How and When to Make Causal Claims Based on Race or Ethnicity*

Maya Sen†‡  Omar Wasow§
msen@ur.rochester.edu  owasow@princeton.edu

August 4, 2012

Abstract

Causal inference is considered the gold standard in social science research. Making causal claims about “immutable characteristics” such as race, however, has been strongly discouraged. In contrast to previous literature, which assumes a fixed conception of race, we propose a different framework that in some cases reconciles race and causation. First, we distinguish those units of analysis in which intrinsic problems of race and causality can be avoided. Second, we demonstrate that race can be defined as a composite measure that has some mutable elements. These extensions allow us to synthesize two areas where causal claims about race may be permissible: (1) studies that measure the effect of exposing an entity to a racial cue and (2) studies that disaggregate race into constituent pieces and measure the causal effect of some mutable element. We demonstrate these techniques via examples from contemporary scholarship.

*Comments and suggestions welcome. We are grateful to Bear Braumoeller, John Bullock, Kevin Clarke, Adam Glynn, Jim Greiner, Jennifer Hochschild, Luke Keele, Gary King, Kevin Quinn, Teppei Yamamoto, Matt Blackwell, Jennifer Brea, Rich Nielsen, and Shauna Shames for thoughtful advice and suggestions. We also thank participants at the 2011 Midwest Political Science Association panel on causal inference for helpful feedback.

†Corresponding author
‡Assistant Professor, Department of Political Science, University of Rochester, Harkness Hall 307, Rochester, NY 14627 (http://www.mayasen.org).
§Department of Politics, Princeton University, 130 Corwin Hall, Princeton, NJ 08544 (http://scholar.harvard.edu/owasow).
1 Introduction

For many political scientists, making causal inquiries about the role of race and ethnicity lies at the core of key research questions. Do majority-minority districts maximize substantive representation? What causes members of certain ethnic groups to be more likely to join rebel militias? Did Barack Obama’s race cause him to lose votes in the 2008 presidential election? Researchers from fields as disparate as labor economics, race and ethnic politics, and public policy have for years focused their exclusive attention on the causal role of race or ethnicity. This has been the case even if the “causal” claim (or language) is not explicit.

On the other hand, scholars of causality have long warned against making any kind of causal inference on the basis of race or ethnicity. Why? Race is commonly understood as an immutable or unchanging characteristic. For centuries, societies have categorized people into different racial and ethnic groups, and a person’s race is generally thought to be resistant to manipulation. Within this fixed framework, race is defined at birth (if not earlier), largely determined by biology and impossible to assign randomly – the key tool for experimental causal inference. Furthermore, because factors like education and class are intimately linked to the distinct historical experiences of each racial and ethnic group, any attempt to isolate a single causal effect associated with race is likely to be deeply confounded. Consequently, experts on causality have long warned that making causal inferences about race or race-based variables is usually a misguided enterprise.

This paper attempts to bridge the divide between applied researchers and scholars of causal inference. In contrast to the immutable characteristics approach, we propose an alternative framework that demonstrates that under certain conditions it is possible to make
causal claims about race, ethnicity, or other seemingly fixed attributes. First, researchers must theorize about the appropriate unit of analysis and, where appropriate, analogize to experimental approaches. Second, empirical work ought to address carefully how race is defined. Within quantitative social science, race is usually a single, monolithic, categorical variable. But as scholars from a wide range of fields have shown (e.g., Appiah (1986); López (1994); Holland (2003)), racial and ethnic categories are typically the product of a complicated amalgam of social, cultural, historical, biological, geographical and legal influences. Rather than gloss over the complex challenges of measuring race (as frequently occurs in quantitative scholarship), we argue in favor of exploiting its composite nature by disaggregating race into constituent elements, some of which can be reasonably manipulated to estimate causal effects.

In particular, by looking at studies that have successfully made causal claims about race, this paper demonstrates that making such inferences is possible provided that researchers understand that race and ethnicity are composites, have some mutable elements, and are divisible into constituent elements. Thinking more flexibly about these considerations opens up the possibility of making causal claims about immutable characteristics like race or gender. To be clear, this is it not to say that making these kinds of causal claims is possible in all instances; what we demonstrate in this paper that there are some limited areas in which making causal claims about race or ethnicity is possible (and has been successful in the past). These include (1) studies that measure the effect of exposing an individual or institution to some signal about race and (2) studies that disaggregate race into constituent pieces and attempt to measure the effect of some mutable element of race within a single racial group. The later sections of the paper develop these ideas through two empirical examples. Throughout, we draw attention those social science studies that have been successful in meeting these

\[1\] Although our focus is race and ethnicity, much of the framework we present also applies to gender, sex, sexual orientation and other seemingly immutable characteristics. We attempt throughout to flag instances where the analogy is particularly appropriate (or inappropriate).
requirements, thereby providing a unifying research rubric for political scientists studying race and ethnic politics through a causal lens.

2 Causal Inference and Potential Outcomes

A brief overview of the Rubin potential outcomes framework helps contextualize the following discussion. (The literature on this topic is voluminous — e.g., Rubin (1974), Holland (1986), Angrist, Imbens and Rubin (1996), Rubin (2005) — and we attempt in this discussion only a bare-bones introduction.) At its core, a causal inquiry involves unpacking the effect of some treatment on some outcome.\(^2\) Does a vaccine cause people to live longer? Is a worker training program effective in helping people go back to work? In all of these cases we see (1) a unit of analysis, (2) a manipulable treatment, and (3) a specific outcome.

Using some simple notation, for unit \(i\), we are ultimately interested in the effect of some treatment, \(T_i\), on an outcome, \(Y_i\). Assuming a binary treatment variable (which does not always have to be the case), a unit \(i\) can either be treated \((T_i = 1)\) or not treated \((T_i = 0)\), yielding two “potential outcomes,” \(Y_i(T_i = 1)\) and \(Y_i(T_i = 0)\). In an ideal world, we would estimate the true treatment effect, defined as the simple difference between the two potential outcomes for unit \(i\):

\[
Y_i(T_i = 1) - Y_i(T_i = 0)
\]  

(1)

The “fundamental” problem of causal inference is, however, that we can never observe the difference between \(Y_i(T_i = 1)\) and \(Y_i(T_i = 0)\) (Holland, 1986; Rubin, 1978). That is, once we assume no convoluted theories involving time travel, unit \(i\) simply cannot receive both the treatment and the control at the same time. This problem extends to all kinds of inquiries — for example, when testing different medicines, or seeing the effects of a work training program.

\(^2\)We note that many interesting questions can be asked and answered with non-causal questions, and much of what we say here (particularly with regard to disaggregating components of race) applies equally to non-causal research designs.
– but it becomes particularly vexing when it comes to immutable characteristics. After all, a person cannot experience the world as being only black and also as being only white, or as being only Native American and as being only Maori, and to think otherwise raises strange hypotheticals. This is an important point to which we return throughout.

In lieu of trying to estimate an unobservable true treatment effect, those interested making causal inferences usually estimate some version of the average treatment effect \( E[Y_i(T_i = 1)] - E[Y_i(T_i = 0)] \), that is, the difference between the outcome means in treated and control populations. An obvious problem is, however, that differences in the outcome variable could be due to inherent differences between the treated and control populations, a problem that some refer to as selection bias (Angrist and Pischke, 2009). For example, we should not be surprised to see that workers who have signed up for a worker training program are more successful in getting jobs – but we also should not be surprised that they are also more ambitious and better educated than non-trained workers, two attributes that would also result in more favorable employment decisions.

The problem is solved in some circumstances by comparing only similarly situated treated and control units. Let \( X_i \) represent the background variables that could affect both the probability of receiving treatment or the eventual outcome. To get at a satisfactory estimate of the average treatment effect, we would like our treatment and control groups to be so similar across \( X \) such that that the only difference between the two groups is that one received the treatment and the other did not. This ignorability assumption is stated as:

\[
P(Y_i(T_i = 1), Y_i(T_i = 0)|X_i, T) = P(Y_i(T_i = 1), Y_i(T_i = 0)|X_i),
\]

i.e., if \( Y_i(T_i = 1) \) and \( Y_i(T_i = 0) \) are independent of \( T_i \), conditional on the covariates \( X_i \) (Holland, 1986). In plain English, the treatment assignment must be independent of the potential outcomes in order for us to assume that the two groups are similar enough to
Many empirical efforts are geared toward satisfying the ignorability requirement – that is, to make the treated and control populations as similar as possible so that the treatment regime could be assumed to be random. By far the easiest way to satisfy the ignorability requirement is simply to assign the treatment randomly – for example, by conducting a randomized experiment. (We discuss some successful experimental designs below; more general discussions are found in Holland (1986) and Imai, King and Stuart (2008).) However, because randomization is rarely an option for political scientists (especially, as we note, for those studying race or ethnicity), researchers have turned to matching or controlling for observed variables to satisfy the ignorability assumption that the treated and control groups are identical on background covariates (Dehejia and Wahba, 2002; Sekhon, 2009). We discuss these the utility of these techniques for race and politics scholars, below.

3 How Potential Outcomes Break Down With Race

Given these key basics of the potential outcomes framework, we now turn to exploring why race and ethnicity present such causal bugaboos. The existing literature has identified two key problems (Greiner and Rubin, 2010): (1) biological elements of race are resistant to manipulation (yielding experimental analogies unidentified), and (2) because race is generally understood to be “assigned” at birth (or conception), the host of characteristics for which most social scientists control (e.g., education, income, etc.) occur after the treatment is assigned and therefore have the potential to introduce post-treatment bias. In addition, we introduce a third problem: (3) building on the idea that race should be viewed as an amalgam of characteristics or a “bundle of sticks,” researchers often misclassify what the race variable actually represents, thus confusing the nature of the “treatment.” We address these three problems and their implications for potential outcomes in turn.
Problem 1: Race Cannot be Manipulated. Making causal inferences usually demands a neatly defined, manipulable treatment variable, one that can be easily documented and manipulated by researchers. Holland (1986), for example, famously admonishes “No causation without manipulation” to bring attention to the idea that all pertinent potential outcomes must be defined in principle in order to make causal estimates possible in practice. Further, to define all potential outcomes, one must be able to conceptualize an experimental analogy that would lead to the possible outcomes. In other words, as Holland puts it, “causes are only those things that could, in principle, be treatments in experiments.” This idea of a manipulative treatment is echoed by others like Cook, Campbell and Day (1979), who argue that “causation implies that by varying one factor I can make another vary”; Pearl (2000), who discusses at length the importance of an intervention in estimating causal treatments; and Gelman and Hill (2007), who warn that “a causal effect needs to be defined with respect to a cause, or an intervention, on a particular set of experimental units.”

The biological dimensions of race and gender are, however, resistant to manipulation. (Imbens and Rubin (2010) refer to them as “currently immutable characteristics,” as future scientific innovations may dramatically ease the effort required to change to certain biological aspects of race or gender – we leave it to science to develop these sorts of techniques.)
Treatment by race and gender also suffer from the problem that it is difficult to think about appropriate counterfactuals. We can imagine how someone lives their life as an African American; much more difficult is imagining what experiment one would design to manipulate the person’s race (and only the person’s race) to check its effect on some outcome. Thus, not only is randomization, the most elegant solution to the fundamental problem of causal inference, beyond the reach of many scholars focusing on the effects of race or ethnicity, but it is difficult in many instances to even conceptualize what an ideal experiment would look like. (We discuss exceptions below.) Ultimately, as Angrist and Pischke (2009) point out, research questions for which there are no experimental analogies (even hypothetical ones, in a world with unlimited time, research budgets, and mildly omniscient powers) are “FUQd” – Fundamentally Unidentified Questions.

Within the causal literature, the immutable (i.e., resistant to manipulation) nature of race and gender has led many to cite race and gender as attributes for which causal inferences are impermissible (e.g., Holland (1986); Rubin (1978); Gelman and Hill (2007)). As noted by Holland (1986): “For causal inference, it is critical that each unit be potentially exposable to any of the causes. As an example, the schooling a student receives can be a cause, in our sense, of the student’s performance on a test, whereas the student’s race or gender cannot.” A more specific admonishment on the topic of gender-based causality is given by Rubin (1978):

[C]onsider the causal effect of sex (male-female) on intelligence. What are the actions to be applied to each experimental unit that define the treatments? Are we to give hormone shots beginning at birth and surgically perform a “sex-change” operation, or at conception “change” Y-chromosomes and X-chromosomes? Even if an “at-conception X- for-Y chromosome change” becomes possible, presumably there will be several techniques developed for effecting the change with potentially different causal effects. Without treatment definitions that specify actions to be performed on experimental units, we cannot unambiguously discuss causal effects of treatments.
Problem 2: With Race, Everything is Post-Treatment. A problem secondary to conceptualizing well-defined potential outcomes is that a person’s race is “assigned” by a combination of social and biological processes at conception or birth. Thus, the host of background covariates that social scientists usually control for or match on (e.g., education, income, age) are determined after a person’s race is assigned.

Taking into account things that happen after the treatment happens or is administered has the potential of introducing post-treatment bias, a pervasive problem within observational social science research (King, Keohane and Verba, 1994; Rosenbaum, 2002). To use a common example, suppose that we are interested in the causal effect of smoking on death, and have a population of randomly assigned smokers and randomly assigned non-smokers. Should we control for lung cancer in the final analysis? Probably not: lung cancer is not only highly predictive of death, but it is also a direct consequence of smoking – probably the key consequence. If we controlled for lung cancer, the effect of smoking on death would essentially be nil, biased downward by the fact that we have controlled for its primary consequence. Race is obviously different from smoking, but the post-treatment issue applies with equal or greater force: race affects deeply how a person is raised and educated, what kinds of employment opportunities (and hence employment experiences) he or she will have, and what kind of cultural and social attitudes he or she will bring to the table. Including any of these attributes would therefore affect our estimates of the causal effect of “race.”

Although perhaps unsatisfactory to many applied researchers, the most appropriate initial approach is to drop any post-treatment variables from an analysis (King, 1991; King, Keohane and Verba, 1994; King and Zeng, 2006; Gelman and Hill, 2007). Thus, in the race context, any factor, attribute, personality trait, or personal or professional experience that could potentially be a consequence of race should be dropped – a practice that would eliminate most of the variables included as standard controls by social scientists. For example, if we were studying the effect of race on employment, we would not control for anything
directly impacted by the subject’s race, e.g., age, education level, income, felon status, zip code, health status, etc. The right-hand side of a regression would simply include race and, possibly, sex.\(^5\) We note that this strategy implies that the researcher is interested in the *total* effect of race (VanderWeele and Hernán, 2012) – which might not be satisfying to both researchers (and reviewers unfamiliar with the causal literature). However, there may be instances where this is not the case, and the researcher is interested in the effects of constituent components of race; we discuss this below. This kind of research design still also fails to address the critique above that experimental analogies are undefined.

Even aside from the post-treatment issue, we note two further problems with controlling for race-related covariates: (1) common support problems and (2) problems with multicollinearity. The common support problem arises when researchers include attributes that vary according to race (e.g., welfare status, participation in programs like Head Start, diseases such as sickle cell anemia or Tay Sachs). Because these traits vary almost exactly according to race, it becomes difficult to find non-minority (or minority) counterparts with which to compare the population of interest. (For example, it would be hard to find a group of whites who have sickle cell anemia (Thomas and Zarda, 2011).) Collinearity becomes a problem when variables or effects vary so closely with race so as to result in (the most extreme case) unconverged calculations of point estimates. The lack of variance in the background variables may also result in small changes having a large impact on the coefficient estimates – thus, standard errors may be large and lead researchers to assume no treatment effects when treatment effects do in fact exist.

---

\(^5\)Sex, which is also assigned at birth, is one of the few standard control variables that could be construed as being pre-treatment or, at the very least, assigned concurrently with the treatment. Evidence suggests, however, that sex ratios can vary by latitude, religion, ethnicity and other factors collinear with race (Gutenberg and Secord, 1983; Navara, 2009). Other possibly pre-treatment factors (e.g., genotype) are discussed in VanderWeele and Hernán (2012).
**Problem 3: Race is Actually a Multi-Faceted Treatment.** The prior two methodological problems associated with race are well-known and well-cited in the causal inference literature. We now introduce a third problem, which is the fact that race is rarely a single, easily defined measure (or treatment). To the contrary, the work of race and ethnic politics, sociologists, anthropologists, and critical theorists (e.g., Appiah (1986); López (1994)) has emphasized repeatedly that race is a complicated amalgam of things like skin color, cultural traits, physical attributes, and educational and sociological components – many of which vary across members of a single group. In other words, race is a composite of many component pieces; metaphorically, it is a “bundle of sticks” (Figure 1). Researchers could never assign the full bundle of factors that constitute a racial identity to some subjects while assigning others to a control; neither could they safely assume that someone identifying with one racial group uniformly shares all components of that racial categorization. It would, for example, be strange to assume that a Mexican American from Los Angeles shares all component attributes of being “Latino,” or shares them to an equal extent, as a Puerto Rican American from the Bronx.

Despite the variable, multi-faceted aspect of race, most quantitative social science nonetheless represents race via a single indicator dummy variables (“0” if white, “1” if black), categorical variables (“1” if white, “2” if black, “3” if Latino, etc.) or percentages (e.g. percent of a majority-minority district that is black). The problem of using such monolithic measures of race is twofold. First, any statistical association will typically offer little or no insight as to which element of race is the key mechanism of action – be it skin color, education, ethnic slang, discrepancies in health measures, facial features, etc. Second, a simple dummy in no way captures the fact that the treatment (“race”) can vary in its components quite radically from observation to observation – for example, our comparison between Mexican Americans versus Puerto Rican Americans. This represents possible ongoing violations of the stable unit treatment assumption (SUTVA), and calls into question whether an accu-
rate estimate of a race-based treatment effect is actually possible. Quite simply, forcing something as complicated as race into simplified categorical variables potentially introduces severe measurement error.

4 Research Design 1: Treatment as Exposure to Race

Although the problems cited by the causal inference literature can never be fully solved, they might in some instances be circumvented with the correct research design. We first consider a type of research design that we call “exposure to race” or exposure studies. These studies examine how subjects respond when exposed to some sort of racial signal or cue and, as such, might be more precisely called “exposure to a racial signal” or “exposure to a racial cue” studies. (These sorts of research designs have been described by Greiner and Rubin (2010) as those that look at the effects of “perceived race” and by VanderWeele and Hernán (2012) as those that look at discrimination.) Among the studies included in this
group would be those that look at how voters respond when presented with advertisements showing black versus white candidates, or those that examine whether employers are more or less likely to interview job applicants with traditionally African American names.

In these sorts of research designs, (1) the treatment of interest is the exposure to a racial cue and (2) the unit of analysis is the individual or institution being exposed; both alleviate the problems of race and causality. Using a simple employment example, suppose we are interested in racial discrimination against black job candidates in today’s job market. We can easily think of an ideal experiment, which would take two applicants, one white and one black, and construct a job profile that is exactly the same for each applicant – except for some signal or cue (such as a name or picture) that one applicant is black and one is white. (Bertrand and Mullainathan (2004), for example, relied on distinctively African-American names to send a signal about the applicant’s race.) The researcher would then send these job applications to employers and check the eventual hiring decisions. A difference would suggest that similarly situated blacks and whites are being treated differently, while no difference would suggest no discrimination. Again, key to this kind of study is that the unit of analysis is actually the prospective employer \( \textit{not} \) the prospective employee, and the treatment is the kind of name attached to the job application. Thus, the research design begins with well-defined potential outcomes, is operationalized via a clean experiment (or a clean experimental analogy), and has a precise moment of treatment. The causal impact of race and ethnicity is identified, alleviating the problems cited above.

\textbf{Experimental Exposure Studies.} These kinds of audit or correspondence studies have been used experimentally to measure race and gender discrimination in a wide variety of fields,\(^6\) including labor economics, (Bertrand and Mullainathan, 2004), sociology (Bobo and Pager (2007) provides a good overview of the literature, critiques, and methods. Although the exact methodology may vary, audit studies usually involve confederates or actors hired by researchers who are then randomly sent out to the field – for example, to different employers or to different lending agents. Partly in response to critiques about potential bias introduced by the confederates (Heckman and Siegelman, 1993;
Johnson, 2004), psychology (Cosmides, Tooby and Kuzban, 2003; Boker et al., N.d.; Steele, 1997; Greenwald, McGhee and Schwartz, 1998). Within political science, a robust public opinion literature (Miller and Krosnick, 2000; Huber and Lapinski, 2006; Valentino, Hutchings and White, 2002; Mendelberg, 2001; Sniderman and Piazza, 1993; White, 2007) has exploited some variant of the exposure research design to estimate race-based causal effects. Sniderman and Piazza (1993), for example, leverage question order to find that the “mere mention” of affirmative action to white survey respondents provokes more negative feelings towards blacks. Mendelberg (2001) creates simulated television news experiments to assess how racial cues might prime racial attitudes among white voters. Within the participation literature, Butler and Broockman (2011) use distinctively black and white names to craft putative “constituent” emails to state legislators; emails from white “constituents” were more likely to be answered by white representatives. Within psychology, Kurzban, Tooby and Cosmides (2001) expose subjects to photos and text to simulate a cross-race conversation and Steele (1997) identifies how internalized racial stereotypes affect women and racial minorities by exposing them to racial and gender cues immediately prior to a mathematics exam. Kurzban, Tooby and Cosmides and Steele also demonstrate how the exposure model can address questions other than concerns about a discriminatory “decisionmaker” (to use the terminology of Greiner and Rubin (2010)).

To step back a moment, although all of these studies exploit different techniques – from simulated avatars to scenarios in surveys – the general approach is the same: randomly present a subject with information that differs only with respect to signals or cues about race. (Though exposure to race is a useful shorthand, it is important to note that the correspondence studies were developed in which matched human applicants were replaced with matched pairs of “paper” applicants. Bertrand and Mullainathan (2004), as noted above, randomly assigned traditionally white and black names to otherwise similar resumes to assess how such signals about the race of the applicant affected hiring decisions. More recently, Adida, Laitin and Valfort (2010) used a similar technique to measure employment discrimination against Muslims in France. Though scholars have viewed audit and correspondence studies as related, we note that all studies employing exposure to a racial cue should be viewed as related and part of a common literature on race and causation. (Heckman, 1998), correspondence studies were developed in which matched human applicants were replaced with matched pairs of “paper” applicants. Bertrand and Mullainathan (2004), as noted above, randomly assigned traditionally white and black names to otherwise similar resumes to assess how such signals about the race of the applicant affected hiring decisions. More recently, Adida, Laitin and Valfort (2010) used a similar technique to measure employment discrimination against Muslims in France. Though scholars have viewed audit and correspondence studies as related, we note that all studies employing exposure to a racial cue should be viewed as related and part of a common literature on race and causation.)
treatment is never race in full (i.e., the whole “bundle of sticks”) but rather only an element of race such as name or physical appearance.) Research designs of the exposure type thus have (1) a randomly assigned treatment, which is the racial signal or cue, and (2) a unit of analysis, which is the subject being exposed to the racial cue. And, as a result, we have (3) well defined potential outcomes and (4) a precise, well defined moment of the treatment is assigned. Accordingly, the causal role of race is appropriately identified, provided thought has been given to the specific experimental design.

**Observational Exposure Studies.** It is possible to import this research design to a wide variety of observational contexts involving how third parties react to once they are exposed to racial signals and cues. In this sense, we could use observational data to understand how mortgage lenders react to Asian American versus white borrowers (Sen, 2012), how juries react to Hispanic versus non-Hispanic death penalty defendants (Greiner and Rubin, 2010), how voters respond to political ads featuring black versus white actors, how universities respond to minority versus non-minority applicants, and how the U.S. government reacts to proposals submitted by minority-owned business in deciding to award contracts. In all of these instances, the interest lies in understanding how exposure to race changes or informs others’ opinions, behaviors, or attitudes – a fact that makes this kind of research design ideal for testing implicit bias or racial discrimination (Greiner and Rubin, 2010; VanderWeele and Hernán, 2012).7

With observational data, researchers must be aware of two attendant issues. First, using

7Greiner and Rubin (2010) refer to this type of observational research as teasing apart the effects of “perceived” race (as opposed to actual race). We use different terminology and draw different analogies, but the research design we suggest here is comparable. Nonetheless, we move away from the “perceived race” language for two reasons. First, we think the best way to think about the “treatment” in these kinds of studies is not as perception but, instead as a signal about race. After all, in an experimental context, the researcher can manipulate the signal to which the subject is exposed but not what the subject actually perceives. Second, perceived race is rarely observed: what a subject perceives occurs within the confines of a mind and is opaque to researchers. As such, focusing on exposure to race rather than perception of race is preferable.
observational data means that researchers lack the ability to manipulate the racial cues and signals received by the subject. It is therefore necessary to include those background variables in the analysis such that the only functional difference between the treated and control groups is that one group is exposed to minority, or other racial cues and that the other is exposed to non-minority racial cues. Second, and perhaps more helpfully, the exposure research design greatly lessens problems of post-treatment bias (Greiner and Rubin, 2010). To illustrate, suppose we are interested in whether a university accepts minority versus non-minority applicants at different rates – perhaps due to affirmative action but, perhaps also, due to discriminatory motivations. The ideal experiment here would be to mimic an audit study and create identical applicants whose profiles differ only with regard to their race. The “treatment” would be administered to the admissions officer at the time he or she reviews the application packet. Anything that happens before is solidly pre-treatment and must be conditioned on; anything that happens after (e.g., decisions about financial aid, work study opportunities) would be post-treatment and should be dropped from the statistical model (Greiner and Rubin, 2010).

This discussion can be boiled down to one key idea: when possible, conceptualizing an experiment or observational study as an exposure study greatly reduces both the theoretical and practical problems associated with making race-based causal inferences. Thus, applied researchers should think carefully about whether an exposure study could provide a well-suited analogy for their research questions and hypotheses. We also note that this is a research design that is particularly apropos to questions involving racial discrimination, disparate treatment, and priming.
5 Research Design 2: Element of Race Designs

Exposure studies offer a useful framework when individuals or institutions have been presented with some signal about race – i.e., we are interested in discrimination, bias, or disparate treatment. However, many research questions do not involve an external actor exposed to a racial cue: Why do African Americans suffer from increased rates of heart disease? Why are certain ethnic groups overrepresented in rebel militias? What explains the educational “achievement gap” between blacks and whites? In these studies, there is generally no treatment by exposure and no “decisionmaker.” (VanderWeele and Hernán (2012) refer to these studies as those focusing on “discrepancies.”) For scholars working on these sorts of topics, the primary research interest – and the appropriate units of analysis – lies in a particular racial or ethnic population itself. And these studies are particularly problematic in terms of having ill-defined potential outcomes and also having post-treatment bias problems.

For these sorts of questions, we suggest a different research design, one that exploits variation within a racial group to extract causal inferences. We call these “element of race” designs. This kind of research design disaggregates the “bundle of sticks” discussed above and singles out a specific element of race that can be manipulated in an experiment (or observed to vary) within a racial group. By identifying a mutable element of race, it is possible to identify well-defined potential outcomes and to assuage potential post-treatment bias problems. The bundle of sticks thus becomes a blessing instead of a curse.

For example, suppose we are interested in understanding disparate educational outcomes for black versus white inner-city youngsters. A naive analysis would be to regress educational outcomes on race (possibly other control variables), taking the group of African Americans as the “treated” group and whites as the “control.” For all the reasons cited above, however, a causal estimate based on this research design would be (1) fundamentally unidentified and (2) biased by any inclusion of post-treatment variables. A better research design would be
one that takes as its starting point the fact that the race variable captures a variety of factors, and, rather than conceive of black youngsters as a treated group and white youngsters as the control, identifies a trait highly collinear with race that is in theory manipulable. One example of such a trait might be the mother’s welfare status. With this in mind, we can re-cast the study as a *within-group analysis* where we compare black youngsters with mothers on welfare versus those black youngsters with non-welfare mothers. (We could include white children, but the cross-race comparisons would only be useful for descriptive purposes, not for meaningful causal inference.) The end result would be an identified causal effect that not only gives us valuable insight into the other descriptive findings but also narrows down the causal mechanism that explains disparate, race-based educational outcomes.

We note several distinct advantages to this research design over more naive cross-race regression approaches. First, limiting the unit of analysis to a single racial group and reconceptualizing the treatment as being something that varies closely, but perhaps not exclusively, with race allows for experimental manipulation (in theory or practice), thus avoiding the critique that no well-defined potential outcomes exist. Second, because the alternate treatment may be “assigned” post-birth, it also allows for the inclusion of all pre-treatment variables (confounders), including traits like mother’s education, health, nutrition, and early educational opportunities. To some extent, we are advocating treating simple biological race (as this is what the race variable now becomes) as a confounding variable that must be controlled for or conditioned on.⁸ Third, with enough data, conditioning on race before moving to a causal analysis resolves the common support problem; it might be difficult to find white matches living in similar areas as black children, but focusing on within-race variation resolves this problem. Lastly, while it would be impossible to meaningfully assign all

---

⁸Here, the approach we are suggesting may in some instances be similar to an effects modification approach. Effects modification would be appropriate in instances where the treatment effect varies according to some different strata or subgroup (i.e., there are heterogenous treatment effects that vary systematically by subgroup). Because the impact of the alternate, non-race treatment may vary according to subgroup, comparing the results between groups may also be useful.
of the components of race as a treatment, disaggregating race allows for the investigation of an effect of a single “stick” or element of race – a much more tractable enterprise.

**Leveraging the Right Kinds of Elements of Race.** Which components of race would make for suitable treatments? This largely depends on both the research question and the researcher’s preliminary hypotheses; our advice is that more manipulable components of race are the most tractable, theoretically and practically. Consider Figure 2, which presents a hypothetical continuum of features that are strongly associated with race but that exhibit varying degrees of mutability. Facial features – such as the shape of one’s eyes or the contours of one’s nose – are fairly immutable, possibly changed through plastic surgery but certainly not something researchers could manipulate easily. (The boom in ethnic-oriented plastic surgery might present some interesting, if far-flung, experimental possibilities – Dolnick (2011); Survey (2004).) These sorts of traits are less useful for researchers, as they present the same conundrums as immutable characteristics do.

A better approach would be to pursue a treatment that is at once more mutable and more likely to be social constructs – e.g., name or neighborhood. First names, for example, are not only quite mutable but provide a strong signal about racial or ethnic background (Chang et al., 2010): one could imagine an experiment in which new parents of the same race and background were randomly assigned to pick a baby name from one of two lists. One list would include names that are not strongly identified with the relevant racial or ethnic group and the other list would include names that do exhibit such an association. This kind of study could then assess the short- and long-run effects of a racially or ethnically specific name on outcomes like education or employment. The key point is that not all of the “sticks” are inherently immutable; neither is the whole “bundle” automatically assigned at birth. For

---

9A large literature in gender studies distinguishes between “sex” and “gender” where “sex” is defined as biological and anatomical while “gender” is defined as the product of psychological, social and cultural forces – see, for example, (West and Zimmerman, 1987; Deaux, 1985). This is similar to what we demonstrate here.
applied researchers, perhaps the best strategy is to conduct a variety of within-group analyses manipulating various of these sticks to more precisely isolate a causal mechanism.

**Studies Manipulating an Element of Race.** A small number of experimental studies have begun using this kind of elements of race approach. For example, one element of race (i.e., one of the sticks in the bundle) is self-worth and self-assessment, which in turn are amenable to experimental manipulation. Walton and Cohen (2011), for example, randomly assigned college freshman a message that all college students struggle to fit in initially but can ultimately succeed. Compared to the black control students, the black treated students exhibited sustained academic improvements over their college careers and later reported being happier and healthier than the black controls. Treated whites, on the other hand, exhibited no significant difference from control whites.

In the observational context, some studies have successfully leveraged additional components of race in order to extract surprising inferences. Cutler, Fryer and Glaeser (2005), for example, explore why African Americans suffer from higher rates of hypertension compared to whites. By more closely examining black subpopulations, they demonstrate that blacks whose enslaved ancestors survived the “Middle Passage” across the Atlantic exhibit higher rates of salt sensitivity compared with blacks whose ancestors were not enslaved (i.e., more recent African immigrants to the United States or the United Kingdom). A possible mechanism is that salt retention – a precursor to hypertension – enabled enslaved African to survive the deadly three-month sea voyage that constituted the Middle Passage. Thus, the appropriate treatment in this study was not race per se; it was treatment by the Middle
Passage, a finding only made clear with within-group comparisons.

Another example is provided by Nisbett and Cohen (1996), which explores why white American men in the South exhibit higher rates of violence than white men in the North. Nisbett and Cohen identify and experimentally test cultural differences they hypothesize are borne of varying immigration patterns. Where a more conventional research design might have compared rates of violence between white and black men, Nisbett and Cohen attempt to disentangle the effects of race and norms by exploiting cultural variation between Northern and Southern white men. The standard cross-race approach takes the appropriate units of analysis to be the person or person(s) of color and his or her white (or non-minority) counterpart.\footnote{Of course, this is not the approach taken by all applied researchers. Some researchers, particularly in race and ethnic politics or in urban politics look at different measures – for example, the percent of a census tract that is minority. However, looking at minority versus non-minority populations does seem to be the general default approach.} Though cross-race comparisons are widely used in fields such as health and education, due to post-treatment bias and immutability, such comparisons are problematic when attempting to provide anything more than a descriptive analysis. In contrast to exposure studies that attempt to measure a contemporary effect of race as a signal, studies exploiting within-group variation are often attempting to identify some trait or quality assigned to members of a population at an earlier historical period.

6 Empirical Example: Explaining the Achievement Gap

We illustrate some of these concepts via an empirical example taken from Fryer Jr and Levitt (2004), which explores the determinants of educational outcomes for white versus black children. Most literature in this field has found that black children consistently and strongly underperform on educational tests, despite researchers controlling for a variety of socioeconomic factors that could potentially affect educational outcomes. This has left policy makers and scholars puzzled as to how and why the gap between white and black
children develops as well as how it could be lessened. In sharp contrast to the previous literature, however, Fryer Jr and Levitt found that the test score disparity between white and black kindergarteners almost entirely disappears after accounting for three factors: (1) participation by students’ families in Women, Infants, and Children welfare programs (WIC), (2) whether the mother was a teenager at time of the child’s birth (or more generally, mother’s age), and (3) an amalgam measure of socioeconomic factors. Thus, Fryer Jr and Levitt conclude that socioeconomic forces are at play, and that they plausibly exacerbate over time, thereby continuing to disadvantage black students.

Table 1 demonstrates our replication of Fryer Jr and Levitt’s core analysis, which used data from the Early Childhood Longitudinal Study (ECLS-K). The key outcome here, and in subsequent analyses, is the students’ performance on kindergarden-level testing. (In the interest of full disclosure, we were able to replicate their results to approximately a tenth of a decimal for each coefficient, although we had different sample sizes due to different approaches to handling missing data; substantively, the results are all identical, and multiple imputation of missing variables does not change the results.) Like Fryer and Levitt, we break up the analysis into test scores of math and test scores on reading. In all instances, whites comprise the baseline group to which (1) blacks, (2) Hispanics (an exclusive category), (3) Asian Americans, and (4) “others” are compared.

Thus, when a simple analysis of race on educational outcomes shows that black and Hispanic students fare worse than whites, while Asians perform better (Model 1 and Model 4). Focusing on black-white differences specifically, the most interesting results come in the way of the scores on the reading tests (Models 4-8). Here, black students begin with the traditional disadvantage against whites – controlling for nothing except student race, black students on average achieve scores that are 0.40 lower than that of white students (Model 4). However, including the three key covariates – (1) a proxy for the age of the mother, (2) a component measuring socioeconomic status, and (3) participation in WIC – not only
erases the effect, actually has the effect of reversing it, resulting in a positive, statistically
significant effect on the black status (Model 8), a surprising result.

Making a causal (or even quasi-causal) claim on the basis of only these results would be misguided, and the methodology discussed here counsels against this sort of kitchen-sink
econometric analysis. First, all of the key variables that erase the effect of the black variable

Table 1: Replication of Fryer and Levitt (2004).
Figure 3: Distribution of Mothers’ Ages (Left) and SES Composite Measure (Right), Disaggregated by Race

are realized post-treatment, meaning that they are fundamentally driven and affected by a students’ race. Second, and relatedly, all three of the key variables fluctuate substantially according to the students’ race. For example, among these students, 76% of black kindergardeners come from families enrolled in WIC, but only 29% of white kindergardeners do. Similarly, black students are more likely to have had younger mothers and are more likely to have low SES measures (Figure 3). Thus, when understanding whether there is a causal effect of race on the educational achievement gap, controlling for these factors that vary substantially (and are affected) by race introduce bias into an estimate of the total effect of race. (In other words, if we control for some components affected by race and not others, what does the remaining coefficient on the race variable actually mean?) A better interpretation of Model 8 would be to take these (as Fryer and Levitt do) as providing intuition for a causal mechanism – i.e., the pathway via which race plays a role in this outcome.

To this extent, a more appropriate analysis (and what Fryer Jr and Levitt do later) is to move forward with within-race comparisons. For example, consider WIC status. We could easily imagine an experiment where black children are born to families, and then these families are randomly sorted into WIC and not WIC-receiving families. We can similarly imagine other experiments for other components of the SES measure, or for the number
of books a family own (another of the variables that Fryer and Levitt control for). Unlike the results presented in Model 8 (or Model 4), these kinds of research designs allow for identification (at least in theory) of a plausible, manipulable treatment(s). Thus, knowing that WIC, teenage mothers, and SES status substantially attenuate the naive “effect of race” gives us a valuable starting point; these are possible alternate treatments that we should explore via more rigorous analysis.

We illustrate some of these notions by presenting within-race analyses, focusing on the three leading suspects of WIC, teen moms, and low SES. For example, suppose that our working hypotheses is that black students are more likely to come from families with low SES measures, and that these SES measures are determinative of their relatively poorer performance. We thus have a treatment (low SES status), a moment of treatment (which is realized post-birth), and an identified hypothetical experiment. We also have well defined potential outcomes – a black student with low SES versus the same black student with high SES. To determine the relative role of SES (i.e., the average treatment effect of low SES), we use matching. Although simple regression would be adequate to tease out preliminary relationships, matching is preferable because it isolates this effect regardless of the possible ways that other variables may be affecting one another. To implement the matching, we use coarsened exact matching (Iacus, King and Porro, 2012). We match on those characteristics that would be considered pre-treatment, including the mother’s age and the student’s weight at birth. We do not match on those characteristics that would be affected by low SES status, such as the number of books owned in the home.

The results are presented in Table 2, and show a clear and convincing evidence that low SES adversely affects educational outcomes. Combined with the fact that black students are more likely to come from low SES backgrounds, this provides clues into the causal mechanism behind black students’ performance. Similar results are found for having a teenage mother and for being enrolled in WIC. Thus, our conclusions dovetails with Fryer Jr and Levitt’s:
<table>
<thead>
<tr>
<th></th>
<th>Change in Outcome</th>
<th>95% CI</th>
<th>Matched n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blacks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Having Low SES</td>
<td>-0.49</td>
<td>(-0.58, -0.40)</td>
<td>903</td>
</tr>
<tr>
<td>Having Teen Mom</td>
<td>-0.24</td>
<td>(-0.34, -0.14)</td>
<td>1292</td>
</tr>
<tr>
<td>Enrolled in WIC</td>
<td>-0.36</td>
<td>(-0.45, -0.27)</td>
<td>1115</td>
</tr>
</tbody>
</table>

Table 2: Change in test scores, after matching, associated with (1) going from high to low SES scores, (2) having a teenage mother, and (3) coming from a family enrolled in WIC.

There is nothing inherent about race that drives the achievement gap. Rather, we can desegregate factors commonly associated with one racial group and attempt to make causal inferences about these sort of characteristics.

7 Empirical Example: Who Fights in African Militias?

Another example is taken from Humphreys and Weinstein (2008), which use survey data to explore which personal characteristics will cause an individual to join civil war militia groups. Previous scholarship has suggested that individuals from marginalized groups have a higher incidence of joining militias. Looking at Sierra Leone’s (1) opposition Revolutionary United Front (RUF) and (2) government-backed Civil Defense Forces (CDF), Humphreys and Weinstein consider, among other hypotheses, whether members of the politically excluded Mende ethnic group are more likely to join militias.

To test this hypothesis, Humphreys and Weinstein include a Mende dummy variable (1 if Mende, 0 otherwise) in a logit regression that has membership in a either militia group as the outcome variable and a host of additional variables as controls. Table 3 presents our replication of the Humphreys and Weinstein results, which is identical to theirs to the tenth decimal place and has the same number of observations. In analyzing more closely the role Mendes play, however, we should recognize that being Mende is considered in Sierra Leone an immutable characteristic, and it is therefore a “treatment” administered at birth.
Accordingly – and acknowledged by Humphreys and Weinstein – an individual’s Mende status is causally prior to a host of other variables included in the model; the inclusion of these other variables (e.g., whether respondent lives in a mud hut) therefore introduces post-treatment bias into the estimate of the total effect of being Mende. Results after removing post-treatment variables are presented in Table 4.

For the RUF membership (the opposition militia), the significance of the Mende variable does not change – self-identifying as Mende is related positively with membership in the RUF, although the size of the effect is reduced (a move that makes sense given the RUF’s status as a Temne-backed organization). The results are, however, different for the government-backed, Mende-supported CDF. The original model (Table 3) shows that the Mende status coefficient, although positive, is not statistically significant. Once we remove the variables directly affected by Mende status, we see that being Mende is positively linked with belonging to the CDF, and that the effect is significant when we control for age. Interestingly, the effect associated with Mende self-identification is about as strong for membership in the RUF and the CDF, a finding makes more substantive sense. The CDF, after all, is thought to represent Mende interests.

To explore the difference between Mendes and non-Mendes further, we focus specifically on membership in the government-backed CDF. Table 5 shows the results of separate logit regressions on Mende and non-Mende populations, with membership in the Mende-backed CDF as the outcome variable. While we don’t advocate this strategy in all instances, it does help to isolate the importance of different traits in the different groups – and to show the effect modification associated with the ethnicity variable. Indeed, what this regression shows is that different variables are important for the different groups (Table 5). For the Mende, living in mud housing (a proxy for poverty) is not a predictor of CDF membership, whereas for the non-Mende, it is. Likewise, having a friend in the CDF is not predictive, but it is for non-Mende. For the Mende group, it is being a boy or a man (as opposed to girl or woman)
<table>
<thead>
<tr>
<th></th>
<th>RUF Membership</th>
<th>CDF Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept                                                       −12.48*        −26.74*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.17)         (3.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mud Walls                                                       0.92*          1.61*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.41)         (0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Access to Education                                     1.09*          0.80*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.30)         (0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supports the SLPP                                               −0.49          −0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.67)         (0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mende Ethnic                                                    2.16*          0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.87)         (0.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does Not Support Any Party                                      1.29*          1.62*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.57)         (0.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offered Money to Join RUF                                        1.77*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felt Safer Inside RUF                                           −0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friend of RUF Members                                           0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Villages Accessible by Foot or Boat Only                        −0.01          0.03*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.02)         (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer                                                          0.32           1.39*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.56)         (0.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student                                                         0.83           1.26*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.54)         (0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male                                                            2.44*          4.06*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.64)         (0.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age                                                             1.03           3.52*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.20)         (1.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Squared                                                     −0.20          −0.46*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.16)         (0.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freetown                                                        −0.15          0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.72)         (0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infant Mortality                                               13.52          16.85*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6.73)         (6.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offered Money to Join CDF                                       3.19*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felt Safer Inside CDF                                           2.34*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friend of CDF Members                                          0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong>                                                           59             64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Logit Regression Replication of Humphreys and Weinstein (2008). Dependent variable is membership in the RUF or CDF militia groups.
that is predictive of membership in the CDF.

Specifically, both Mende and non-Mende individuals are receptive via factors associated with economic grievances—poverty (mud housing) and less education are more likely to lead to membership in the CDF. On the other hand, the Mende people are less susceptible to selective incentives—e.g., friendship, and, to a lesser extent, money and safety. These results suggest that many Mendes already have a natural affinity for the Mende-backed CDF—not necessary through friendships, but through shared ethnicity. By contrast, for non-Mendes, for whom no pre-existing affinity exists for the CDF, monetary and social incentives are more salient. This inclination is borne out by a simple interacted model, represented in Table 6, although we note that the interaction term is not significant.\footnote{Mediation analyses using the \texttt{mediation} package were less conclusive.}

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
 & RUF & RUF & CDF & CDF \\
(Intercept) & \(-6.27^*\) & \(-3.49\) & \(-8.31^*\) & \(-7.22^*\) \\
 & (0.75) & (1.87) & (0.84) & (1.65) \\
Mende & 1.47* & 2.08* & 1.31 & 1.63* \\
 & (0.70) & (0.73) & (0.69) & (0.71) \\
Gender & 0.87* & 1.53* & 3.63* & 3.92* \\
 & (0.30) & (0.41) & (0.45) & (0.57) \\
Age & \(-1.46\) & \(-0.56\) & & \\
 & (0.87) & (0.85) & & \\
Age Squared & 0.09 & 0.03 & & \\
 & (0.10) & (0.09) & & \\
\hline
\textit{N} & 59 & 59 & 64 & 64 \\
\end{tabular}
\caption{Logit Regression Replication of Humphreys and Weinstein (2008). Dependent variable is membership in the RUF or CDF militia groups. Post-treatment variables have been removed from the model. Standards errors in parentheses \* indicates significance at \(p < 0.05\)}
\end{table}
<table>
<thead>
<tr>
<th></th>
<th>Mende Only</th>
<th>Non Mende Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-21.71^*$</td>
<td>$-25.25^*$</td>
</tr>
<tr>
<td>(3.73)</td>
<td>(9.12)</td>
<td></td>
</tr>
<tr>
<td>Mud Walls</td>
<td>2.06*</td>
<td>2.18*</td>
</tr>
<tr>
<td>(0.61)</td>
<td>(0.83)</td>
<td></td>
</tr>
<tr>
<td>Lack of Access to Education</td>
<td>0.81</td>
<td>0.86*</td>
</tr>
<tr>
<td>(0.52)</td>
<td>(0.29)</td>
<td></td>
</tr>
<tr>
<td>Supports the SLPP</td>
<td>0.00</td>
<td>$-1.78^*$</td>
</tr>
<tr>
<td>(0.76)</td>
<td>(1.05)</td>
<td></td>
</tr>
<tr>
<td>Does Not Support Any Party</td>
<td>1.75*</td>
<td>2.03*</td>
</tr>
<tr>
<td>(0.84)</td>
<td>(0.58)</td>
<td></td>
</tr>
<tr>
<td>Offered Money to Join CDF</td>
<td>3.33*</td>
<td>3.66*</td>
</tr>
<tr>
<td>(1.09)</td>
<td>(1.17)</td>
<td></td>
</tr>
<tr>
<td>Felt Safer Inside CDF</td>
<td>2.20*</td>
<td>2.72*</td>
</tr>
<tr>
<td>(0.43)</td>
<td>(0.37)</td>
<td></td>
</tr>
<tr>
<td>Friend of CDF Members</td>
<td>$-0.82^*$</td>
<td>2.09*</td>
</tr>
<tr>
<td>(0.74)</td>
<td>(0.65)</td>
<td></td>
</tr>
<tr>
<td>Villages Accessible by Foot or Boat Only</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>2.10*</td>
<td>1.03*</td>
</tr>
<tr>
<td>(0.76)</td>
<td>(0.50)</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1.03</td>
<td>0.38</td>
</tr>
<tr>
<td>(0.86)</td>
<td>(0.69)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6.15*</td>
<td>2.01*</td>
</tr>
<tr>
<td>(1.03)</td>
<td>(0.61)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.41</td>
<td>1.46</td>
</tr>
<tr>
<td>(1.28)</td>
<td>(1.26)</td>
<td></td>
</tr>
<tr>
<td>Age Squared</td>
<td>$-0.16^*$</td>
<td>$-0.15^*$</td>
</tr>
<tr>
<td>(0.15)</td>
<td>(0.13)</td>
<td></td>
</tr>
<tr>
<td>Freetown</td>
<td>2.55</td>
<td>1.47</td>
</tr>
<tr>
<td>(1.35)</td>
<td>(1.48)</td>
<td></td>
</tr>
<tr>
<td>Infant Mortality</td>
<td>19.22*</td>
<td>26.47</td>
</tr>
<tr>
<td>(6.22)</td>
<td>(41.49)</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>41</td>
<td>44</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 5: Comparing models fitted on the Mende population versus on the non-Mende population. Coefficients are logit estimates (standard errors in parentheses). Outcome variable is whether an individual joined the CDF or not.
(Intercept)  $-8.523^*  -8.140^*$  
          (0.850)   (1.601)  
Mende  1.411  1.722$^*$  
          (0.740)   (0.780)  
Male  3.474$^*$  3.722$^*$  
          (0.435)   (0.527)  
Friend of CDF Members  2.122$^*$  2.383$^*$  
          (0.652)   (0.652)  
Mende:Friend of CDF Members  $-0.708  -1.133$  
          (0.992)   (1.025)  
Age  $-0.134  
          (0.799)  
Age Squared  $-0.024  
          (0.089)  
AIC  22.771  26.713  
BIC  65.949  87.162  
log $L$  8.614  14.644  

Standard errors in parentheses  
$^*$ indicates significance at $p < 0.05$

Table 6: Simple interacted model. Logit coefficients. Outcome variable is membership in the CDF.

8 A Unified Framework for Race and Causality

In this paper, we have highlighted both the pitfalls and the possibilities associated with trying to extract meaningful causal inferences about race in a quantitative framework. Most quantitative social scientists try to gain leverage on the causal impact of race by including simple dummy variables, along a standard battery of control covariates. As we note in this article, however, race presents unique challenges for quantitative scholars. First, race is resistant to manipulation and, hence, potential outcomes are ill-defined and research questions fundamentally unidentified. Second, because race is “assigned” at birth, the host of characteristics that most social scientists control for (age, education, income, etc.) occur after the treatment is assigned and therefore potentially introduce bias into the estimate of interest.
Third, an equally meaningful problem is that race is too complex to be synthesized into one neat variable. To the contrary, how a person is categorized by society or self-identifies is inextricably intertwined with tangible measures such as education, income, health, diet, economic status as well as intangible factors as culture, traditions, and political and social attitudes. Thus, the introduction of a race “dummy” variable – along with attendant background covariates – oftentimes does a disservice to queries that look to make causal inferences about race-based characteristics.

The framework described in this article may help researchers interested in race or ethnicity extract those kinds of inferences that capture a causal effect. First, we suggest that researchers interested in race begin by thinking whether their research design may be appropriately captured by an exposure study. This kind of research design may be particularly appropriate for those studying public opinion, political behavior, law, and public policy – fields in which questions of interest frequently involve how institutions or individuals view and interact with racial signals and cues. Because the exposure research design avoids the pitfalls outlined above, it is serves as an extraordinarily useful (yet underused) research design.
Second, when it comes to research designs focusing on minority populations themselves, researchers may actually be able to focus on some alternate manipulable treatment regime that varies closely (perhaps exclusively) with race. Here, we find the analogy to the “bundle of sticks” a useful one; and even though biological race itself may not be subject to manipulation, things like name, culture, neighborhood, dialect, and diet – i.e., those variables that define the contours of racial identification – may be experimentally manipulated and observationally assessed. We do not attempt to say that such an alternate treatment may be found in all instances; rather, the takeaway is that (a) such an alternate treatment may vary closely with race, (b) may not already be included in the analysis, and (c) may explain away much of the effect previously attributed to race. Focusing on treatments other than the biological race of a subject not only solves problems with fundamentally unidentified research questions and ill-defined potential outcomes, but it also forces researchers consider exactly what is being captured by the racial identification variable. Both of these are welcome considerations – both in terms of increasing statistical rigor and also in terms of increasing substantive engagement with developments in the racial and ethnic politics literature.

Two further issues are worth flagging. First, we suggest throughout that researchers think carefully about post-treatment bias issues. This is not a new warning (e.g., King, Keohane and Verba (1994)), but it carries particular urgency when it comes to race and causality. Race, which is assigned in part at birth, has immutable components, which means that the host of variables that social scientists routinely control for may be determined post-treatment and could therefore introduce bias into the causal estimate. To rectify this issue, researchers interested in the causal impact of race should think carefully and what is and what is not post-treatment. Our suggestions for exposure studies and within-race analyses substantially help in this regard. Second, we note that many possible alternate treatment regimes vary almost exclusively by race and, therefore, comparisons between whites and blacks, Hispanics and Asians, etc., may be of limited use due to problems with collinearity and a substantial
(and persistent) lack of common support among key covariates. As a result, a useful way to explore whether alternate treatment regimes could be capturing some of the effects of “race” is to conduct within-group comparisons.

We conclude by emphasizing the importance of experimental analogies. This is a point that has been made by the causal inference and econometrics literature (Angrist and Pischke, 2009), but is particularly worthwhile for those specifically interested in race. Keeping an eye on what the ideal experiment would look like (and what factors would or would not have to be controlled for) is essential for thinking clearly about potential identification strategies and problems. Ultimately, keeping experimental analogies in mind – while also keeping in mind what precisely is measured with the inclusion of a “race” variable – will help scholars to reconcile race and causation.

References


Boker, SM, JF Cohn, BJ Theobald and I Matthews. N.d. “Something in the way we move: Motion dynamics, not perceived sex, influence head movements in conversation.” Experimental Psychology: . . . Forthcoming. 14

Chang, Jonathan, Itamar Rosenn, Lars Backstrom and Cameron Marlow. 2010. ePluribus: Ethnicity on Social Networks. 19
Griffin, J.H. 1996. Black Like Me. NAL. 7
Hochschild, JL and V. Weaver. 2010. Perspectives on Politics 8(3). 7


URL: http://www.pnas.org/content/98/26/15387.short 14

Locke, J. 1847. *An essay concerning human understanding*. Troutman & Hayes. 7


Nisbett, R.E. and D. Cohen. 1996. *Culture of honor: The psychology of violence in the South*. Westview Press. 21, 32

URL: http://www.princeton.edu/~pager/pager_ajs.pdf
URL: http://www.synovate.com/consumer-insights/infact/issues/200406/
URL: http://www.nytimes.com/2010/04/12/sports/12sickle.html