” which attests to the relationship between the interactive conduct of research and the enlargement of the set of stakeholders who potentially benefit from research.

Hale (2001:76) offers the lucid guidance that, “[t]he scientific investigation of a given language cannot be understood in isolation. In carrying out field research, linguists are inevitably responsible to the larger human community which its results could affect.”

Milroy and Gordon (2003) was the only textbook that actually included a section explicitly on “Race and Ethnicity.” In the fourth chapter, entitled, “Language Variation and the

Social World (2003:108),” the authors talk about race and ethnicity directly. The general stated goal of the chapter is to discuss the types of regularities that researchers in variationist sociolinguistics look for in a corpus of language data. Their purpose in the chapter, however, is to explore “analytic and interpretative issues” that arise when analysts seek to account for differential usage by different “kinds of speaker” (2003:88). They do not speak of how to *sample* ethnicity-related groups, but of how to sample responsibly given the sociocultural context, and how to interpret ethnicity-related patterns.

Citing Giddens (1989), Milroy and Gordon talk about ethnicity and race as systems underlying the distribution of social power and inequality. They discuss the distinction between race and ethnicity in a way other texts did not. Ethnic groups are

formed by persons who share, or believe they share, common cultural characteristics. These are wholly learned, typically very early in life, and usually involve a sense of place and of a common history and destiny, a shared religion and social ideology, and a shared language or set of communicative conventions (109).

They go on to describe race as a problematic construct resulting from the still-common belief that “humans can be separated by biological criteria into different races.” The main problem they associate with these constructs is that “racial differences are best understood socially as being closely associated with ethnicity, in referring to *physical variations singled out by members of a community and treated as ethnically significant*” (Giddens 1989: 246; italics in original).

Ultimately, the focus of this discussion is on urging the analyst-in-training to understand that ethnicity is locally constructed, and to guard against making correlations between what may be construed as race and language variation. In line with Heller et al. (2018) above, ethnicity means different things in different communities, and the researcher must look for the significance of linguistic variation in local conditions and local social practices which have arisen from historical events.

The chapter summarized above is concerned with understanding the *effects of ethnicity*, not with how to measure ethnic group membership. However, in an earlier chapter of the book, entitled “Locating and Selecting Subjects,” Milroy and Gordon (2003) talk about the relationship between research question and method as regards specific studies looking at ethnicity. They offer Hazen’s (2000) study of Warren County, North Carolina as a worthy example of a project that goes beyond a binary sampling into groups of African Americans and Whites, which is sometimes done in studies of language variation in the US South. Hazen included a third, Native American group. The primary research question being addressed concerned the social changes in the community and the linguistic consequences of these. So, sampling with a view to the intersectionality of age and ethnicity was reasonable: in this case, stratification was based on the sociocultural experience of school segregation. He crossed ethnicity and age in a manner allowing him to partition into groups those who: 1) experienced segregated schools, 2) those who were in school at integration, 3) those who began school after integration (2003:33). This text centers the value of sociocultural knowledge, combined with a life-stage approach. Ethnicity is included because it was an important factor related to different experiences of social change that might have linguistic consequences.

This section has reviewed a sizable sampling of textbooks about research design in a range of the subfields of linguistics. A (relatively) quick hunt for book-length resources was used to make clear what types of sources a researcher is likely to “turn up” on a focused search, perhaps without being pointed to particular (e.g., article-length) resources. We have seen that different considerations of what constitute the primary data of linguistic research produce different types of design process. Crucially, texts that focus on linguistic data abstracted away from people and communities emphasize design and collection of linguistic forms to address research questions about the properties of, relations, and interactions between linguistic structures. The texts that deal with Hale’s notion of a larger human community tend to have a focus on documentation, endangerment, ethics, and the construction of local identities. From this survey, most texts a researcher might find will be those with a language data focus. It is more difficult to find methods texts that center approaching speech communities with a view to thoughtfully collecting demographic information about locally meaningful ethnic distinctions. The point is not to detract from the need for training in the study of language data and the rich set of tools (including experimental methodologies) that are available to linguists. The hope is that the above review might provide researchers-in-training some guidance regarding texts that might help in this regard. It is tempting to ask whether future texts in linguistics might make more room for the types of more nuanced consideration of race and ethnicity that seem to be centered in texts that emphasize ethics, empowerment, researcher positionality, the role of identity, power relations and intersectionality in language form and variation. The remaining sections of this white paper are intended to offer additional recommendations to this end.

2. Research question design

2.1 How do linguists go about asking questions? Inductive v. deductive reasoning in linguistics research design

As the previous section demonstrates, many linguists think of their primary task as measuring languages or language structures (Kiesling 2011:3-6, Milroy 1987: 3-5, Kibrik, 1977). For others, the principal task is studying the interrelationship between language structure and social structure. This is an important point, because the former approach places the analyst’s focus and scholarly rigor on representing orderly, abstract structures, rather than on studying human behaviors. We should be aware of how our particular discipline within linguistics informs how we think about the relationship between the investigator and the speech community. Language structures are objects of study in both approaches, but crucially, the path to conceptualizing how we represent individuals or groups is different in the two. In some approaches, thinking about how, and how well, we represent people and groups is not a central part of the process of study design. To incorporate such considerations, for example, if we wish to address Hale’s concern in the previous section that the linguists conduct their work with a view to their responsibility to communities, it will be necessary to work in a way that seems tangential to the main task. In other approaches, representation and ethical treatment of people is centered in study design and the path to incorporating considerations about justly representing people is much clearer.

It may be instructive to consider that some subfields of linguistics go about investigating their objects using deductive methods, while others use inductive ones. *Deductive* approaches begin with general principles that represent a current, best understanding (or theory) of our linguistic object of study. We then gather new specific cases whose fit into our theory is tested.

Our research questions often are formulated “if a proposition is true for some structure A, then it is true for another B with the same critical property.” We seek to predict the patterning of new structures that share the same critical linguistic property. Theory, in deductive reasoning, is built from the general down to the specific. If the theory does not hold as new data is introduced, it is revised or refined. It is easy to see that the focus in such approaches remains firmly fixed on the structures involved in linguistic relations.

Inductive reasoning takes essentially the opposite approach. We begin with a database of observations of many individual cases of linguistic forms or features from a carefully chosen community, and once we reach a satisfactory level of these forms (token counts), we make inferences about how those cases are distributed in that community. We then extend our model to test the generality of these inferences to new data drawn from other communities, until we are able to state general principles about what critical features best characterize their occurrence. As these principles grow stronger, we then make deductions (formulate our theory). The process iterates as we select a new community, bring in more data to continue to extend our principles in generality, allowing us to make general statements about the social or stylistic levels at which variation occurs, and about the routes, mechanisms and causes of linguistic change (Labov, 1994:4). Research questions under inductive approaches, then, can center the notion of human community as they build generalizations about language structures.

Below are examples of research questions demonstrating these two approaches:

Inductive approach: Variationist Sociolinguistic study. “What is the association between age and (r)-deletion in the speech of working class New Hampshireites?”

We might say that many variationist research questions typically take the form of a wh-question to test the association between a social fact and a linguistic one. A strength of such approaches is that they allow room for consideration of race and ethnicity in constituting an appropriate level of social structure for study, and in the consideration of how to sample that community.

Deductive approach: Syntactic study designed under a Parametric view from (Minimalist/Principles & Parameters): “What features are on the AspP in AAE that explain habitual be, aspectual adverb stay verb-ing, and stative zero copulas?”

This research question assumes that there is some stable dialect that can be differentiated from others based on the presence or absence of certain parameters in the grammar, theorized sometimes as features on functional heads in a tree but also in other ways. This question does not on its surface allow for asking about why that premise is questionable considering wide multilectalism, or between-community variation.

2.2 What makes a good research question?

Often, the type of research question driving an investigation will determine how we treat macrosocial variables (such as racial or ethnic identity) in linguistic study. Without careful examination of relevant identities and labels used by community members within a given community (see Section 5), a research question may inappropriately represent the speakers that

are being studied. Charity Hudley (2017) explains how antiquated theoretical frameworks of race and ethnicity lead researchers to broadly categorize speakers during research question formulation. Identifying speech communities (and by extension ethnic groups) by how they sound or by their linguistic or cultural traits may result in circular, inaccurate, even harmful portrayals of certain groups of people. For example, Newman and Wu (2011) suppose only mutually exclusive identities in their research question: ‘are [judges of different ethnic backgrounds] able to distinguish Korean from Chinese Americans?’ The *a priori*, static conceptions of racial identities in this question impedes any interpolation of nuance into research design. Moreover, the study essentializes the Chinese and Korean American communities with only two speakers from each category. Opposing these conceptualizations are research questions that formally include stakeholders situated within the community being studied. For example, Johnson and Johnson (2015) studied perceptions of bilingualism in Latinx middle school students, and used educational consultants employed at various middle schools as principal interviewers of Latinx students. The inclusion of these stakeholders allowed for research questions to be developed through discussions with consultants as to the most important aspects of ethnicity and bilingualism policy for the students being interviewed.

3. Methods of collecting data

This part of the white paper focuses on methods of collecting demographic data used in a wide range of quantitative research studies, including multiple choice, open-ended (free response questions) and then interview-style prompts. We address issues that highlight the challenges with closed-response question types and recommend ways to address them. We note, however, that some of the thornier issues, pertaining to use of questions that do not easily allow detailed, subjective, or complex self-identifications, and intersectionality need to be treated with care, using the techniques recommended below in Section 4. The section closes with recommendations regarding best practices using these techniques.

3.1 Method: demographics collected through multiple-choice survey

Collecting demographic information through multiple-choice survey or questionnaire questions is perhaps the most common method by which researchers in the quantitative subfields of linguistics collect information about participant ethnicity. Categories may be established by prior ethnographic fieldwork in the community. Such time can be explicitly used to learn about groupings salient within the community (Heller et al. 2018).

3.1.1 Strengths of multiple-choice surveys

There are many strengths to using a multiple-choice survey. First, it can be used in large scale studies and smaller scale studies. Second, although the workload can be heavier initially due to the increased effort the researcher must devote to establishing relevant categories, analysis is made easier later by having already established categories. Third, since participants are provided with already existing categories, participants might feel less confused as to how to answer the question. Fourth, if the researcher also allows participants to name other categories if they wish

to do so, this method provides some of the benefits described using free responses as described in Section 3.2.

3.1.2 Weaknesses of multiple-choice surveys

There are weaknesses to multiple choice surveys. The researcher may choose labels that are irrelevant in the community, which may result in higher nonresponse rates from participants and which may result in the demographic categories not representing the community well. The researcher may need to devote a significant amount of time to researching which labels are used in a particular community, potentially resulting in more time dedicated to study design. Additionally, there may be issues with how to categorize participants who pick the “Other” option, and depending on the study size, and number of people who chose that option, this could potentially lead to a lot of work and introduce issues discussed in Section 3.2. If choosing this method, keep in mind issues described in Section 4.2.4, as well.

An obvious first choice that researchers in the United States might consider as a model for categories to use in multiple-choice questions for surveys and questionnaires is the US Census. As mentioned in the introduction, the race and ethnicity questions were revised in 2020. There are five “race” categories from which to choose, including: ‘White,’ ‘Black or African American,’ ‘American Indian or Alaska Native,’ ‘Asian,’ and ‘Native Hawaiian or Other Pacific Islander.’ Participants may check more than one box if desired, and optionally write in additional ancestry information (see Section 4.2.4). For ‘Asian,’ separate checkboxes allow selection of a select number of alternatives, including: ‘Chinese,’ ‘Filipino,’ ‘Asian Indian,’ ‘Vietnamese,’ ‘Korean,’ ‘Japanese,’ and ‘Other Asian.’ For ‘Native Hawaiian and Pacific Islander,’ choices include: ‘Native Hawaiian,’ ‘Samoan,’ ‘Chamorro,’ and ‘Other Pacific Islander.’ Aside from these questions, a separate prompt asks whether the person is of Hispanic, Latino, or Spanish origin. In considering whether these questions provide suitable models for survey design, several points are worth noting. First, for historic reasons, the Census categories are designed to allow trends to be compared to earlier surveys, and “useful for statistical analysis” (Office of Management and Budget, 1997). This means that categories are slow to change and may not reflect current practice in the social sciences, or the increasing diversity of the nation. And, as mentioned above, they may not fit a researcher’s community well. Second, several changes recommended to the Office of Management and Budget by the Census bureau were not adopted in time for the revised census and are not reflected in the current categories. These include a new ‘MENA’ category--Middle East and North American, and reassignment of ‘Hispanic, Latino or Spanish origin’ to a racial category (Jurjevich, 2020). Third, it is perhaps apparent that the census uses the term *race*. The only *ethnicity* included in the census is “Hispanic, Latino or Spanish.” Clearly, this usage of the terms *race* and *ethnicity* does not align with the current definitional practices outlined in Section 1. Again, it is important, if choosing this method, to keep these issues in mind.

3.1.3 Example of a study which uses a multiple-choice survey

A strong example of utilizing a multiple-choice question to obtain information on ethnicity is Conrod (2019), whose research focused on pronouns. In particular, Conrod (2019) was interested in singular *they*, and they created an online sentence acceptability task to obtain information

about the acceptability of singular *they*. Participants were asked to complete a short demographic questionnaire at the end of the task and were asked to choose a label from a predetermined list that best describes their own ethnicity. The choices came from the US Census, as enumerated above in Section 3.1.1. There was also an option for participants to choose ‘Other’ and name a different label, and participants could pick more than one label. Additionally, participants could opt not to answer the question. Although there are some drawbacks to using this approach as discussed in Section 4.2, this method was sufficient for the purposes of Conrod (2019). Ultimately, their research focus was on pronouns, but by adding this question to their demographic survey, the researcher could see that their sample consisted of mostly White people. This information regarding ethnicity in Conrod (2019) could help inform future research regarding singular *they*, given that this sample was unbalanced for ethnicity. The strengths of this approach in Conrod (2019) include allowing participants to choose multiple options, allowing participants to not answer the question, and including another option. The weaknesses mostly involve issues with using Census categories and issues with using predetermined categories that will be discussed in Section 4.2.

3.1.4 Summary and Recommendations

In summary, collecting demographic information through multiple-choice surveys has both strengths and weaknesses. Multiple choice surveys can make analysis easier for researchers, and the predetermined categories might make it easier for participants to understand what the researcher means by ethnicity. The weaknesses involve issues regarding the relevance of the labels in the community of study. Our recommendations include that researchers:

1. make an effort to understand which labels might be relevant in the community of interest.
2. allow participants to not answer questions.
3. allow participants to choose multiple options.
4. include an option where participants can name one or more labels not already included which are relevant to them.

An interesting aspect of multiple-choice surveys in sociolinguistics, is that they seem to be seldomly used on their own for studying ethnicity. One reason for that may be that ethnicity is infrequently included in demographic surveys. When a researcher does include ethnicity, the researcher tends to ask more questions than just a multiple-choice question. In end, much like in Conrod (2019), if one’s research is not focused on ethnicity, it is still important to ask information about ethnicity in order to understand who was included and who was not included in one’s study and because it can play a significant role in understanding language variation; in addition, it is important to prioritize your participant’s well-being in mind while asking these questions (see Sections 3.4.3, 3.7, and 4). Due to the issues regarding multiple-choice questions, it may be best to include a free-response question in addition to a multiple-choice question. This allows for the researcher to obtain more information which could better inform future research in the community and which could inform future multiple-choice questions.

3.2 Method: demographics collected through free-response written or online survey

Collecting demographic information through free response, written or online, is another method by which researchers might collect information regarding ethnicity. The strengths to this method relate to topics that will be discussed in Section 4 more extensively (e.g. fluidity and multifaceted nature of self-identifications, Section 4.2, intersectionality, Section 4.3.1.). One noteworthy advantage is that using this method helps avoid some issues regarding researcher-imposed labels, or categorizations, on participants because the participants provide the labels themselves. Labels are an important issue in variationist sociolinguistics because many studies rely on categories or labels for sociolinguistic analysis (e.g., Census-based categories, see Sections 3.1.2, 4). It can be challenging to choose the best label for a questionnaire. For example, Wolfram & Schilling (2016) describe the extensive variation in identity labels associated with Latinx-identifying people when discussing ethnicity and language variation in the United States. Latinx are not a monolithic group, and as described in Wolfram & Schilling (2016) the differences in labels used, “reflects the diversity of situations involving speakers, social situations, and political conditions” (p. 188). Indeed, Afro-indigenous communities of the Garifuna living in New York must negotiate between ethnic self-identities of “Black” and “Hispanic”, which are further mutable depending on geographic locale (see England 2010 for a discussion on the sociopolitical implications of these multifaceted identities in America). The variety of self-identifying terms that participants use can vary within one community. For example, in research conducted on Chicano English in Los Angeles, Fought (2003) encountered terms for ethnicity such as: ‘Mexican-American,’ ‘Chicano,’ ‘Latino,’ and ‘Mexican.’ Additionally, in the community, speakers had different understandings of what these terms meant. For example, some of the speakers, “had mixed feelings about ‘Chicano,’ which was sometimes associated with radical politics, or surprisingly even with gang members” (Fought 2003, p. 17). However, “others thought ‘Chicano’ was a neutral term, and to some it had an important and positive ethnic significance” (Fought 2003, p. 17).

3.2.1 Strengths of the free-response method

The free-response method also has the potential to make participants feel more represented, because forced choice demographic questions might not accurately represent the person’s identity and could result in not completing a survey leading to higher nonresponse rates. The issue of nonresponse rates and how to reduce them is discussed in great detail by Dilman, Smyth, and Christian (2014). One survey tactic that Dilman et al. (2014) identifies as being problematic is requiring participants to answer every question in a survey. Given the situation a participant might think: “none of these answer categories fit me; I don’t know what to do. Should I quit or just make something up?” (p. 20). This issue is more salient when including participants from groups who do not fit perfectly into more traditional demographic categories. For example, the Garifuna An additional benefit for researchers is that this is a method that has been used before in sociolinguistic research such as by Cheng, Faytak, and Cychosz (2016), D’Onofrio (2019), and Hilton (2018). Another benefit for researchers is that the inclusion of questions can potentially add minimal time to the study design compared to other methods. In other words, asking a question such as, “how would you describe your ethnicity?” is potentially less labor intensive for the researcher than a similar multiple-choice question where one would need to understand which terms are relevant in community.

3.2.2 Weakness of the free-response method

Researchers considering employing free-response questions might have one of several types of concern. Researchers may receive a large variety of responses that resist being distilled into a small set of groups, making it challenging to aggregate participants for linguistic analysis. This issue is particularly acute for larger, more quantitative samples in study designs that prioritize the need for groupings to be systematic, and the need for groups to be justifiable to other researchers and members of the community that is being studied. Aggregation under such circumstances does mean an increased workload. Additionally, participants might not know how to answer the question without some examples because they might not know what the researcher means by ethnicity. If participants do not know how to answer a question, they might quit or fabricate a response, neither of which is a good outcome for the researcher. Free response also opens the potential for harm in disclosing personally identifying information, and should be treated with care (see Section 3.7). Also, grouping participants could become challenging if some participants choose very broad categories and others choose very narrow categories, which has the potential to delay and/or impede accurate data analysis. Another significant concern is that if the researcher chooses to group together participants who used different labels, they could potentially racialize their respondents by assuming that a label that was not reported is relevant to that participant. For example, a participant who wrote down ‘Polish American’ might also consider the term ‘White’ to be relevant to them if the researcher chooses to place them in such a category. Conversely, another Polish American respondent might feel unhappy to be categorized as ‘White’ and might feel like that category is irrelevant to them. The researcher would not know unless the participant is asked, so it might be better to ask more questions, so that the researcher is not assuming labels. This would also increase the validity of groups that researchers are using as well. Furthermore, the participant might have a very different understanding of ethnicity than the researcher, resulting in the introduction of some bias by the researcher for asking questions that perhaps do not capture accurately the participants identity and lived experience. The potential for mismatch between the understanding of ethnicity between researcher and participants might mean that having a truly free-response question without any examples or explanations from the researcher might be not possible (see Sections 1 and 4.2 for reference).

3.2.3 Example of study which uses a free-response question

An example of utilizing a free-response question to obtain information on ethnicity is Cheng et. al. (2016). One particular strength of Cheng et. al. (2016), is the detail provided regarding how they decided to group participants. In this study, the researchers were interested in analyzing the relationship between ethnicity and participation in a language change phenomenon often known as the California Vowel Shift (CVS). They used an existing corpus, the Voices of Berkeley for their analysis. The corpus included 786 participants. In this corpus, ethnicity was elicited with an open-ended question, labeled by the researchers as the Self-Determined Ethnicity (SDE) question. From the 786 participants, the researchers identified 535 as being from California and having provided enough demographic information. The researchers categorized the participants into twelve groups: White, Latino, Chinese, Korean, Middle Eastern, South Asian, Vietnamese, Filipino, Hapa, Black, Japanese, and Native American. These groups were, “based on their response and the experimenters’ reasoning” (Cheng et al. 2016, p. 65). Cheng et al. (2016) provide some examples of the categorization process, such as:

Participants who reported ‘White’, ‘Caucasian’, or a combination of European heritages such as ‘Irish/German’ were categorized as ‘White’. ‘Chinese’ and ‘Chinese American’ were both considered ‘Chinese’—keeping in mind that all participants were Californian. Participants who identified as mixed-race Asian and European were combined into a ‘Hapa’ SDE group. (p.65)

Some of the 535 participants were not included because of responses that were too broad for the researchers’ purposes, representing nonspecific ‘multiracial’ responses. Ultimately, the study included data from 506 participants grouped into these categories and found evidence that ethnicity is a significant predictor “of how an individual’s vowel space will adhere to previously-identified CVS patterns” (Cheng et al. 2016:76). Some criticisms of this work include the previously discussed issues with grouping people at all. If a participant writes that they are ‘Irish/German,’ wouldn’t that label be more relevant to that person’s identity than the label ‘White?’ Does this person ever identify themselves as ‘White?’ Additionally, not including multiracial responses may have made statistical analysis easier but also ignores a growing number of people who identify as ‘multiracial’ (Pew Research Center, 2015). To the researchers’ credit, they note that, “‘ethnicity’ for the purposes of sociolinguistic inquiry is more complex than the traditional ethnic categorizations we are all accustomed to” and that, “‘Chinese’ as one big ethnic group is neither accurate nor ideal” (Cheng et. al. 2016: 76)

3.2.3 Summary and recommendations

Although there are many strengths to the free-response method, there are many weaknesses as well. Adding a short free-response question, adds minimal time to study design, but has the potential to be challenging to analyze later. This method is more likely to provide information on categories which are relevant in the community, which has the potential to inform studies by researchers who wish to use a multiple-choice survey method. As mentioned above, grouping people using terms that were not explicitly identified by participants runs the risk of racialization. However, many sociolinguistic studies rely on large groups for statistical analysis in order to make broad claims regarding language. If using this method, our recommendations are:

1. Consider whether asking for this information has the potential to harm participants (see Sections 3.4.3, 3.7, and 4)
2. Consider whether this is the best method based on your study design due to the relatively low effort needed by the researcher to add the question, “how to describe your ethnicity” to their demographic survey.
3. Allow participants to not answer the question.
4. Discuss the categorization schema used with respondents, using many examples.
5. Address concerns regarding racialization and ethnicization of participants in research papers
6. Understand the limitations of category labels (see Section 4.2)
7. Remember when aggregating participants, to overtly describe what your labels mean (see Section 3.2.3) and to describe how you made those methodological choices and the limitations of those choices.

If grouping respondents by ethnicity is necessary in a study, it is recommended that the free-response not be used alone but rather in conjunction with another method, such as a multiple-choice questionnaire. Using these two methods together allows for researchers to conduct a quantitative analysis while also gaining both an understanding of which labels are relevant to each participant individually and to the community as a whole.

3.3 Method: demographics collected through interactive interview

The interactive interview, which traces its roots to traditional, Labovian variationist analysis, offers us a vehicle through which to fully explore the complexities of research participants' identities. Done correctly, it supports our goals of appropriately and ethically gathering demographic data during linguistic research. However, we acknowledge that this method can be complicated by multiple factors such as identity co-construction or the sometimes taboo notion of discussing ethnic self identity. Its effectiveness will therefore depend on the individual study in question and its implementation. As such, we refrain from specifying a single 'right way' to solicit these data from research participants during an interactive interview, instead supplying considerations which we hope will generalize broadly to many types of study.

The primary reason for *interviewing* participants for these data is to deeply understand an identity's impact on the linguistic variable in question. Berg (2007:96) states that such interviews allow participants to "speak in their own voice and express their own thoughts and feelings." Indeed, participants can report their ethnic and racial self-identities however they choose, and according to their current membership. As race is a social construction (and ethnicity is often conflated with race), an interview scenario also allows participants to ask for clarification from researchers if they feel that their internal models of ethnic identity are incongruent with the questions we may ask. Obviously, allowing participants to report their own identities during the interview (instead of racializing them *a priori*) is crucial.

We must be aware of the interaction between (perceived) researcher identity and participant responses to questions regarding self identity. Adida et al (2016) details how interactive interviews constitute a social interaction, and how in African cultures these perceptions can influence survey responses. In their analysis spanning 14 different African countries, responses from survey participants differed markedly when ethnic identities of the interviewer and interviewee were *shared* versus when they were *different*. Likewise, Rockquemore and Arend (2002:55) find that biracial individuals can understand their racial identity as "Sometimes Black, sometimes White...depending on the social context," e.g. with researchers presenting different ethnic identities. We therefore must acknowledge that our own ethnic and racial identities can influence our participants' responses to questions about their own identities during face-to-face interviews. Understanding this allows us to craft appropriate questions to ask our participants. We reiterate that there is no single right set of questions for this task, only that it is critical to deeply understand why and how linguists approach their question formulation.

As alluded to above, discussing these topics is often taboo in the United States (Paulker et al. 2015). Therefore, participants' willingness to discuss intimate topics such as their own ethnic or identity experiences often varies greatly, especially in-person and with an unknown researcher. Cukor-Avila and Bailey (2001) present evidence that the familiarity between interviewer and interviewee matters enormously to the volume of data gathered. They highlight

stylistic differences between a participant's responses during beginning and final interviews, with the latter offering incredibly rich personal information.

In certain circumstances, several changes may be needed in the study design and procedures in order to face emergent ethical concerns. A good example showing how to handle unexpected and difficult situations is illustrated in Section 5, which presents a decision to alter a linguistic prompt in the Pacific Northwest English Study materials borne of the need to show respect for a speech community (Wassink, 2015, PNWE hereafter).

Our recommendations:

1. Acknowledging (if not discussing) the identities of the researchers participating in the interactive interview
2. Give participants ample space to self-identify as they feel able to during interviews
3. Do NOT assume participants' identities *a priori*
4. Remain flexible in case participants' identities change throughout the research process (and beyond)
5. Use proxies for ethnic identity such as family history or neighborhood segregation if participants show reticence and demographic data is truly requisite to the experiment (Charity Hudley 2017)

3.4 Method: perception experiments

While much of this white paper is focused on language production research, our recommendations are also relevant to studies of language perception. One of the challenges of conducting language perception experiments is the high potential for confounding factors, and as described in Section 1.1, race and ethnicity are often confused for each other or conflated entirely. Nevertheless, the authors especially want to highlight that race and/or ethnicity data should be collected for reasons described in Sections 3.1, 3.2, 4.2, 4.3. Although we want to encourage the study of race and ethnicity in linguistic research, we do acknowledge that responsibly including race and ethnicity in aspects of perception research imposes an additional burden on the researcher and participants. Our focus in this section is on the future improvement of research design, rather than a critique of previous research designs. Here, we aim to provide a broad overview of methodological considerations unique to perceptual studies, particularly those in sociolinguistics, while focusing on the issues of confounding factors, generalizability, and balancing the risk and benefits of perception studies involving race and/or ethnicity.

Perception research can include investigations of the role of the perceived race of a speaker and subsequent perception of the language produced by that speaker. Such research, when taking a critical approach towards race, has the potential to contribute to greater understanding of the mechanisms and motivations underlying bias against racialized language varieties and their users (Baugh 2017; Purnell, Idsardi & Baugh 1999). Nevertheless, this line of inquiry must take caution in balancing these potential gains with the potential harm that may arise through study designs that can reinforce race essentialism in the eyes of their participants (see Sections 3.4.3 and 4.2).

3.4.1 Types of perception studies in sociolinguistics

Perception studies conducted by sociolinguists can be said to have one of two broad objectives:

1. To better understand how language varieties activate beliefs about the social groups who are believed to use these varieties, and how activation of these beliefs affects human behavior
2. To better understand how actual or perceived differences between language varieties affect language processing

Generally speaking, the methods and assumptions underlying studies with the former objective are similar in approach to research in social psychology, while methods and assumptions underlying studies with the latter objective more closely resemble research in cognitive psychology. This is an important distinction because these fields of inquiry have historically differed in terms of assumptions and methodologies. Examples of both kinds of linguistic studies are illustrated in Table 1.

Table 1: Social Beliefs and Language Processing Study Examples

Study type	Research Paradigm	Example Study	Research question
Social Beliefs Study	Linguistic Profiling	Purnell, Isardi, & Baugh (1999*)	Do speakers suffer discrimination based on a listener's association of the phonetic qualities of their voice with a racialized language variety?
	Linguistic Stereotyping	Lambert, Hodgson, Gardner, & Fillenbaum (1960)	Does the language a person speaks determine how that speaker will be evaluated by listeners?
	Regional Grouping and Identification of Speakers	Plichta & Preston (2005:107)	"Can speakers use degree of a continuous feature in variety identification?"
	Folk Linguistics	Niedzielski & Preston (2010)	What are non-linguists' attitudes towards language variation?
	Reverse Linguistic Stereotyping (RLS)	Kang & Rubin (2009:441*)	"[Do] attributions of a speaker's group membership trigger distorted evaluations of that person's speech[?]"

Language processing	Cross-Dialectal Comprehension	Labov & Ash (1997)	How does the perception of speech of certain dialects vary based on listener dialect?
	Expectations-based	Hay & Drager (2010)	Can indirect primes of national identity shift perceptual boundaries?

*Indicates a study that explicitly investigates race or ethnicity

3.4.2 Confounding factors and generalizability

Researchers who do experimental work in perception are rightly concerned both about generalizability of their results and about potential confounding factors. These two reasons are part of why contemporary perception researchers conventionally provide at least some information about the participants in their studies, most often including prompts eliciting age and gender information. It is also common for perception researchers to include participant age and gender as factors in their analysis, either by subsetting data for descriptive presentation or by including them as variables in a statistical model, in order to understand and quantify their potential influence. Given that racial disparities in US society with respect to criminal justice, education, housing, banking, health care, and social welfare are well-documented (Lee, Esposito, Edwards, Chun & Grinstein-Weiss, 2020), the authors make the following recommendations for the purpose of ensuring that the knowledge learned in a study is applicable to members of non-majority racial groups:

1. We recommend that researchers collect and report participant race and/or ethnicity data in addition to participant age and gender data. We recommend that they do so following the guidance offered in Sections 3.1, 3.2, 3.3, and 4.2.
2. We encourage perception researchers to more broadly consider whether race and/or ethnicity could play a more central role in their research question.
3. We ask researchers to consider if their experimental designs are continuing to uphold the idea that the “default” person in a given society is a member of that society’s majority race (see Purdie-Vaughns & Eibach 2008).
4. We recommend that researchers who opt to exclude race and/or ethnicity from their experimental design overtly discuss their decision to do so.

The authors believe that inclusion and overt discussion of race and ethnicity in research design may bring into focus gaps in experimental work and may lead to a more diverse canon of work that is more representative of all individuals in society. If a researcher is including human subjects, it should be assumed that race and ethnicity are potential confounding factors, and we believe that strong, experimental research not only discusses all likely potential confounding factors but also is transparent about the limitations of its generalizability.

3.4.3 Balancing risks and benefits in perception studies

Some scholars try to answer their research questions by artificially manipulating the race or ethnicity of a hypothetical person or group (e.g. Squizzero 2020). In doing so, these experiments could be designed in a way that risks validating or promoting essentialist beliefs to participants. To provide an example, in a study that offers empirical support for the idea of *reverse linguistic stereotyping*, i.e. the process by which a speaker's perceived group membership influences perception of their language proficiency or style, Kang & Rubin (2009) asked respondents to listen to portions of a recorded lecture while projecting a photograph of an Asian face or a Caucasian face, ostensibly the face of the speaker, and asked respondents to evaluate the degree of speaker accentedness. By manipulating the race of a hypothetical speaker - a method used to evoke stereotypes associated with certain phenotypes or phenotypical cues (cf. Sen & Wasow 2016) - and then asking about accentedness, participants might inaccurately infer that a speaker's race can be a valid, reliable predictor of accentedness. However, we first note that the potential harm of this form of experimental design can be mitigated by debriefing participants at the conclusion of the experiment, which is common practice for studies that involve deception of participants. We also note that this study laid the groundwork for the authors' future work in developing and implementing an intervention that successfully improved ratings of comprehensibility and instructional efficacy of non-native English-speaking international teaching assistants by challenging biases held by undergraduate students (Kang, Rubin & Lindemann 2015).

In sum, we recommend that researchers carefully consider how participants might interpret their experimental methods and materials, take steps to mitigate potential harm, and balance the risks and benefits of their research. We also recommend that perception researchers note their own positionality as analysts and architects of their studies (see Section 4.1). We echo the concluding statement of Section 1.1 that linguists should never refer to biological dispositions towards language use, including in perception studies.

3.5 Method: demographic info was not collected

Some linguistics studies do not collect any demographic information, and some collect demographics but omit race and ethnicity. This category includes various kinds of studies:

1. Grammatical intuitions of the researcher are reported and demographics are not collected
2. Experimental studies (ERP, eye-tracking) sometimes collect/report age and gender/sex but not race/ethnicity
3. Some demographic details (including race/ethnicity) are sometimes strategically omitted for purposes of protecting the anonymity or safety of participants
4. Rapid anonymous surveys sometimes only can include demographic information that is researcher-imposed

These types of studies have different underlying reasons for failing to collect demographic data, and data about race and ethnicity in particular. Some of these underlying reasons include the inability to accurately collect the data (as in rapid anonymous surveys) or considerations of preserving anonymity and safety for human subjects. However, in other methodologies, it is uncommon to discuss race or ethnicity at all; this section will focus briefly on the first two methods listed above and explain why it may be important for researchers in those areas to adjust current practices.

3.5.1 Grammatical intuitions and the source of example sentences

Grammatical intuitions of the researcher or of unnamed informants are commonly used as the primary data type in formalist linguistic research, particularly syntax and semantics research. There has been some empirical validation of the reliability of these reports (e.g. Mahowald, Graff, Hartman, and Gibson 2016), but most such papers do not focus directly on social aspects of language at all. Indeed, it is often an un-discussed assumption that the intuitions of the author (or their informants) reflect a mostly unified stable state of the grammar in a language-using population. As an example of an extremely typical syntax paper, Hsu and Syed (2020) discuss the cross-linguistic state of the DP hypothesis; examples that are not cited to other sources are (presumed) to reflect the judgments of the authors or unnamed informants. This paper was randomly chosen as an example because it is very recent, very typical of papers in theoretical syntax, and is published in the flagship journal of the LSA. Presumably neither the authors nor any editors nor reviewers would ask whether judgments (of * or ?, notations which denote ill-formedness) are in fact typical of all speakers of any given language, or whether any of the reported constructions should vary by social identity.

The reason such methods are worth discussing in a paper about race and ethnicity in linguistics is because, if those aspects of language use are unexamined, it is possible to either overlook or misreport the state of the grammar of interest. While the data reported in Hsu and Syed (2020) are not known to vary sociolinguistically, they are also not empirically proven *not* to vary. Because the authors do not give any positional statement of their own language backgrounds, a reader cannot know whether the judgments in the Mandarin data are reflective of all Mandarin speakers, or whether the authors' judgments should be taken to represent a particular subset (such as social or regional dialects).

Other recent formal work (also sampled from the 2020 volume of *Language*) approach these issues with varying degrees of detail and specificity. Of notable detail is Legate, Akkuş, Šereikaitė, and Ringe (2020), who include details about their Turkish informants in a note in the paper:

We had ten primary Turkish consultants, ranging in age from mid-twenties to early forties, four from Bitlis, and one each from Adıyaman, Bursa, Denizli, Hatay, Isparta, and Mersin. We had two additional Turkish consultants, in their thirties, from Bitlis and İzmir, whose grammar is systematically different; we discuss their grammar when relevant. We had eight Lithuanian consultants, five in their late twenties, three in their late thirties to forties, six from Vilnius and Kaunas, two from Šiauliai. (Legate et al. 2020:771)

This detail can be contrasted with Hsu and Syed (2020), discussed above, and with Davis, Smith, Super, Super Sr., Thom Sr., and Mikkelsen (2020) in the same volume; the latter authors thank their informants and clarify that their consultants retain copyright to the material, but do not provide further sociolinguistic details about either the authors or the consultants. It is noteworthy that the three papers discussed in this section are published in the same year, in the same journal, because this demonstrates that there is not a standard expectation for acknowledgement or positionality practices in the field, even for a single journal.

We once again emphasize that no papers we review here were selected as “bad examples,” but rather are reflective of the norms and standards of the subdiscipline within which

they are located. Instead, take our discussion as critical of those norms of the subdiscipline, which reflect no fault of individual authors or researchers but which require a field-wide paradigm shift.

One important, and relatively minor, change that researchers in fields like theoretical syntax can make is to include a statement about the positionality of the authors and the sources for all grammaticality judgments as a part of the introduction of any paper. We provide an example here, written by a syntactician on our team for an in-progress manuscript:

All data that are not otherwise cited are reflections of the author's own grammaticality judgments. Examples which are adapted from attested forms are marked to indicate this with a °, with context given when possible. The author is a monolingual White American nonbinary person who grew up in California and has lived in Taiwan and Washington state (U.S.) as an adult. Examples which are known to reflect sociolinguistically active constructions are marked with a %.
(Conrod, unpublished manuscript)

Such positionality statements go a long way towards de-mystifying syntactic data; readers and reviewers are much more able to identify where the author might be unaware of existing relevant sociolinguistic variation. Including the positionality of the author also de-normalizes the presumed 'default' assumption that all authors writing on a language are reporting on a prestige or standardized variety. Without these statements, syntax articles on (for example) English are presumed to be discussing White, mainstream American English unless explicitly stated otherwise. Indeed, it is unlikely that a syntax paper discussing AAE syntax would "get away" with presenting examples without any comment on social identity whatsoever. This asymmetry is indicative of an entrenched assumption of whiteness as default that is endemic in formal fields and bringing attention to it in our writing is a basic and necessary first step to fixing that (see Charity Hudley et al. 2020 for discussion).

Strategies that we have suggested in this section include:

1. Including written acknowledgement in research publications in formal linguistics that describe the demographic makeup of language informants or consultants
2. Explicitly stating the sociolinguistic background of authors (or informants) whose grammaticality judgments are represented in the publication
3. Explicitly describing how grammaticality judgments were collected, and including discussion of whether syntactic constructions were investigated for evidence of sociolinguistic variation (even if the answer is negative)

Other strategies that researchers in formal syntax and related fields may consider include:

1. Intentionally and explicitly seeking out informants of the language of interest with significantly different racial or ethnic identities from the researcher
2. Including small-N surveys, as in Mahowald et al. (2016) to validate judgment data
3. Intentionally and explicitly including citations of literature which report on similar syntactic phenomena from a sociolinguistic perspective

3.5.2 Experimental studies

The inclusion (or not) of participant demographic data in experimental linguistics (which broadly spans fields not called linguistics by name -- cognitive science, as well as psycholinguistics and neurolinguistics) is more mixed, but we still observe that many works in these fields either do not collect or do not report data on race and ethnicity.

As an example (again selected randomly for recency, typicality for its subfield, and publication in LSA's flagship journal *Language*), Kwon, Kluender, Kutas, and Polinsky (2013) present ERP (event-related brain potential) measurements to test the processing cost of different types of relative clauses. The paper reports the participant demographics thus:

Twenty-two native Korean speakers were paid \$10/hour for their participation in the ERP study (fourteen females, eight males). At the time of the experiment, all participants were between the ages of twenty-two and thirty-one (mean age: twenty-five) and were enrolled in graduate school or in English classes at UCSD Extension. The average length of stay in the US was thirteen months (range of length of stay: two months to 3.5 years). All participants were right-handed with no neurological disorders and normal or corrected-to-normal vision. (Kwon et al. 2013:552)

No further information about the participants is given; the participants' social identity, including ethnic and racial identity, is not discussed. While it may be the case that this data was collected and then not reported (due presumably to irrelevance), we suggest that it may be preferable to report all collected demographic information (even in a footnote) rather than leave social identities un-reported.

A failure to collect information on social identity or demographic information about participants can lead to significant oversights in research; likewise, the inclusion of this information can be a route towards better representation in the scientific literature. By not reporting ethnicity or racial identity whatsoever, experimental linguists may have inadvertently glossed over important biases in the physical technology used to collect this type of data. Refer, for example, to recent innovations in EEG technology that were necessary for equipment to accurately record readings from participants with natural hair textures (Etienne, Laroia, Weigle, Afelin, Kelly, Krishnan, and Grover 2020). If linguists working with EEG and other technology had tracked and reported race and ethnicity, it would have been significantly easier to spot biases like a relative lack of Black participants across many studies. Without this data, the biases are harder to report and harder to correct.

Our recommendations for researchers carrying out this type of research include:

1. Incorporating already-existing tools for collecting demographic data. Refer to Sections 3.1-3.3 for more details on these tools and how to use them.
2. Include race and ethnicity in the data you collect. Note that reporting the language spoken by participants does not equate to reporting this information.

3.6 Method: Computational methods and large datasets

Computational methods in linguistics allow linguists to scale up the size of the datasets they are working with and therefore test linguistic hypotheses more rigorously (Bender & Langendoen 2010). Especially when working with datasets collected from social media (e.g. Twitter, Reddit, and so on) and exploring hypotheses which involve connections between language and race/ethnicity and/or racialized language varieties, researchers may find themselves in the position of needing information about race from the speakers whose language appears in their sample. In addition, as we strive to ensure that language technology works for varieties beyond mainstream, standardized, prestige dialects it is important to construct datasets to train and test automated systems which represent a broad range of racialized varieties (Jørgensen et al 2015, Tatman 2017, Sap et al 2019, Koenecke et al 2020, Wassink et al under review). In that case as well, researchers sometimes find themselves in the position of wanting to be able to attribute racialized identities post hoc to speakers represented in their dataset.

Faced with a large dataset lacking labels for an attribute of interest, a typical computational approach would be to train a classifier based on a smaller dataset where the labels are known (here racialized identities) to then ‘predict’ the labels in the larger dataset. The features used by the classifier might be linguistic features (e.g. lexical items, spelling variations, inflectional patterns), metadata (e.g. names from user profiles), or social network information (i.e. information about the user's connections in the social network, their patterns of ‘liking’, or other similar behavioral information).

Creating a classifier can be harmful to the people it is used to classify and to the goals of the research project for several reasons:

1. It represents a case of racializing others, rather than allowing for self-identification.
2. This same classifier could be used beyond the purpose of labeling the dataset at hand, extending the harms of racialization.
3. The classifier will have some degree of inaccuracy when compared to the ground truth of the speaker’s own identities and this inaccuracy will impair the researcher’s ability to draw conclusions based on the dataset.

Alternative, better approaches to take in these contexts include:

1. Being transparent about the lack of information regarding racialized identities in large datasets.
2. Contextualizing large datasets based on relevant studies of samples of the user population (e.g. surveys of Wikipedia editors where the participants were allowed to self-report their identity categories).
3. Intentionally constructing datasets for testing the generality of language technology in such a way that speakers can be asked about their racial identities (and furthermore have the chance to opt-in to the study and give informed consent).
4. Conduct error analyses in terms of linguistic features rather than language varieties, with reference to the linguistic literature about the varieties that such features are characteristic of. However, when taking this path, researchers should take care to avoid essentializing linguistic varieties as characteristic of racialized groups (Charity Hudley 2017) and similarly beware of linguistic appropriation/digital blackface (see, e.g. Abreu 2015).

3.7 Collection of data that potentially endangers participants

While we work on incorporating new antiracist practices in our research, towards the complementary aims of moving away from uncritical use of racial, ethnic (and linguistic) categories, paying greater attention to respondents' self-identification, and eliciting appropriate amounts and types of ethnicity data, we must be aware that there are times when collecting certain types of data may potentially endanger participants. Below, one example is presented.

Ethnic categories are locally defined and can be "loaded" due to their linkage to national identity and socio-political concerns such as immigration status. Smith (2019) investigated multilingualism among Senegalese expatriates in Rome, Paris, and New York, with particular attention to the powerful way that speakers achieved layered social meanings by code-switching between Italian, French, and Wolof. Some of her respondents in Rome were caught in a shift taking place in Italy in the early 2000s, away from a time when integration was facilitated by rules put in place in the immigration laws of 1998, to a new environment in which rhetoric from the highest levels of government was working to dehumanize and deter immigrants from remaining in Italy. Permanent residency was difficult to achieve for some of her Senegalese respondents.

While discussion of difficulties obtaining permanent residency in Italy was not the focus of elicitation, narratives of travel and mobility were invited as a principal method of gathering connected speech and included the possibility of discovering undocumented status (2019:103). Smith made a deliberate choice, born of an interest in protecting the anonymity of her respondents, to avoid gathering citizenship information or immigration status. If these topics came up in an interview, she de-identified the recordings and transcripts of respondents who might be in danger of deportation or detention. There was no need for information about citizenship or immigration status given the research aims of the study, and thus no reason to gather or to store it in any form (e.g., on recordings or demographic surveys). This is a case of information being volunteered in an interview. As one solution, the analyst can turn off the recorder (and instead write notes free of compromising details, as Smith did), or remove material from the recording after the session is over by excising audio or attenuating segments to silence. But, it is also common in sociolinguistic surveys to directly request neighborhood (Milroy and Milroy 1978), town (Labov (1966) or region of residence (Wolfram and Dannenberg 1999), or nationality information as part of the sampling protocol or screening process that determines eligibility for participation in a study or in a demographic survey. It is important to understand that such data might have negative repercussions for our respondents and consider when it may be best NOT to gather sensitive data that we never plan to use (e.g., in summary reports of usage of linguistic forms), or to present.

For field data collection settings, we recommend giving subjects the freedom to ask *why* we are gathering the information we are, how we plan to use it, and answering truthfully. After asking them to divulge personal information, some amount of discussion (and even reciprocation from us) may be an appropriate step to take to build trust and give subjects control over their own demographic details. For online settings, we can provide a link to a webpage explaining how respondent data will be used and describing more about the aims of the study. This is part of informed consent. Our Institutional Review Board (IRB) protocols likely say that subjects should be given the choice to opt of participating in any study, whether before or after procedures are undertaken. We must take this seriously. In summary, for fieldwork settings, we recommend:

1. If the fieldwork involves migrants, carefully consider whether there are potential harms in reporting citizenship status.
2. Incorporate prose into the consent form allowing respondents to determine how their data may be used, and with whom their data may be shared. This may include allowing them to listen to the interview after the recording session is done and exercise their right to wipe any part they do not desire to be kept.
3. If potentially identifying material is not required as part of the research (for linguistic or demographic analysis), consider de-identifying these portions of the recordings

The sections above have been primarily focused on quantitative types of study design. Many of the questions that we have raised, and prompt design challenges we have discussed relate to representation, essentialization, and the inherent challenges in utilization of fixed categories. For many such designs, however, a mixed methods approach that introduces qualitative methods provides a way of resolving some of these issues. The following section of the paper details some of the theory and practice that might prove useful in allowing for richer, more complex information to be captured, and for greater nuance to be used in its analysis. Such methods also lend themselves to more ethically appropriate research practices, more generally.

4. Potential Issues and Our Recommendations Regarding How to Address Them

4.1 Analyst positionality

All social research is founded on the human capacity for participant observation. We act in the social world and yet are able to reflect upon ourselves and our actions as objects in that world. (Hammersley and Atkinson 2006:22).

We take from the above that social researchers, yes even *linguists*, are part of the social world they study. We recognize in our field the tendency to believe it is possible to isolate a body of data uncontaminated by the researcher (by standardizing research procedures or attempting to deliver these with presumed “neutrality”). *Reflexivity* is a term that means two things in the field of ethnography (Hammersley and Atkinson’s audience): 1) the researcher’s critical awareness of their positioning within their field of study, and 2) ways that the analyst can introduce consciousness and commentary on their practices and activities in a cultural context. Reflexivity implies that how we orient ourselves to our work is influenced by the socio-historical location in which we are entering the field as a practitioner. Our work is not isolated from our academic heritage, including the earlier foci, findings, and limitations of our field. This also applies to the knowledge we produce and the way we produce it. Some scholars view Social Research as potentially supporting aspects of the political status quo in Western societies (Hammersley and Atkinson, 2006:17). Miles-hercules (2020) echoes this criticism, pointing out that Linguistics as a field has often viewed itself as unlinked to, and its interests quite apart from, other disciplines that encourage critical reflexivity on systems of power and oppression. It seems all the more important that researchers in our field (not just our sister fields of ethnography, sociology, anthropology, etc.) be trained to practice reflexivity.

Giddens (1979) and Hammersley and Atkinson (2006:17) both consider all social research as bearing hallmarks of participant observation: “it involves participating in the social

world, in whatever role, and reflecting on the products of that participation.” Our emphases are reflexes of that location. The production of knowledge therefore has consequences. What to do? We can work with the knowledge we have, recognizing it may be limited. When addressing respondents who we hope will inform us about the language of a particular group, we can call out our own lack of knowledge about the community and include knowledgeable locals to help us with study design. We can ask how we can treat people in the community in ways that they recognize as respectful. In this way, we can actively consider ways that harm may be avoided. Smith (2019) also invites respondents to ask her questions about *her* background, identity, and work, in a manner intended to diffuse the power differential her interviews may cause. She remains accessible to them during the fieldwork period, where reasonable (also see PNWE Case study, Section 5.1.2).

In a similar vein, when it comes to conceptualizing fieldwork, Milroy and Gordon (2003) emphasize that analysts need to recognize that the sociolinguistic interview is a speech event, with socially defined roles, format and expectations. Respondents will automatically attempt to comply with the “cooperative principle”, assuming power relations to be asymmetrical (and not in their favor), with unequal distribution in the right to talk, and that turn-taking rules control the interviewer-interviewee relationship. This often has the consequence of reducing the volume of talk received, even when the desired outcome is the collection of large amounts of natural conversation: “Just getting people to talk.” Respondents may view invitations to talk freely with suspicion if a non-familiar researcher asks personal questions when these seem to be at odds with a demonstrated interest in particular linguistic structures (the data, not the person). Milroy and Gordon (2003) encourage the interviewer to make clear the relevance of their questions since it is one of the expectations that speakers bring to the interview. Some researchers, including Briggs (1984) and Labov (1984) suggest that a researcher honestly acknowledge her position as a learner, “in a position of lower authority” than the person she is talking to. This allows the researcher to learn about the cultural norms of the community as someone new to the community, and to respect the respondent as an expert regarding their community.

More directly to the topic of this white paper, we recognize that we may plan for a research study in ways that more closely reflects earlier practice in our field, than represent the community, or the phenomenon, of interest. These earlier approaches impact how we draw our respondent samples or conceptualize speakers who might be suitable to represent the speech of the community. This can lead to including some types of people but omitting others, based on who researchers before us studied. We may need to ask the community how it defines itself, educate ourselves regarding the diversity that is typical in the community with a view to representing, rather than essentializing away from it.

Our Recommendations:

1. Before conducting fieldwork, the analyst should consider their social role explicitly. One excellent reference on this point is Briggs (2012). If you are planning to work as an outsider in a community whose social norms you don’t know or likely don’t understand, position yourself as a learner. Learn about role relations in the community and consider where a respondent might position you. Consider the impact you will have on the data.
2. Consider the power dynamic your interview sets up. Are there actions you can take to diffuse the power differential, avoid oppression, threat, or the appearance of coercion?
3. Strengthen your *reflexivity* chops. Read critiques of our field and consider how we might mitigate them in our personal research practices.

4.2 Issue type: over-reliance on category labels

Race and ethnicity, like any other categorizations of humans within a social world, are necessary shorthands for a complicated multitude of experiences and phenomena, as evidenced throughout this paper. In order to gain ‘big picture’ insights, scientists will often rely on these shorthands; however, the categories and their labels do not accurately represent the diversity and uniqueness of human lives and are not often interpreted the same way (see Sections 4.3.1 and 4.3.2 for more information).

“Racialization is the act of ascribing and prescribing a racial category or classification to an individual or group of people. The process of racialization includes differentiating and assigning group membership based on racial attributes including but not limited to cultural and social history, physical features, and skin color” (Charity Hudley, 2017:1).

Race essentialism is the tendency to view race as biologically based, immutable, and informative (e.g., Haslam, Rothschild, & Ernst, 2000; Prentice & Miller, 2007) (see also Section 1.1) and has been linked to racial stereotyping and prejudice, as well as a strong reliance on rigid racial categories/labels (e.g. Levy, Stroessner, & Dweck, 1998; Prentice & Miller, 2007; Williams & Eberhardt, 2008).

In describing the difference between racial identity, racial category, and racial identification, Rockquemore, Brunnsma, and Delgado (2009) explain that a racial identity is the extent to which an individual claims an identity through an act of self-identification based on contingent factors and immediate social context as well as broader cultural practices which can shift over time (Markus, 2010). On the other hand, a racial category refers to the racial choices institutionally available; while a racial identification is an outsider’s categorization of the self (Rockquemore, Brunnsma, & Delgado, 2009). Nowadays, racial categorization is believed to be attained by relying not only on visible phenotypic cues but also on multiple alternative sources like expectations, activated stereotypes, social perception, etc. (Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011). All these interacting aspects guide how a subject is categorized by different people, in different contexts, and at different times (for more detail, see Sections 1.1 and 3.4).

Researchers in many fields fall back on analysts’ classification of people into social categories, emphasizing a discrete view of race that may conflict with individuals’ racial identity, its fluidity over time, and its social perception in a given context. The variety of possible self-identifications, as well as the cultural and phenotypic ambiguity that makes these categories not always easily discernible, often results in racial categorizations that are incongruent with individuals’ racial identities (Sanchez, Shih, & Wilton, 2014). In other words, racial categories/labels in surveys often do not authentically portray subjects’ racial identities. If race is such a dynamic and interactive process, it seems extremely wrong to rely on fixed racial categories when conducting research. Yet, category labels are needed to allow certain types of analysis, as evidenced in Section 3.1. Not to mention that study participants might perceive their ethnicity and racial self as highly intertwined, and, as a result of that, race can inform researchers about linguistic practices and behaviors (see Section 1.1). It thus seems crucial to signal the advantages and disadvantages of relying on racial categories/labels, in order to expand and refine our categorization methods (Pauker, Meyers, Sanchez, Gaither, Young, 2018) as well as develop ethically appropriate research practices. In doing this, we cannot reiterate enough that, while

collecting data about race can be useful, linguists should not refer to a biological determination of race in their enquiries, nor use biological dispositions to explain linguistic phenomena (see Section 1.1).

What categories/labels help with:

1. They allow large-scale analysis
2. They can provide a ‘big picture’ of social phenomena
3. They can reveal social inequities across groups of people

What categories/labels fail to grasp:

1. People’s authentic self-identifications and intergroup variation
2. Outsiders’ categorization of the self
3. Fluidity of racial identities
4. People with multiracial/racially ambiguous identities
5. Intergroup perceptions and dynamics

Given these shortcomings, over-relying on fixed racial categories/labels might lead to inaccurate representations of social experiences and phenomena and nevertheless unethically foster stigmatization or identity denial (Albuja, Sanchez, & Gaither, 2019; Cheryan & Monin, 2005). We will thus address each point in detail, providing some suggestions and novel ideas. We do not claim to solve this issue entirely. The purpose of this section is to stimulate the reader’s reflection on how to inquire about race in a scientifically and ethically appropriate way when using racial categories/labels.

4.2.1 People’s authentic self-identifications and intergroup variation

Assigning a racial category to a self is not an act of self-identification. Identities are multifaceted and racial categories need to be flexible to leave space for subtle nuances. The development of the self is a dynamic, relational process, which is constantly negotiated, highly context-sensitive, and operates on various levels (De Fina, 2007; Tajfel & Turner, 1979). This also applies to racial identities, especially when intra-race intersectionality applies (Albuja, Sanchez, & Gaither, 2020) (see also Section 4.3.1), and language generally plays a crucial role in this (Clément, & Noels, 1992; Yip, 2005). Liebkind, Mähönen, Varjonen, and Jasinskaja-Lahti (2016), recommend exploring multiple dimensions when determining a person’s identity and suggest incorporating an individual level, an interactive level, and a societal level. In particular, the individual analysis assesses people’s degree of identification with the group; the interactive analysis examines their level of involvement in both their own and dominant group in the social context; and the societal level considers their political and ideological orientations (Liebkind et al., 2016). These are good practices also when examining people’s racial identity as well.

Our recommendations:

1. Allow subjects to create their own racial label, acknowledging further specification (see also Sections 3.1, 3.2)
2. For analytical purposes, if you are certain that what you wish to measure *is* race, and you can justify this, we still suggest eliciting participants’ responses using independently standardized

and validated racial categories or an open-ended response technique first, as described in Sections 3.1, 3.2. Further subcategories can be provided as well as extra space to let subjects express their own self-identifications or comment on the label(s) selected. It is crucial to remember that when self-categorization happens, it can place risk of harm on some respondents (i.e. on their immigrant status, see Section 3.7) or cause psychological discomfort. Subjects need to have the option to decline responding to any question in the questionnaire and need to be reminded of this option. Novel survey platforms offer the opportunity to store identifiers separately from the rest of the data (i.e. REDCap) in a way that drastically decreases the occurrence of sensitive data exposure. For more information about potential risk, we refer to Sections 3.4.3 and 3.7.

3. Include a question measuring subjects' sense of belonging to the selected racial category(ies) in order to evaluate the degree of in-group self-perceived variation and detect the level of intergroup affiliation (i.e. 'to what extent do you feel like you belong to the racial category(ies) you have selected above?').

4.2.2 Outsider's categorization of the self

As mentioned in the previous section, when defining one's identity, we should also examine the psychological engagement with in-group members and within the wider social context (Liebkind et al., 2016). For example, identity questioning from others (i.e. 'what are you?') might impair belonging and signal exclusion (Albuja, Sanchez, & Gaither, 2019, 2020; Cheryan & Monin, 2005; Wang, Minervino, & Cheryan, 2013). Also, the extent to which people's physical features appear prototypical of one racial group in the eyes of others are dependent on the type of social context in which they are immersed (Doyle & Kao, 2007; Feliciano, 2016; Good, Chavez, & Sanchez, 2010; Rockquemore & Brunsma, 2002). Social perception, which is based upon several contextual factors (socio-economic condition, verbal behaviour, presentation, intergroup exposure...) is strongly linked to racial categorization and impacts on the way individuals perceive themselves, too (Davenport, 2016). Hence, allowing for extra determination beyond standard racial categories/labels will help capturing a better picture of social phenomena. When doing so, it is advisable to inquire about both in-group and out-group perceptions.

Our recommendations:

1. Measure self-perceived racial categorization by asking subjects how their racial identity is perceived by others of same, similar, or different racial background. Extra contextual factors (i.e. social status, social networks...) can inform the researcher about the type of 'perceivers' the subject refers to when answering the question.

4.2.3 Fluidity of racial identities

A racial identity may vary situationally in response to a social context or as influenced by long-term experiences such as migration, or heritage connections (Davenport, 2016). Furthermore, people (especially if multiracial) are prone to shift their identity over time (Doyle & Kao, 2007; Harris & Sim, 2002; Hitlin, Brown, & Elder, 2006). Racial categories/labels should be able to grasp – at least to a minimal extent, identity variation.

Our recommendations:

1. Gather feedback from survey participants, especially over time (i.e. rating their confidence in selecting the racial category(ies), testing participants a second time, including space for comments...)
2. Allow further specifications through subcategories and open questions. Fluidity can be better detected through self-identification.
3. According to the research design, use multiple questions prompting subjects to think about different points in time, contexts, geographical locations, social connections... (i.e. ‘which racial group do you feel like you belong to, and when...?’)
4. Ask subjects to rate the perceived variation of their sense of belonging to the selected racial category/label (i.e. their perceived change over time...)
5. Use qualitative insights to supplement quantitative investigations. Interviews are a good tool to capture the subtle nuances in subjects’ identity shifting, while ethnographic observations could also inform the researcher about intergroup categorization and dynamics. For better practices we refer to Section 3.1.

4.2.4 People with multiracial/racially ambiguous identities

The 2000 US Census was the first one to allow respondents to select more than one racial identification. Nearly 7 million people (2.4% of the population) selected more than one race that year, and the multiracial population has grown extensively since then (Jones & Smith, 2001). By the next Census, the multiracial population had increased by 32%, whereas the monoracial population only grew by 9.2% (Jones & Bullock, 2012). These data trends make evident that multiracial populations are fast-growing and increasingly prominent groups in the US (Section 3.1.2).

It is important to highlight that not all multiracial individuals acknowledge their multiple racial identities. Some simply claim one identity (Chen, Benet-Martínez, & Bond, 2008). Many mixed-race adults do not identify as multiracial, and approximately 30% report having changed their identification throughout their lives (Jones & Bullock, 2012). Hence, flexibility is even more imperative when considering this type of subjects, as multiracial identities change across perceivers, contexts, and time (Bonam & Shih, 2009; Chen & Hamilton, 2012; Cheng & Lee, 2009; Gaither, 2015; Nguyen & Benet-Martínez, 2010; Shih, Bonam, Sanchez, & Peck, 2007) and this is strongly related to individuals’ psychological well-being (Yip, 2005).

The literature has used racial ambiguity (e.g., a mixture of racial features, not prototypical of one specific race) as one proxy for multiracial identity, especially when examining how individuals perceive and categorize social targets based on visual cues (Tran, Miyake, Martínez-Morales, & Csizmadia, 2016). The experience of dealing with ambiguous features that do not fit into stereotypical or traditional categories and the exposure to people with a racial identity that spans across categories can help develop more appropriate categorization methods.

Our recommendations:

1. Allow multiple answers when including racial categories/labels and do not include a ‘multiracial’ category as such. It is crucial to keep in mind that when ‘multiracial’ is included as a category option, perceivers tend not to categorize in line with hypodescent (Chen et al.,

- 2014, Chen & Hamilton, 2012, Peery & Bodenhausen 2008). Also, a general ‘multiracial’ category is not well defined and might result in incongruent identifications or identity denial.
2. Ask feedback from multiracial participants. Multiracial subjects understand better how race changes across situations, time, and depending on a number of contingent factors. From the perspective of multiracial individuals as perceivers, research showed that identity flexibility influences the process of social perception, determining a reduced belief in the essential nature of racial categories (Pauker et al., 2018; Wilton, Sanchez, & Giamo, 2014).
 3. Ask subjects to rate the perceived variation of their sense of belonging to each racial category/label previously selected.
 4. Allow further specifications by providing subcategories and open questions, or by prompting subjects to think about their identity shifting with time and contexts (i.e. ‘which racial group(s) do you feel like you belong to, and when...?’) (See Sections 3.2 and 3.3 for further reference).
 5. Enquire about the perceptions of others. This will help distinguish between general racial ambiguity (i.e. some individuals might be unaware of their full racial background) and explicit multiracial and intersectional self-identification and could help create references with activated stereotypes and social perception. As mentioned above, and reiterated in Section 4.3.1, extra contextual factors (i.e. social status, network of relationships, etc.) can inform the researcher about the type of ‘perceivers’ the subject refers to when answering the question. This is extremely important here as it helps verify potential patterns of stigmatization or identity denial.
 6. Check patterns of racial categorization in other cultures/countries, as this should reflect the dominant narrative of race in that culture. This practice can help stretch the categorization potentials, better eluding a monoracially biased approach (Chen, Paula Couto, Sacco, & Dunham, 2017).

4.2.5 Intergroup perceptions and dynamics

People’s involvement in both their ethnic and dominant culture on an everyday basis is part of racial identity practices (Liebkind et al., 2016). The underlying idea is that individuals generally strive for a positive self-image and they may adopt different strategies to attain this goal. For instance, they might endeavour to leave their ethnic group in favor of a more positive affiliation and/or they may attempt to make their group more positively perceived by their peers (Tajfel & Turner, 1979; Pauker et al., 2018). Interpersonal denial refers to lived experiences of having the racial or cultural identity unacknowledged or questioned by others (i.e. ‘you don’t look like a member of your racial group’ or ‘where are you really from?’) (Albuja, Gaither, Sanchez, & al. 2019, Albuja, Gaither, & Sanchez, 2019). Experiences of this kind are most common for less prototypical in-group members or multiracial subjects when forced to make a single racial identification choice (Cheryan & Monin, 2005; Townsend, Markus, & Bergsieker, 2009). The researcher should keep in mind that identity denial can be experienced as an identity threat and devaluation of the self that leads to consequences (Albuja, Gaither, & Sanchez, 2020, Steele, Spencer, & Aronson, 2002; Sanchez, Chavez, Good, & Wilton, 2012). Being denied a relevant social identity, for example, may result in reassertion (Branscombe, Ellemers, Spears, & Doosje, 1999) or self-rejection through intragroup marginalization (Albuja, Gaither, Sanchez, & al. 2019; Albuja, Gaither, & Sanchez, 2019). Empirical evidence indicates that participants, especially if racially ambiguous, value others’ perceptions of their race as a form of self-verification, further

confirming that identity denial, where identity is not perceived accurately, is a negative experience (Remedios & Chasteen, 2013). The researcher needs to be aware of the fact that individuals, especially if multiracial, racially ambiguous, or ethnically marginalized, may undergo a strong pressure both from in-group members and the external social context to perform a certain social script that aligns with a specific racial identity (Castillo, Conoley, Brossart, & Quiros, 2007).

Our recommendations:

1. Ask subjects to rate the perceived variation of their sense of belonging to each racial category/label previously selected and ask for general feedback (i.e. rating their confidence in the selection).
2. Ask subjects how their racial identity is perceived by others. Allow space for subjects to comment on their racial categorization.
3. Collect more information about subjects' socio-economic status, cultural affiliations, psychological integration in the society, linguistic practices... (See also Section 4.3.1) Understanding subjects' cultural orientation, their status in the society, their type of socio-psychological engagement with the dominant community can help detect inter-group dynamics and assess the validity of the categories/labels used.

4.2.6 General good practices when using categories

We conclude this section by presenting some general advice that might prove helpful when having to rely on racial category/labels:

1. Before creating a survey, let the community guide the racial categorization used. Regardless of whether researchers will elicit information about racial self-identification or not, learning more about the composition of the community and what the categories in use are can ensure fairness in the sampling process, as evidenced in Section 1.3. Piloting, and socially contextualizing the survey is a great way to avoid having sampling biases and homogenizing groups.
2. Ask about subjects' self-perceptions when piloting the survey in order to identify potential flaws in the categorization system adopted.
3. Supplement survey data with ethnographic/qualitative data. This practice is probably the best option to avoid categorization systems. In this process, specifically seek for multiracial subjects' feedback. Multiracial identities are indeed a reminder that categories can fluctuate (Wilton et al, 2014).
4. Lastly, being guided by the research questions, verify whether race and ethnicity categories/labels are really needed in the research.

We hope that this section will guide researchers in developing practices able to better capture the diversity and uniqueness of subjects' experience when it comes to race. Other useful readings that illustrate the complexities of racial categorization in relation to identity shifting and multiracial identities are: Feliciano, (2016); Kang & Bodenhausen, (2015); Sanchez, Shih, & Garcia, (2009); Shih, Sanchez, & Ho, (2010).

4.3 Issue type: possible confounds in your data (and how to discuss them)

This section discusses issues that may arise from incomplete or skewed data samples, as well as particular statistical methods. We approach this section as a practical attempt to improve the ability of sociolinguistic research to navigate intersectional social systems.

4.3.1 Intersectionality should inform your understanding of social categories

The term “intersectionality” originated in legal scholarship. Crenshaw (1989) coined the term in order to discuss how, for example, Black women may experience discrimination not only in similar ways as other women or as other Black people (which would mean that existing anti-discrimination laws would be applicable) but also be discriminated against as *Black women*, beyond a simple additive effect. Since its inception, intersectionality theorizing has expanded beyond simply describing the lived experience of intersection identities; the dynamism and change of intersectional identities and mutually constitutive nature of the identities that make up the intersections as well as the intersections themselves are all much discussed topics (Crenshaw 2017; Davis 2008; Choo and Ferree 2010; see Levon 2015 for a literature review tailored to linguistics). Levon (2015) clearly spells out the need for intersectionality to be considered more broadly in sociolinguistics research; coming at this idea as a researcher in the subfield of language, gender and sexuality, they state:

“I propose that we can achieve the goal of integrating intersectionality more fully by combining a focus on marginalized lived experiences with a sustained examination of the ways that linguistic practices linked to one category are used to constitute another. What I mean by this is that we need to go beyond an investigation of solely ‘gendered’ or ‘sexual’ features to also include an analysis of features normally associated with other social systems (e.g., race/ethnicity, social class, and place) and of how those features are recruited in the construction and perception of different sexual and gendered positionings.” (Levon 2015:301)

The insight that Levon identifies -- that linguistic features associated with other social systems may be recruited to construct linguistic gendered and sexual identities -- goes both ways. For the purposes of this paper, we acknowledge that linguistic features associated with other social systems (gender, sexuality, class, place) may also be recruited in the construction of ethnic and racial identities and linguistic performances; see Sections 4.2.4 and 4.2.5 for further discussion.

Furthermore, failure to take into account intersectionality when investigating social phenomena closely tied to identity can have adverse effects on participants (and later readers of the research); Albuja, Sanchez, and Gather (2020) show the stressors placed on individuals who occupy multiply marginalized identities, including identity denial and identity questioning. These stressors are associated with poorer mental health outcomes, including lower feelings of autonomy and belonging and greater feelings of conflict. Linguists should keep in mind that insufficiently intersectional design of research questions and materials can exacerbate the stressors of identity denial and identity questioning in participants and should endeavor to address this proactively as part of our core ethical responsibility to our informants and communities.

In order to adequately capture the complex relationship between multidimensional social positions that are multiplicative (not simply additive), we use this section to make specific recommendations on how to collect and analyze (socio)linguistic data so as not to obscure, misrepresent, or otherwise under-explain the system of intersectional social relationships and positions. This section reviews how some quantitative methods, data collections, and modes of theorization may ‘flatten’ what is in reality a complex web of social relations, and suggest ways to instead include this nuance which is crucial to understanding language and race and ethnicity in their full context.

4.3.2 Certain quantitative methods can obscure insight

The effects of multiple intersecting marginalized identities, including those related to race and ethnicity, can be obscured or erased by some common quantitative analysis techniques in social science research. In these techniques, analysts exclude certain predictive variables, measured as part of a study, from statistical models constructed to answer a study’s research questions. This subsection outlines two common situations in which this happens and offers best practices for avoiding obfuscation of intersecting marginalized identities.

One situation in which one or more of an individual’s several marginalized identities is commonly obscured is in cases of suspected (multi-)collinearity, i.e. correlation of two or more predictor variables in a statistical model. If two predictor variables, such as race and socioeconomic status, are thought to be highly correlated, it might be assumed that they measure the same thing, and one variable is then sometimes removed from the model. Another common scenario in which predictor variables -including race and ethnicity- are excluded is during stepwise regression, a technique for fitting statistical models in sociolinguistic studies in which predictive variables or their interactions are selected for inclusion in a final model based on the outcome of statistical tests or other criteria. Stepwise regression persists despite a general opposition to its use by statisticians and its high likelihood of producing spurious results in sociolinguistic studies (Johnson 2010). The exclusion of predictive variables judged to be statistically insignificant is misunderstood to be beneficial; in actuality, it is possible that a causal effect of a factor such as race or ethnicity exists despite its not reaching statistical significance (Smith 2018). Whether removed because of collinearity, stepwise regression, or for another reason, an additional drawback of removing predictor variables from a model means that they are no longer controlled for, and there is potential for overstating or understating the effect of the remaining predictive variables (O’Brien 2007). For example, in a sociolinguistic study of vowel merger which includes age, gender, and ethnicity, if gender were to be removed, the effect size of ethnicity could be exaggerated or downplayed, leading to inaccurate conclusions about the social distribution of a linguistic change. What can be done instead?

Before data collection:

1. Determine the social variables which are likely to be relevant for a particular object of linguistic study through qualitative work, by conducting a pilot study, or, ideally, both.
2. Use the most appropriate sampling techniques available.
3. Collect a sufficient amount of data, both in terms of respondent demographics and number of participants; including more independent variables in a study will increase the minimum number of participants needed.

After data collection:

1. Instead of stepwise regression, run and report a full model. Consider additional data collection if there are too many predictors in the model for the number of observations collected.
2. For collinearity, Gordon (2010:390-393) suggests that a researcher take two ideas into consideration: whether the two predictor variables measure the same or different constructs, and whether they are strongly correlated. If the answer to both questions is yes, it can make sense to create an index from the multiple variables, or to conduct a confirmatory factor analysis or use item response theory.
3. Consider statistical techniques that take into account multiple demographic categories, such as linear discriminant analysis.
4. Only include in a statistical model the sociodemographic information that you have collected as potential predictors. Sociodemographic information collected for other reasons, such as for investigating potential sampling bias, should not be included in a model.
5. Rather than including interactions between all variables in your model, include interactions on known associations between categories, from existing work in your context, or through your own preliminary work. Interaction effects are important to include because the effects of membership in two or more particular social categories are not always accurately captured simply by summing them up (Crenshaw 1989).

4.3.3 Sampling bias may affect your findings

Whether you are using any of the statistical methods above or not, the distribution of data you collect may lead to confounding variables in your findings. If you collected data on race and/or ethnicity as well as gender and/or sex or other demographic categories, you may find that sampling bias may create an irregular distribution of data such that you cannot use a single variable to explain behaviors. For example, if your participants included mostly older White men and younger Black women, you might not be able to explain differences between those groups using only one of those variables. Such an uneven distribution of data may reduce the efficacy of certain statistical methods (as discussed above). Even without any inferential statistical analysis, however, summarizing or drawing conclusions from irregularly distributed data may produce misleading results.

A hypothetical summary table of the demographics for the example experiment above is shown in Table 2 below. Table 3, following, shows the hypothetical results of a variable based on racial category in this hypothetical experiment.

Table 2: *Demographic makeup of Hypothetical Experiment 1 (respondent counts)*

	Black		White	
	Men	Women	Men	Women
Young	1	30	1	0

Mid	2	2	2	2
Old	0	1	48	4

Table 3: *Percent bag-raising by race for Hypothetical Experiment 1*

Black speakers (n=36)	White speakers (n=57)
44%	79%

Notice that, while no inferential statistical analysis is presented, it is still impossible for these data to realistically support any hypothesis about only one of the social variables (ethnicity, age, or gender). If the data are skewed in this way due to recruitment or other practical concerns, this is important to discuss as researchers contextualize their findings. For this hypothetical experiment, it would be misleading to report the percentages by only one demographic variable as Table 3 above does. This is because Table 3 implies that there is a difference in the rates of bag-raising for Black and White speakers; while this appears to be trivially the case, it is not at all clear based on these data that race is the explanatory variable, since gender and age are also so irregularly represented among the participants.

In order to responsibly report findings like the ones in Hypothetical Experiment 1, it is necessary to directly address the distribution of demographic data, and to avoid presenting misleading summaries like the one in Table 3.

What to do instead?

1. When summarizing or reporting findings, be explicit about the distribution of all categories about which data was collected; if the distribution is highly skewed, include a clear summary like the summary shown in Table 1 to make that distribution easy to see at a glance.
2. Discuss the skewed distribution of the findings explicitly in your discussion section, and be up-front about what the data can and cannot explain

4.3.4 Insufficient data collection may affect your findings

The previous section assumes that the researcher has collected all the relevant demographic information from participants; however, it is always possible that there are social categories that may not have been directly addressed in data collection, which may nevertheless be relevant to the research question. Incomplete demographic data may hide confounds just as well as skewed demographic data. In Hypothetical Experiment 1, discussed above, data is included for racial categories but appears to conflate race with ethnicity -- or else, ethnicity data was simply not collected. Conflating variables or otherwise failing to collect relevant demographic data may be insufficient to explain within-group or between-group differences.

In Hypothetical Experiment 1, Table 2 shows summaries of two racial categories (Black and White). However, there is no inclusion of ethnic categories. If, for example, White

participants in Hypothetical Experiment 1 were ethnically heterogeneous, there may be unexplained within-group variation. Hypothetical Experiment 1 may have taken place in a community with a large Ashkenazi population, where Ashkenazi people largely self-identify as White. If Hypothetical Experiment 1 *had* asked about ethnicity in addition to race, more information would become evident. Table 4 shows that difference (compare to Table 3 above).

Table 4: *Percent bag-raising by race and ethnicity for Hypothetical Experiment 1*

Black - African American (n=30)	Black - all other ethnicities (n=6)	White - Ashkenazi (n=39)	White - all other ethnicities (n=18)
50%	15%	90%	55%

The within-racial-group ethnic identities shown in Table 4 would be unexplained variation to the researcher of Hypothetical Experiment 1 if that researcher had collected no information on ethnic identities. This is one of many possible ways that the researcher could have missed important or relevant variables, of course, but is of particular interest because the tendency (across linguistics and many other social sciences) to conflate race and ethnicity severely hampers our ability to explain *why* differences in linguistic behaviors exist.

Our first and most important suggestion for preventing confounds like the one in Hypothetical Experiment 1 is that researchers collecting information on race should also collect information on ethnicity, and vice versa. The data in Table 4 do not completely de-mystify the bag-raising situation in Hypothetical City, but give the researcher an important head-start on how to direct future work (such as: what’s going on in the Ashkenazi community in Hypothetical City that leads to such high levels of bag-raising? Is this variable actually linked to Jewish identity in this city? Are different Jewish ethnicities significantly different from each other in this community? And of course, many other questions that arise from this richer data.)

4.3.5 Main takeaways

There are many ways that data collection, research question design, and quantitative (or other) analysis may obscure conflated variables in sociolinguistic research; our interest in this section has been discussing the particular ways that this arises around the collection of data about race or ethnicity. This subsection has covered particular statistical methods to be cautious with, as well as possible consequences of skewed distributions or incomplete data. In general, we recommend several ways of protecting research design from these pitfalls:

1. Discuss these issues directly in the Discussion section and include discussion of possible conflation or confounds in the data; this should not be an ‘admission of guilt’ but rather an intentional effort towards transparent science.
2. Include a thorough discussion of future directions for the research in order to fill the gaps; no research design is perfect, but science is a collaborative accumulation of knowledge, and researchers can directly lay groundwork for themselves and other others who can further investigate places where the data don’t tell the whole story.

3. Directly discuss interaction effects stemming from intersectionality in the Discussion section. In quantitative work, include multiple demographic variables in the analysis (but only if the information was collected!).

Adopting a clear and appropriate study design will also prevent many of the pitfalls described here; refer to Section 1.3 for discussion on research design.

4.4 Issue type: how to identify and talk about potentially harmful implications of one's work

The previous sections (2 and 3) have discussed primarily ways that a researcher can design linguistic studies to best mitigate harm, and to most appropriately treat race and ethnicity in linguistics. This section is focused on the ‘afterlife’ of that research; i.e, how a researcher may anticipate what issues may arise *after* a piece of research is published (in a scholarly outlet or otherwise). This section has two main goals:

1. Discuss how to embed intentional anti-racist framing into the writing of a publication, to (ideally) prevent bad-faith actors from citing the publication in racist writings
2. Present some potential ways to mitigate harm if one's publication *does* get cited in support of racist causes

This section is only a preliminary approach towards what must be a larger discussion in linguistics and our sister sciences: if we (as linguists) prefer that our science not be used to harm others, we should continue to investigate, adopt, and expand existing approaches towards a citational ethics more generally.

4.4.1 Writing to prevent bad-faith citation

It will not always be possible to anticipate all possible routes for bad-faith actors to cite our research in ways that misrepresent the research, enact harm, or are fundamentally oppositional to our social and political aims as scholars and authors. However, making an intentional effort to frame our writing and science communication in ways that root it in anti-racist perspectives can make it harder for people to take it out of context. Here are some strategies to consider:

1. Look through your writing for specific references to race. How would each such sentence sound if taken out of context?
2. When you refer to your participants, avoid using generalizing phrases: if your study includes Black people and your writing refers to that subgroup of your study participants, “Black participants” or “Black study participants” are less likely to make it easy for people to use direct quotes out of context in a harmful way than “Black people”, while likely being more accurate in terms of what you can claim based on your study.
3. Check your writing for ‘White gaze’ and ‘othering’: Do you only mention race when describing racialized participants? Do you use third person plural pronouns to refer to racialized participants but first person for White participants? Do you write about Indigenous people primarily in the past tense?
4. Check your writing for being overly accommodating of hegemonic viewpoints. That is, there are definitely times when it is worth directly engaging with and refuting ideologies of race and

language. On the other hand, outside of such discussions, it is preferable to just assume as common ground such notions as ‘White supremacy is bad’ and not defensively motivate that position.¹

5. Consider including an ‘ethical considerations’ or ‘broader impacts’ section in your paper. You can use this space to specifically warn against misappropriation and explain why certain uses of your results would be invalid and/or inappropriate. This practice is gaining ground in computational linguistics conferences and related fields.²

4.4.2 Handling bad-faith citation

When a bad-faith actor uses your work in ways that reinforce or support racist systems, it may be appropriate to respond. Here are some general tips for handling such cases:

1. Decide whether a reply is warranted, or would just give something fairly obscure more visibility. If a reply is not warranted, then nothing needs to be done immediately. However, it is nonetheless worth keeping an eye on whether the bad-faith piece is gaining traction, and you might want to draft a reply to have ready.
2. Analyze how your work was misappropriated. What jumps in reasoning are being made to connect your research to the point the bad-faith actor is trying to make?
3. Draft your response and share it with a trusted colleague, to check for tone and accessibility.
4. Claim the academic high ground and avoid taking a defensive stance. This should be written for a general audience since it is likely that the bad-faith piece you are responding to was as well.
5. Create a platform (if you don’t have one already). This can be a website, blogging platform, or social media presence associated with your research. Sites like medium.com are useful for putting up content if you don’t already have web space.
6. Publicize the response, keeping in mind that you are trying to reach both the linguistics audience as well as other people who may have seen the bad-faith piece and not your own research. Channels that you can access directly include posting to social media (e.g. Twitter) and pitching op-eds to news outlets.
7. If there is media attention, and reporters are contacting you, set aside time to talk with them. Before agreeing to an interview, look at previous work by the reporter making the query and consider the angle they approach you with. Before giving an interview, write down your main talking points, so can you be sure to include them in your answers.

¹ A paper that takes this tack---assuming ‘white supremacy etc. are bad’ unapologetically---is Bender, Gebru et al 2021.

² See, for example, the guidelines from NAACL 2021 (<https://2021.naacl.org/ethics/faq/>), ACL 2021 (<https://2021.aclweb.org/ethics/Ethics-FAQ/>), and NeurIPS 2021 (<https://neuripsconf.medium.com/introducing-the-neurips-2021-paper-checklist-3220d6df500b>).

5. A Case Study: The Pacific Northwest English Study (PNWE)

In this section, we include a real-world example of conducting research which includes ethnicity as a sociolinguistic variable of interest, the Pacific Northwest English Study (PNWE). The main aim of this section is to give researchers who are unfamiliar with including ethnicity in their research design, an inside view into the process of including this variable and perhaps some insight into the kinds of methodological decisions that a researcher would need to make. The present section builds on Sections 3.1, 3.2, 3.3, 4.2, where we discussed different methods by which one might want to include questions about race and ethnicity.

5.1 Context of case study

Here we present an example of research design that does not neatly fit into the categories free-response, multiple choice, or interactive interview. This hybrid approach to ethnicity data elicitation can offer researchers the flexibility to combat issues surrounding each method highlighted in Sections 3.1, 3.2, 3.3, and 4.2.

5.1.1 Introduction to the PNWE

The PNWE study broadly aimed to understand the development of English in the Pacific Northwest 200 years after the introduction of the English language into the region, especially given the overall lack of dialectology research in the region. An interesting aspect of the Pacific Northwest is the near constant migration into the region of people with diverse backgrounds who bring with them different ways of speaking. In order to adequately study English in the Pacific Northwest, one needs to include groups of people representative of the population of the region. This means taking into account the ethnic composition of the region and not just focusing on Non-mobile Older Rural Males (NORM) of the majority ethnicity. This work was conducted by Dr. Alicia Beckford Wassink and her research team in Washington state beginning in 2006. Currently there have been three waves of the study hereafter referred to as PNWE I, PNWE II, PNWE III. Race was not included in the study because the Principal Investigator's theoretical model of race precluded self-identification for racial categories (see Sections 3.2 and 4.2). Because embedding in local life (via participation in the speech community and cultural practice) was the focus, *ethnicity*, not race was the focus. Although including a balanced representation of ethnicity was prioritized in the design of PNWE I, the National Science Foundation (NSF) considered prioritizing ethnicity risky. The work on the geographic region was underrepresented in the dialectological and sociolinguistic literature, and more generally, work on race and ethnicity was also underrepresented in the dialectological and sociolinguistic literature. The original study was meant to include a multiethnic sample, but funding was only provided for including White participants. The NSF allowed for some non-White participants to be included as a proof of concept. The sociolinguistic literature at the time did not include clear models for analyzing racial and ethnic minority participation in regional language change, which likely contributed to this funding decision. Funding opportunities for PNWE II and PNWE III allowed for included non-White participants more robustly.

5.1.2 PNWE design considerations

From a research design perspective, the inclusion of ethnicity as a variable in the design of PNWE II and III provided many challenges. Firstly, the specific groups of ethnicities included in the study were not chosen arbitrarily. With the desire to ethically represent the demographics of the region, Wassink chose to create a historical profile of the region's demographics to discover which groups of people demonstrated long standing settlement within the Pacific Northwest region. For example, the Yakama (also known as the Yakama Nation) represent a confederation of Native American tribes indigenous to the lands in the eastern part of Washington State. They participated in the 1850's division of land in what was then the Washington Territory, and thus were justifiably included in the study. Then, drawing on recommendations from both the field of linguistic anthropology and the American Anthropological Association (AAA), Wassink conducted research locally for PNWE II to understand the conceptualization of ethnic identity by different communities themselves. Self-identified terms formed the categories that Wassink later quantified in phonological analysis. It should be noted that participants were allowed to identify themselves as belonging to multiple categories, again following examples from the AAA, and were given space to identify themselves as belonging to certain communities in lieu of an a priori categorization. The choices made to exclude specific groups were considered with intention. For example, Wassink chose not include members of the Duwamish Tribe (dx^wdəwʔaš) because procuring potential participants proved nearly impossible given the lack of Duwamish registry.

When collecting self-identified and self-reported ethnicity data during the study, PNWE II and PNWE III used a demographic survey delivered orally in an interview format. This was to emulate the structure of a discrete questionnaire but allow the participants to fully respond to the various questions or expand if they desired. Examples of questions included: "what is your own family's ethnic background?" and "what were the main racial and ethnic groups of your school?". Wassink never directly asked about race in the questionnaire, acknowledging that in this way the open-ended questions that allowed for ethnic identity to be collected were not wholly systematic. However, they were helpful, as the model of ethnicity grounding the experiment was based on shared cultural experiences instead of "putting yourself in a racial bin." It should be noted however that some participants did self-identify using terms that more closely resembled racialized groupings, an effect of the non-standardized collection method. Additionally, the tension between qualitative data and quantitative analyses was felt throughout the research study. Lastly, Wassink specified that her own external facing identity (see Positionality Statements at the beginning of the document) during interviews, and openness to discuss her own minoritized identity, likely prompted participants from non-White backgrounds to eagerly share their own family's histories with the wider academic world, whereas European-background research assistants who administered interviews reported feeling "talked to" about participants' backgrounds. This is another example of why positionality is important to share, and an example of the considerations that must be taken into account when deciding how to procure ethnicity data, however it shows the researcher's thoughtfulness in these study decisions.

In summary, PNWE is a real-world example of including ethnicity in sociolinguistic research. PNWE is a strong example of including ethnicity in studies because of how intentional the study design decisions were regarding who was included and the methodology. We believe that it is appropriate for researchers to make methodological decisions that do not directly match free-response or multiple-choice methods discussed in Sections 3.1 and 3.2.

5.1.3 Handling unexpected situations in the field

Even with rigorous and thoughtful preparation, it is likely that researchers will encounter unexpected situations in the field. While conducting PNWE research, several changes were made to the study's protocols in order to continue to responsibly conduct research. One such change involved seeking additional approval from local tribal authorities and undergoing additional researcher training. Prior to commencing fieldwork, the study received approval from the University of Washington IRB and sought approval from the local tribal IRB, the Northwest Portland Area Indian Health Board (NPAIHB), of which the Yakama Nation is a member. NPAIHB determined that their review was not necessary for this particular study, and fieldwork was about to commence. However, upon contacting a local tribal organization for assistance in locating study participants, the researchers learned that the Yakama Nation had begun the process of creating their own IRB and had made the decision to review studies and to require specific research ethics training before agreeing to endorse or permit the initiation of data collection. The PNWE research team underwent the training, which was focused specifically on research with Native American respondents and the history of ethical failings making this specific training necessary. The researchers also learned that the Yakama Nation had begun to require that tribal leaders be presented with the opportunity to review any research findings prior to publication in order to safeguard the intellectual knowledge of the Yakama people.

Another necessary change pertained to the study's elicitation materials. As part of the study, adapted from a large-scale sociolinguistic survey (Labov, Ash, & Boberg 2006), respondents were asked to judge the acceptability of phrases containing certain syntactic constructions. One such construction was *needs* + past participle, for which respondents were asked, "What if a mother said to her child, 'Your hair needs cut?'" It emerged during the course of the research that this phrase was a trigger--evocative of the traumatic involuntary assimilation practices to which Native American children were subjected at government boarding schools, which included forcible cutting of long hair. This particular prompt was removed and was later replaced with a non-triggering alternative. Thus, with any study involving specific communities, the onus falls on the researchers to be aware of differences between communities' cultural practices and carefully consider ethical issues arising from each aspect of the study.

Our recommendations:

1. When using ethnicity as a core linguistic variable, remain flexible to necessary design modifications
2. Ensure that the ethnic groups chosen for study are appropriate for the research question instead of assumed to have importance *a priori*
3. Research possible ethical issues surrounding linguistic data elicitation from the communities that are the subjects of the research

6. **Conclusion**

The purpose of this work is both to clarify the ways in which linguistic research has incorporated

(and often misinterpreted) conceptualizations of race and ethnicity in previous investigations, and to provide practical advice for addressing pressing issues related to the collection, analysis, use, and interpretation of these data. Drawing on very recent work in linguistics and on recent work in our sister fields in the social sciences and humanities, we have aimed to develop a more critical, ethical, and widely applicable representation of race and ethnicity in linguistic studies. While a distinct definition of race and ethnicity can be stipulated, it is often hard to operationalize these concepts. This operationalization is even more challenging if we consider the multiple psychological implications that ascribing race or ethnicity can have for each and every individual and how these concepts often intertwine in subjective self-identifications. These potentially harmful implications exist for research study participants, for policy makers, and for researchers themselves.

We believe that we have successfully created a comprehensive guide illustrating why it is important to collect such data, why analysts need to be able to understand and justify their research practices, why it is important for discussions of how ethnicity and race are conceptualized within the field of linguistics, what the current ethical standards are, and how to avoid potential harm and incongruencies. We recognize the sensitivity and relevance of the topic and strongly counter the use of biologically determined definitions of race in any linguistic investigation. In the bigger picture, we hope that this work not only sets the path for improved research practices and stronger scientific validity, but also helps foster linguistic and social justice, empower different ethnic communities, and dismantle patterns of inequality.

Acknowledgement

The authors would like to thank Laura Munger for help with editorial tasks related to this paper.

References

- Abreu, A. M. (2015). Online Imagined Black English. *Arachne*. Available at: https://arachne.cc/issues/01/online-imagined_manuel-arturo-abreu.html
- Adida, C. L., Ferree, K. E., Posner, D. N., & Robinson, A. L. (2016). Who's asking? Interviewer coethnicity effects in African survey data. *Comparative Political Studies*, 49(12), 1630-1660.
- Albuja, A. F., Sanchez, D. T., & Gaither, S. E., 2019. Identity questioning: Antecedents and consequences of prejudice attributions. *Journal of Social Issues*, 75(2). 515–537.
- Albuja, A. F., Gaither, S.E., Sanchez, D.T., Straka, B., & Cipollina, R. 2019. Psychophysiological stress responses to bicultural and biracial identity denial. *Journal of Social Issues*, 75, 1165-1191
- Albuja, A.F., Sanchez, D.T., & Gaither, S.E. 2020. Intra-race intersectionality: Identity denial among dual-minority biracial people. *Translational Issues in Psychological Science*, 6(4), 392–403.
- Baugh, J. 2017. Linguistic profiling and discrimination. *The Oxford Handbook of Language and Society*, January. 349–368.
- Bender, E. M., & Langendoen, D. T. 2010. Computational linguistics in support of linguistic theory. *Linguistic Issues in Language Technology*, 3(2), 1-31.

- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In Proceedings of FAccT, 610-623.
- Berg, B. L. 2007. *Qualitative research methods for the social sciences*. London: Pearson.
- Bobo, Lawrence. 2001. Racial attitudes and relations at the close of the twentieth century. In *America becoming: Racial Trends and Their Consequences*, vol. 1, ed. Neil J. Smelser, William Julius Wilson, and Faith Mitchell, 264-301. Washington, DC: National Academy Press.
- Bolton, K., & Hutton, C. 2000. Orientalism, linguistics and postcolonial studies. *Interventions: Orientalism and Linguistics*, 2(1), 1-5.
- Bonilla-Silva, E. 2006. *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States* Rowman & Littlefield Publishers.
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. 1999. The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje, (Eds.), *Social identity: Context, commitment, content*. 35-58. Oxford, England: Blackwell Science.
- Briggs, C. L. 2012. *Learning How to Ask*. Cambridge: Cambridge University Press.
- Bucholtz, M., & Hall, K. 2004. Language and identity. *A companion to linguistic anthropology*, 1, 369-394.
- Bucholtz, M. 2020. Race, Research, and Linguistic Activism. *The Routledge Companion to the Work of John R. Rickford*.
- Bucholtz, M., Lopez, A., Mojarro, A., Skapoulli, E., VanderStouwe, C., & Warner-Garcia, S. 2014. Sociolinguistic justice in the schools: Student researchers as linguistic experts. *Language and Linguistics Compass*, 8(4), 144-157.
- Castillo, L. G., Conoley, C. W., Brossart, D. F., & Quiros, A. E. 2007. Construction and validation of the Intragroup Marginalization Inventory. *Cultural Diversity & Ethnic Minority Psychology*, 13, 232–240.
- Charity Hudley, Anne H. 2017. Language and racialization. *The Oxford Handbook of Language and Society*. Oxford, UK: Oxford Handbooks.
- Charity Hudley, Anne H., Christine Mallinson & Mary Bucholtz. 2020. Toward racial justice in linguistics: Interdisciplinary insights into theorizing race in the discipline and diversifying the profession. *Language* 96(4). 200–235.
- Chen, S. X., Benet-Martínez, V., & Bond, M. H. 2008. Bicultural identity, bilingualism, and psychological adjustment in multicultural societies: Immigration-based and globalization-based acculturation. *Journal of Personality*, 76, 803–837.
- Chen, J. M., & Hamilton, D. L. 2012. Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology*, 48, 152–164. <http://dx.doi.org/10.1016/j.jesp.2011.10.005>.
- Chen, J. M., Moons, W. G., Gaither, S. E., Hamilton, D. L., & Sherman, J. W. 2014. Motivation to control prejudice predicts categorization of multiracials. *Personality and Social Psychology Bulletin*, 40(5), 590–603.
- Cheng, A., Faytak, M., & Cychosz, M. 2016. Language, race, and vowel space: Contemporary Californian English.
- Cheng, John. 2003. Race: The Power of an Illusion. <https://www.racepowerofanillusion.org/qa/what-difference-between-race-and-ethnicity>.
- Cheng, C. Y., & Lee, F. 2009. Multiracial identity integration: Perceptions of conflict and distance among multiracial individuals. *Journal of Social Issues*, 65, 51–68.

- Cheryan, S., & Monin, B. 2005. "Where are you really from?": Asian Americans and identity denial. *Journal of Personality and Social Psychology*, 89, 717–730.
- Choo, Hae Yeon, and Myra Marx Ferree. 2010. Practicing intersectionality in sociological research: a critical analysis of inclusions, interactions and institutions in the study of inequalities. *Sociological Theory*. 28(2). 129–49
- Comaroff, J., & Comaroff, J. (2009). *Ethnicity, Inc.* Chicago: The University of Chicago.
- Conrod, K. 2019. *Pronouns raising and emerging* [Doctoral dissertation, University of Washington].
https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/44842/Conrod_washington_0250E_20754.pdf?sequence=1&isAllowed=y
- Clément, R., & Noels, K. A. 1992. Towards a situated approach to ethnolinguistic identity: The effects of status on individuals and groups. *Journal of Language and Social Psychology*, 11(4), 203–232.
- Crenshaw, Kimberlé. 1989. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *u. Chi. Legal f.*, 139.
- Crenshaw, K., Harris, L. C., HoSang, D., Lipsitz, G., 2019. *Seeing race again: Countering colorblindness across the disciplines*. Oakland, California: Oakland, California: University of California Press.
- Cukor-Avila, P., & Bailey, G. 2001. The effects of the race of the interviewer on sociolinguistic fieldwork. *Journal of Sociolinguistics*, 5(2), 252-270.
- Czaykowska-Higgins, E. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within canadian indigenous communities. *Language Documentation & Conservation*, 3(1), 182-215.
- Davenport, L. D. 2016. The role of gender, class, and religion in biracial Americans' racial labeling decisions. *American Sociological Review*, 81, 57–84.
- Davis, Kathy. 2008. Intersectionality as buzzword: a sociology of science perspective on what makes a feminist theory successful. *Feminist Theory* 9(1) 67–85.
- Davis, C. S., Smith, V., Super, N., Super Sr, P., Thom Sr, C., & Mikkelsen, L. (2020). Forms and functions of backward resumption: The case of Karuk. *Language*, 96(4), 841-873.
- Delgado, R., & Stefancic, J. 2017. *Critical race theory: An introduction*. NYU Press.
- DeGraff, M. (2020). Toward racial justice in linguistics: The case of Creole studies (Response to Charity Hudley et al.). *Language* 96(4), e292-e306. doi:10.1353/lan.2020.0080.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. 2014. Reducing People's Reluctance to Respond to Surveys. In *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. 19–55. John Wiley & Sons.
- D'Onofrio, A. 2019. Complicating categories: Personae mediate racialized expectations of non-native speech. *Journal of Sociolinguistics*, 23(4), 346-366.
- Dovchin, S. (2019a). Language crossing and linguistic racism: Mongolian immigrant women in Australia. *Journal of Multicultural Discourses*, 14(4), 1-18.
- Dovchin, S. (2019b). The politics of inequality in translanguaging and linguistic discrimination. In T. Barrett & S. Dovchin (Eds.), *Critical inquiries in the studies of sociolinguistics of globalization* (pp. 84–102). Bristol: Multilingual Matters.
- Doyle, J. M., & Kao, G. 2007. Are racial identities of multiracials stable? Changing self-identification among single and multiple race individuals. *Social Psychology Quarterly*, 70(4), 405–423.

- Eckert, P. 1989. *Jocks and burnouts: Social categories and identity in the high school*. Teachers college press.
- Eckert, P. 2005. Variation, convention, and social meaning. In *Annual Meeting of the Linguistic Society of America*. Oakland CA, 7.
- Eckert, P. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41, 87-100.
- England, S. 1999. Negotiating race and place in the Garifuna diaspora: Identity formation and transnational grassroots politics in New York City and Honduras. *Identities Global Studies in Culture and Power*, 6, 5-53.
- Errington, J. 2001. *Colonial linguistics*.
- Espinosa, A., Tikhonov, A., Ellman, L., Kern, D., Lui, F., and Anglin, D. (2018). Ethnic identity and perceived stress among ethnically diverse immigrants. *Journal of Immigrant and Minority Health*, 20(1), 155-163.
- Etienne, A., Laroia, T., Weigle, H., Afelin, A., Kelly, S. K., Krishnan, A., & Grover, P. 2020. Novel Electrodes for Reliable EEG Recordings on Coarse and Curly Hair. *bioRxiv*.
- Feliciano, C. (2016). Shades of race: How phenotype and observer characteristics shape racial classification. *American Behavioral Scientist*, 60(4), 390–419. doi: 10.1177/0002764215613401
- Fishman, Joshua A., 1977. Language and Ethnicity. In *Language, Ethnicity and Intergroup Relations*, ed. H. Giles. 15-58. New York: Academic.
- Flores, N. 2016. A tale of two visions: Hegemonic whiteness and bilingual education. *Educational Policy*, 30(1), 13-38.
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. 2011. *Looking the part: Social status cues shape race perception*. PLoS One, 6, 1–10.
- Fought, C. 2003. *Chicano English in context*. New York: Palgrave Macmillan.
- Gaby, A., & Woods, L. (2020). Toward linguistic justice for Indigenous people: A response to Charity Hudley, Mallinson, and Bucholtz. *Language* 96(4), e268-e280. doi:10.1353/lan.2020.0078.
- García, J. D. 2020. Race and Ethnicity. In N. Brown, T. McIlwraith, & L. Tubelle De González (Eds.), *Perspectives: An Open Invitation to Cultural Anthropology*, 2nd edn. 444–455. American Anthropological Association. <http://perspectives.americananthro.org/>
- Gaither, S. E. 2015. “Mixed” results: Multiracial research and identity explorations. *Current Directions in Psychological Science*, 24(2), 114–119.
- Giddens, A. 1979. Positivism and its critics’. In T. Bottomore and R. Nisbet (eds.) *A History of Sociological Analysis*. London: Heinemann.
- Giles, H., & Johnson, P. 1987. Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language*, 68, 66-99.
- Good, J. J., Chavez, G. F., & Sanchez, D. T. 2010. Sources of self-categorization as minority for mixed-race individuals: Implications for affirmative action entitlement. *Cultural Diversity and Ethnic Minority Psychology*, 16(4), 453–460.
- Gordon, Rachel A. 2010. *Regression Analysis for the Social Sciences*. Routledge: New York.
- Gurin, P., Peng, T., Lopez, G., & Nagda, B. A. 2001. Context, Identity, and Intergroup Relations. In D. Prentice & D. Miller (Eds.), *Cultural Divides*. 133–171. Russell Sage Foundation.
- Hamilton, D. L. 2012. Natural ambiguities: Racial categorization of multiracial individuals. *Journal of Experimental Social Psychology*, 48, 152–164.

- Haslam, N., Rothschild, L., & Ernst, D. 2000. Essentialist beliefs about social categories. *British Journal of Social Psychology*, 39(1), 113–127.
- Hill, J. H. 1998. Language, race, and white public space. *American anthropologist*, 100(3), 680–689.
- Hilton, K. 2018. *What Does an Interruption Sound Like?*. [Doctoral dissertation, Stanford University]. ProQuest Dissertations Publishing.
- Hsu, B., & Syed, S. 2020. Cooccurrence of nonarticle determiners as evidence for split DPs and bundled projections. *Language* 96(3), 114–134.
- Hudley, A. H. C., Mallinson, C., & Bucholtz, M. (2020). Toward racial justice in linguistics: Interdisciplinary insights into theorizing race in the discipline and diversifying the profession. *Language*, 96(4), e200–e235.
- Hurtado, H., Gurin, P., & Peng, T. 1994. Social Identities—A Framework for Studying the Adaptations of Immigrants and Ethnics: The Adaptations of Mexicans in the United States. *Social Problems*, 41(1), 129–151.
- Johnson, Daniel Ezra. 2010. Why stepwise isn't so wise. Invited panel presentation at *New Ways of Analyzing Variation (NWAV)*, 39.
- Johnson, E. J., & Johnson, D. C. (2015). Language policy and bilingual education in Arizona and Washington state. *International Journal of Bilingual Education and Bilingualism*, 18(1), 92–112.
- Jones, N. A., & Bullock, J. 2012. The two or more races population: 2010. *2010 Census Briefs*. Retrieved from <https://www.census.gov/prod/cen2010/briefs/c2010br-13.pdf>
- Jones, N. A., & Smith, A. S. 2001. The two or more races population: 2000. *Census 2000 Brief*. Retrieved from <https://www.census.gov/prod/2001pubs/c2kbr01-6.pdf>
- Jørgensen, Anna, Hovy, Dirk and Søgaard, Anders. 2015. Challenges of Studying and Processing Dialects in Social Media. In *Proceedings of the Workshop on Noisy User-generated Text*, pages 9–18, Association for Computational Linguistics.
- Jurjevich, J. 2020. Race/ethnicity and the 2020 Census. *Census 20/20*. Retrieved from: <https://www.census2020now.org/faces-blog/same-sex-households-2020-census-r3976>.
- Kang, S.K., & Bodenhausen, G.V. 2015. Multiple identities in social perception and interaction: Challenges and opportunities. *Annual Review of Psychology*, 66. 547–574.
- Kendi, I. X. 2019. *How to be an antiracist*. One world.
- Kiesling, Scott F. 2011. *Linguistic Variation and Change*. Edinburgh: Edinburgh University Press.
- Kibrik, A. E. 1977 *The methodology of field investigations in linguistics*. Janua Linguarum, Series Minor. The Hague: Mouton.
- Koenecke, Allison, Nam, Andrew, Lake, Emily, Nudell, Joe, Quartey, Minnie, Mengesha, Zion, Toups, Connor, Rickford, John R, Jurafsky, Dan and Goel, Sharad. 2020. Racial Disparities in Automated Speech Recognition. *Proceedings of the National Academy of Sciences* 117(14), 7684–7689.
- Kubota, R. 2002. The author responds:(Un) raveling racism in a nice field like TESOL. *TESOL Quarterly*, 36(1). 84–92.
- Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. 2013. Subject/object processing asymmetries in Korean relative clauses: Evidence from ERP data: Color versions of Figures 2–4, 6–8. *Language*, 89(3), A1–A7.
- Labov, William. 1994 *Principles of Linguistic Change: Internal Factors*. Cambridge, MA: Blackwell.

- Labov, William; Sharon Ash; and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. De Gruyter.
- Labov, W., Cohen, P., Robins, C. & Lewis, J. 1968. A study of the non-standard English of Negro and Puerto Rican speakers in New York City. *Cooperative Research Report 3288. Vols. I and II*. New York: Columbia University. (Reprinted by U.S. Regional Survey, 3812 Walnut Street, Philadelphia, Pa. 19104.)
- Legate, J. A., Akkuş, F., Šereikaitė, M., & Ringe, D. (2020). On passives of passives. *Language*, 96(4), 771-818.
- Leonard, W. Y., & Haynes, E. 2010. Making “collaboration” collaborative: An examination of perspectives that frame linguistic field research. *Language Documentation & Conservation*, 4, 269-293.
- Levon, E. 2015. Integrating intersectionality in language, gender, and sexuality research. *Language and Linguistics Compass*, 9(7). 295-308.
- Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development*, 70(5), 1163-1180.
- Levy, S.R., Stroessner, S.J., Dweck, C.S. 1998. Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology*, 74, 1421–1436. doi: 10.1037/0022-3514.74.6.1421.
- Liebkind, K., Mähönen, T. A. E., Varjonen, S. A., & Jasinskaja-Lahti, I. 2016. Acculturation and Identity. In D. L. Sam, & J. W. Berry (Eds.), *Cambridge Handbook of Acculturation Psychology*. 2 edn. 30-49. Cambridge University Press.
- Linguistic Society of America. 2018. The state of linguistics in higher education: Annual report 2018. Online: <https://www.linguisticsociety.org/resource/state-linguistics-higher-education-annual-report>.
- Linguistic Society of America. 2019. Statement on Race. <https://www.linguisticsociety.org/content/lsa-statement-race>
- López, N., Vargas, D. E., Juarez, M. Cacari-Stone, L., & Bettez, S. 2017. What’s Your “Street Race”? Leveraging Multidimensional Measures of Race and Intersectionality for Examining Physical and Mental Health Status among Latinxs, *Sociology of Race and Ethnicity*. 4(1):49-66.
- Mahowald, K., Graff, P., Hartman, J., & Gibson, E. 2016. SNAP judgments: A small N acceptability paradigm (SNAP) for linguistic acceptability judgments. *Language*, 92(3), 619-635.
- Mallinson, C., & Charity Hudley, A. H. (2010). Communicating about communication: Multidisciplinary approaches to educating educators about language variation. *Language and Linguistics Compass*, 4(4), 245-257.
- Markus, H. R. 2010. Who am I?: Race, ethnicity and identity. In H. Markus & P. Moya (Eds.), *Doing race: 21 essays for the 21st century*. New York: W.W. Norton & Co.
- Mendoza-Denton, Norma. 2008. *Homegirls: Language and Cultural Practice among Latina Youth Gangs*. New York: Wiley-Blackwell.
- Miles-Hercules, D. 2020. Scoping the Scope: revisiting the aims and paradigms of Linguistics. Online webinar, Linguistic Society of America “This IS Linguistics: Scope, Positionality, and Graduate Apprenticeship when Diversifying the Linguistics Curriculum.” 11 September, 2020.
- Milroy, Lesley. 1987. *Observing and Analysing Natural Language: a critical account of sociolinguistic method*. Oxford: Basil Blackwell.

- Milroy, L. & M. Gordon 2003. *Sociolinguistics: method and interpretation*. Oxford: Blackwell.
- Motha, S. 2020. Is an antiracist and decolonizing applied linguistics possible? *Annual Review of Applied Linguistics*, 40, 128-133.
- Newman, M., & Wu, A. (2011). "Do you sound Asian when you speak English?" Racial identification and voice in Chinese and Korean Americans' English. *American speech*, 86(2), 152-178
- Nguyen, A., & Benet-Martinez, V. 2010. Multicultural identity: What it is and why it matters. In R. Crisp (Ed.), *The psychology of social and cultural diversity*. 87–114. Hoboken, NJ: Wiley-Blackwell.
- Noels, K. A. 2014. Language variation and ethnic identity: A social psychological perspective. *Language and Communication*, 35(1), 88–96.
- Noels, K. A., Leavitt, P. A., & Clément, R. 2010. "To See Ourselves as Others See Us": On the Implications of Reflected Appraisals for Ethnic Identity and Discrimination. *Journal of Social Issues*, 66(4). 740–758.
- O'Brien, Robert M. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41(5), 673-690.
- Office of Management and Budget. 1997. Revision to the Standards for the Classification of Federal Data on Race and Ethnicity. *Federal Register*, 62(210), Thursday, October 30.
- Pauker, K., Apfelbaum, E. P., & Spitzer, B. 2015. When societal norms and social identity collide: The race talk dilemma for racial minority children. *Social psychological and personality science*, 6(8), 887-895.
- Pauker K., Meyers C. K., Sanchez D. T., Gaither SE, Young DM. 2018. A review of multiracial malleability: Identity, categorization, and shifting racial attitudes. *Social and personality psychology compass*.
- Phinney, J. S., & Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 54(3), 271–281. doi: 10.1037/0022-0167.54.3.271
- Pew Research Center. 2015. "Multiracial in America: Proud, Diverse and Growing in Numbers." Washington, D.C.: June.
- Prentice, D. A., & Miller, D. T. 2007. Psychological essentialism of human categories. *Current Directions in Psychological Science*, 16(4), 202-206.
- Purdie-Vaughns, V., & Eibach, R. P. 2008. Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, 59(5–6), 377–391.
- Purnell, T., Idsardi, W., & Baugh, J. 1999. Perceptual and phonetic experiments on American English. *Journal of Language and Social Psychology*, 18(1), 10–30.
- Remedios, J. D., & Chasteen, A. L. 2013. Finally, someone who "gets" me! Multiracial people value others' accuracy about their race. *Cultural Diversity & Ethnic Minority Psychology*, 19, 453–60.
- Rockquemore, K. A., & Arend, P. 2002. Opting for white: Choice, fluidity and racial identity construction in post civil-rights America. *Race and Society*, 5(1), 49-64.
- Rockquemore, K. A., Brunsma, D. L., & Delgado, D. J. 2009. Racing to theory or retheorizing race? Understanding the struggle to build a multiracial identity theory. *Journal of Social Issues*, 65, 13-34.
- Rosa, J., & Flores, N. 2017. Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society*, 46(5), 621–647.

- Rzepnikowska, A. (2018). Racism and xenophobia experienced by Polish migrants in the UK before and after Brexit vote. *Journal of Ethnic and Migration Studies*, 45(1), 61-77.
- Sanchez, D. T., Shih, M., & Garcia, J. A. 2009. Juggling multiple racial identities: Malleable racial identification and psychological well-being. *Cultural Diversity & Ethnic Minority Psychology*, 15, 243–254.
- Sanchez, D. T., Chavez, G., Good, J. J., & Wilton, L. S. 2012. The language of acceptance: Spanish proficiency and perceived intragroup rejection among Latinos. *Journal of Cross-Cultural Psychology*, 43, 1019–1033.
- Sanchez, D. T., Shih, M., & Wilton, L. S. 2014. Exploring the Identity Autonomy Perspective: An integrative theoretical approach to multicultural and multiracial identity. In V. Benet-Martínez & Y. Y. Hong (Eds.), *The handbook of multicultural identity: Basic and applied psychological perspectives*, 139–159. New York, NY: Oxford University Press.
- Sap, Maarten, Card, Dallas, Gabriel, Saadia, Choi, Yejin and Smith, Noah A. 2019. The Risk of Racial Bias in Hate Speech Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678, Florence, Italy: Association for Computational Linguistics.
- Sen, M., & Wasow, O. (2016). Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics. *Annual Review of Political Science*, 19, 499–522. <https://doi.org/10.1146/annurev-polisci-032015-010015>.
- Shih, M., Bonam, C., Sanchez, D., & Peck, C. 2007. The social construction of race: Biracial identity and vulnerability to stereotypes. *Cultural Diversity & Ethnic Minority Psychology*, 13, 125–33.
- Shih, M., Sanchez, D. T., & Ho, G. C. 2010. Costs and benefits of switching among multiple social identities. In R. J. Crisp (Ed.), *Social issues and interventions. The psychology of social and cultural diversity*, 62–83. Wiley Blackwell.
- Smith, Gary 2018. Step away from stepwise. *Journal of Big Data*, 5(1), 32.
- Smith, Maya A. 2019. *Senegal Abroad: linguistic borders, racial formations, and diasporic imaginaries*. Madison: University of Wisconsin Press.
- Squizzero, Robert. 2020. Attitudes towards L2 Mandarin Speakers of Chinese and non-Chinese Ethnicity. Paper presented at the 32nd meeting of the North American Conference on Chinese Linguistics, Storrs, CT.
- Steele, C. M., Spencer, S. J., & Aronson, J. 2002. Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, 379-440. San Diego, CA: Academic Press.
- Tatman, Rachael. 2017. Gender and Dialect Bias in YouTube’s Automatic Captions. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 53–59, Association for Computational Linguistics.
- Tajfel, H., & Turner, J. C., 1979. An integrative theory of intergroup conflict. In W.G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations*, 33- 47. Monterey, CA: Brooks/Cole Publication Company.
- Townsend, S. Markus, H. & Bergsieker, H. 2009. My choice, your categories: The denial of multiracial identities. *Journal of Social Issues*, 65(1), 185-204.
- Tran, A. G. T. T., Miyake, E. R., Martinez-Morales, V., & Csizmadia, A. 2016. “What are you?” Multiracial individuals’ responses to racial identification inquiries. *Cultural Diversity and Ethnic Minority Psychology*, 22(1), 26–37.

- Wang, J., Minervino, C., & Cheryan, S. 2013. Generational differences in vulnerability to identity denial: The role of group identification. *Group Processes & Intergroup Relations*, 16, 600–617.
- Wassink, Alicia B. 2015. Sociolinguistic Patterns in Seattle English. *Language Variation and Change*, 27, 31–58.
- Wassink, A.B., Gansen, C., & Bartholomew, I. (under review) Uneven success: automatic speech recognition and ethnicity-related dialects, *Speech Communication*.
- Williams, M.J., & Eberhardt J.L. 2008. Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, 94, 1033–1047. doi: 10.1037/0022-3514.94.6.1033
- Wilton, L., Sanchez, D. T., & Giamo, L. 2014. Seeing similarity or distance? Racial identification moderates intergroup perception after biracial exposure. *Social Psychology*, 45(2), 127-134.
- Wolfram, W., & Schilling, N. 2016. *American English* (3rd ed., Language in society (Oxford, England); 25). Chichester, West Sussex, England: Wiley Blackwell.
- Yip, T. (2005). Sources of situational variation in ethnic identity and psychological wellbeing: A palm pilot study of Chinese American students. *Personality and Social Psychology Bulletin*, 31(12), 1603–1616.

Appendix 1: Author positionality statements including: linguistic subfields, race/ethnicity [and gender], and more

Dr. Emily Bender

I am a computational linguist with training in syntax and sociolinguistics, interested (among other things) in how a linguist’s understanding of language, including sociolinguistic variation, can be used to help mitigate the potential for harm inherent in machine learning systems trained over large datasets. I am also a white cis woman who has experienced racialization while living abroad (in Japan), but not the same kind of racialization that racialized groups are subjected to in the US. I am currently a member of the LSA’s Ethics Committee and also am co-chairing the Ethics Committee for NAACL 2021. The LSA committee is a general advisory committee, with the goal of “encourag[ing] linguists to actively engage with the ethical issues that arise in the course of their professional activities, and to foster increased discussion of ethics within the discipline.” ([source](#)) The NAACL committee is responsible for reviewing submissions to the conference that general review has flagged for ethical concerns. Its procedures are documented [here](#).

Dr. Kirby Conrod

I am a sociosyntactician, working primarily in variationist and Minimalist generative frameworks (and at their intersections); my primary research interests are in pronouns and morphosyntactic gender, and transgender and nonbinary issues in linguistics more generally. I am a white gay transgender nonbinary person. I grew up in California, lived in Taiwan for three years (in Hsinchu and Kaohsiung), and have lived in Seattle, WA for seven years. The racialization I experienced while living and working in Taiwan was primarily centered around dominant

ideologies of white supremacy in English language education, and was not comparable to the experiences of non-white foreign English teachers in those same contexts.

I am also autistic and disabled, which informs my experiences and perspectives around marginalized identities and relations. I am an active member of the LSA's Committee on LGBTQ+ [Z] Issues in Linguistics (COZIL), and am a temporary part-time lecturer at the University of Washington.

Martin Horst

I am a linguist. Currently working in the computational linguistics industry, I am curious about linguistic identity in multicultural settings, first and foreign language acquisition, and the intersection of sociolinguistics and NLP systems regarding racial bias. I identify as a white cis hetero man native to Seattle with a propensity towards foreign language acquisition. My limited time in Morocco afforded me a taste of racialization, but not the same kind of racialization native to the US. I humbly recognize that nearly every possible privileged identity is beholden to me, and am striving to use these privileges to learn and grow.

Monica Jensen

I'm a sociolinguist, syntactician and a sociosyntactician, working primarily in English on regional syntactic variation in the United States. Beyond that I'm a queer white cis woman. While my grandmother and her family were Hispanic, the racial and ethnic realities of early 20th century California meant that it was safer and easier for her to pass as white; subsequently my family only discovered our history later on. Since I have none of the culture or history and I'm very white-presenting, I don't see my Latine background as part of my identity; I do however spend probably too much time thinking about the benefits and detriments of passing as white in American culture.

Dr. Anna Kristina Moroz

I am a variationist sociolinguist with some training in phonetics. My current research focus is on applications of rootedness in sociolinguistics and understanding vowel systems in Southern Oregon. In general, I lean towards more qualitative methods because of my lived experience, even though I think there is a place for more quantitative methods. I was born and raised in the Pacific Northwest and I am a cis woman. My parents are from Poland and I spent a significant amount of time in Poland and was raised bilingual. I currently describe myself as American with a Polish bonus. I think in general, I have been told what my identity is regarding being Polish and American more often than I have had the opportunity to voice, and I often have my claim challenged when it is shared. I think my lived experience has resulted in me not wanting to participate in research studies or answer surveys. I often feel annoyed about the demographic portion due to questions regarding my languages and ethnicity. Being able to pick one option for what is your native language is hard and inaccurate when you have two native languages. Writing white, when you eat Polish food at home, speak Polish at home, and the decorations you

have are in other Polish homes often feels not specific enough. Not knowing how a researcher will interpret the most accurate phrase, “American with a Polish bonus,” makes me uncomfortable to write that. So I just choose to remove myself from situations where I get compartmentalized into a box. These are probably the kinds of scenarios that would be good for researchers to avoid.

Dr. Alex Panicacci

I am a Postdoctoral Researcher in receipt of a British Academy Fellowship, affiliated to the Linguistic Department at Queen Mary University of London and with the Psychology Department at the University of Washington. The various strands of my research are united by a fundamental interest in identity practices in modern multicultural societies. This has led me to work in the areas of multilingualism, sociolinguistics, cultural and social psychology, paying particular attention to emotion and affective socialization, acculturation, and personality in migration contexts. I am a cis white woman, and a first-generation college graduate. My academic path has been unusual under several aspects. I have studied different subjects and moved a lot during my educational up-bringing. I was born and raised in Italy, with an innate sense of *wanderlust*. I spent most of my adult life in London, a very multiethnic city, and I have now temporarily relocated in the Pacific Northwest. I have never been a proper full time scholar in the conventional sense. In the past, among other things, I worked as a Citizenship Consular Officer with the Italian Ministry of Foreign Affairs and International Cooperation and trained as a ballet dancer. Experiencing diversity in different forms and contexts changed me at a psychological and cognitive level. In this picture, my interest in linguistic, cultural, and racial diversity stems from first-hand experiences of excitement and struggle in managing my British-Italian identity and has been constantly nurtured by the variety of contexts I had the chance to be involved in.

Rob Squizzero

I am a variationist sociolinguist, phonetician, and language teacher with research interests at the convergence of language and psychology. My research focus came about from a deep fascination with the way that perceived social group membership can affect actual intelligibility and comprehensibility of spoken language. I currently work primarily with communities in the United States and in Greater China, and work on intersecting effects of ethnicity, regionality, gender, and perceived native speaker status on L1 English and L2 Mandarin Chinese. I am a gay White man who experienced a form of racialization while living in Taiwan, though this racialization differs in many ways from that experienced by minorities in ethnically heterogenous societies as well as by other ethnic minority group members in Taiwan. I identify ethnically as Italian-American and Irish-American and grew up just outside of Providence, Rhode Island, often curious about the world outside of my small (but unique) home State.

Dr. Alicia Wassink

I am a sociophonetician and creolist, with research interests in (among other things): vowel system dynamics, ethnicity-related bias in speech recognition systems, social network modeling, language attitudes, and the linguistic outcomes of both language and interethnic contact. I grew up between two communities: an urban, mixed Hasidic and West Indian working-class community in Philadelphia, and a rural working-class community outside of Kingston, Jamaica. When asked, I say that I am half Jamaican and half African-American, although my sister and I like to use a longer term when we talk to each other, saying we're "Jam-Afri-Merican" because labels are silly. I am a cis black woman. I am a first-generation college graduate. Growing up on the margins between my African-American and Jamaican communities, I regularly found (and still find) my claim to both identities contested. I learned young that the ability to use African American English and Jamaican Creole natively was a critical part of ingroup ethnic identification, but that this covert value is part of a complex attitude system, in which overt praise for these varieties is withheld and sometimes flatly denied. The desire to explore competing attitudes to language is why I became a linguist. I sit on the executive committees of the American Dialect Society and the Linguistic Society of America. In my neighborhood in a northern exurb of Seattle, I lead a community-based group exploring racial justice and reconciliation.

Appendix 2: Linguistics methods texts surveyed for guidance regarding treating race and ethnicity in study design

Item #	Editor/Author	Year	Title	Publisher
1	Winter	2020	Statistics for Linguists: An Introduction Using R	Routledge
2	Loerts	2020	Essential statistics for applied linguistics: using R and JASP	Red Globe Press
3	Rose	2020	Data collection research methods in applied linguistics	Bloomsbury
4	McKinley and Rose, eds.	2020	The Routledge handbook of research methods in applied linguistics	Routledge
5	Litosseliti	2018	Research Methods in Linguistics	Bloomsbury
6	Rauch & Carr	2018	Linguistic Method: Essays in Honor of Herbert Penzl	De Gruyter Mouton
7	Heller, Pietikäinen & Pujolar	2018	Critical Sociolinguistic Research Methods: Studying Language Issues That Matter	Routledge
8	Brezina	2018	Statistics in corpus linguistics: a practical guide	Cambridge
9	Drager	2018	Experimental research methods in sociolinguistics	Bloomsbury
10	de Groot & Hagoort, eds.	2018	Research methods in psycholinguistics and the neurobiology of language: a practical guide	John Wiley & Sons
11	Desagulier	2017	Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics	Springer
12	Bonami & Sagot	2017	Computational methods for descriptive and theoretical morphology: a brief introduction	Springer Verlag

13	Jenset	2017	Quantitative historical linguistics: a corpus framework	Oxford
14	Ranganathan, ed.	2017	The Bloomsbury research handbook of Indian ethics	Bloomsbury
15	Zhu, ed.	2016	Research methods in intercultural communication: a practical guide	Wiley Blackwell
16	Paltridge and Aek eds.	2015	Research Methods in Applied Linguistics : a Practical Resource	Bloomsbury
17	Meyerhoff, Schleaf & MacKenzie	2015	Doing Sociolinguistics: A practical guide to data collection and analysis	Taylor and Francis
18	Levshina	2015	How to do linguistics with R: data exploration and statistical analysis	John Benjamins
19	Orfanidou, Woll, & Morgan, eds.	2015	Research methods in sign language studies: a practical guide	Wiley-Blackwell
20	De Costa, Peter I	2015	Ethics in Applied Linguistics Research: Language Researcher Narratives	Routledge
21	Connor-Linton & Amoroso, eds.	2014	Measured language: quantitative approaches to acquisition, assessment, and variation	Georgetown University Press
22	Cohen & Demner-Fushman	2014	Biomedical natural language processing	John Benjamins
23	Holmes & Hazen	2013	Research methods in sociolinguistics: a practical guide	Wiley
24	Xu	2013	Linguistic Decision Making: Theory and Methods	Springer-Verlag
25	Hemforth	2013	Experimental Linguistics	Oxford University Press
26	Krug & Schlüter, eds.	2013	Research methods in language variation and change	Cambridge

27	Podesva & Sharma, eds.	2013	Research methods in linguistics	Cambridge
28	Müller & Ball	2012	Research Methods in Clinical Linguistics and Phonetics: A Practical Guide	Somerset: Wiley-Blackwell
29	Ender, Leemann, & Wälchli	2012	Methods in contemporary linguistics	De Gruyter Mouton
30	Graeme Porte, ed.	2012	Replication research in applied linguistics	Cambridge
31	Thieberger, ed.	2012, 2018 (online)	The Oxford handbook of linguistic fieldwork	Oxford University Press
32	Hoff, ed	2012	Research methods in child language: a practical guide	Wiley-Blackwell
33	Faust	2011	The Handbook of the Neuropsychology of Language	Wiley-Blackwell
34	Coffin, Lillis, & O'Halloran	2010	Applied linguistics methods: a reader (subtitle: systemic functional linguistics, critical discourse analysis and ethnography)	Routledge
35	Grenoble & Furbee, eds.	2010	Language documentation: practice and values	John Benjamins
36	Franzosi	2010	Quantitative narrative analysis	Sage
37	Gries	2009	Quantitative Corpus Linguistics with R: A Practical Introduction	Taylor and Francis
38	Johnson	2008	Quantitative Methods In Linguistics	Wiley
39	Li & Moyer	2008	The Blackwell Guide to Research Methods in Bilingualism and Multilingualism (and companion text)	Wiley-Blackwell
40	Rasinger	2008	Quantitative research in linguistics: an introduction	Continuum

41	Centre for Applied Linguistics	2008	Readings in quantitative linguistics - International Symposium on Linguistics, Quantification, and Computation	Indian Institute of Language Studies
42	Baayen	2008	Analyzing linguistic data: a practical introduction to statistics using R	Cambridge
43	Hale	2007	Historical Linguistics: Theory and Method	Wiley-Blackwell
44	Dörnyei	2007	Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies	Oxford University Press
45	Bayley & Lucas	2007	Sociolinguistic Variation: Theories, Methods, and Applications	Cambridge
46	Milroy & Gordon	2003	Sociolinguistics: method and interpretation	blackwell
47	Bod, Hay & Jannedy, eds.	2003	Probabilistic linguistics	MIT
48	Čermák	2002	Research methods in linguistics: essential principles, based on a general theory of science	Karolinum Press, Charles University
49	Paolillo	2002	Analyzing linguistic variation: statistical models and methods	Center for the Study of Language and Information
50	McMenamin, ed.	2002	Forensic linguistics: advances in forensic stylistics	CRC Press
51	Wengraf	2001	Qualitative research interviewing: biographic narrative and semi-structured methods	Sage
52	Wetherell, Taylor & Yates, eds.	2001	Discourse as data: a guide for analysis	Sage
53	Manning	2000	Foundations of statistical natural language processing	MIT Press

54	Campbell	1997	American Indian languages: the historical linguistics of Native America	Oxford
55	Klavans & Resnik	1996	The balancing act: combining symbolic and statistical approaches to language	MIT
56	Těšitelová	1991	Quantitative linguistics	John Benjamins
57	Švejcer	1986	Contemporary Sociolinguistics: Theory, Problems, Methods	John Benjamins
58	Woods	1986	Statistics in language studies	Cambridge
59	Bell	1981	An introduction to applied linguistics: approaches and methods in language teaching	Batsford
60	Politzer	1972	Linguistics and applied linguistics: aims and methods	Heinle & Heinle Enterprises
61	Zellig	1951	Methods in structural linguistics	University of Chicago Press