Measuring Social Change as Categorical Change

by

Jeffrey A. Smith

Department of Sociology
Duke University

Date: ______________________

Approved:

___________________________

J. Miller McPherson, Co-Supervisor

___________________________

James Moody, Co-Supervisor

___________________________

Peter J. Mucha

___________________________

Lynn Smith-Lovin

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Sociology in the Graduate School of Duke University

2013
ABSTRACT

Measuring Social Change as Categorical Change

by

Jeffrey A. Smith

Department of Sociology
Duke University

Approved:

___________________________
J. Miller McPherson, Co-Supervisor

___________________________
James Moody, Co-Supervisor

___________________________
Peter J. Mucha

___________________________
Lynn Smith-Lovin

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Sociology in the Graduate School of Duke University

2013
Abstract

Sociologists often depict demographic categories as socially constructed, non-essential, and fluid. The language of fluid, contingent categories has not, however, translated very well into the practice of describing social change in a population. There are notable exceptions, but the typical approach is still to take fixed demographic categories, such as Black/White, and follow their outcomes over time. The first goal of this dissertation is to bridge the gap between theory and practice by offering a formal framework for measuring categorical fluidity. The second goal is to use changes in categorical meaning to explore the macro features of a social system.

First, I develop a model of categorical change rooted in interaction patterns, such as marriage or friendship rates. Interaction patterns offer an ideal metric to measure fluid demographic categories: they concretely capture social distinctions without relying solely on pre-defined labels. I consider two categories to be equivalent if the observed behavioral implications of group membership are the same, even if the labels are different. If College graduates now interact in the same manner as High School graduates used to, then College is the new HS. To formalize this idea, I place categories into social locations based on observed rates of interaction. Categories are close if interaction is likely and far if interaction is unlikely. I then ask which categories occupy the same locations over time, or have the same range of interaction partners.
Second, I apply this model to the question of racial change in America. I interpret two macro level changes, one demographic and one political, through the lens of categorical change. Demographically, there have been large increases in the number of Hispanics identifying as Other racially, as opposed to White or Black. Using Census marriage data, I find that this increase in Other-Hispanics reflects a schism in the meaning of Hispanic. The shifts in social locations point to a growing divide between those that see Hispanic as another race and those that do not. Politically, there have been large changes in the measurement of race, with individuals now allowed to claim multiple races in the census. I ask how these “new” mixed race categories fit into the existing racial order. I find that the Hispanic mixed race categories create distinct categories in a way that the mixing of traditional racial/ethnic categories does not.

Third, I use the model of categorical change to test theories of power and influence. I argue that the meaning of a category amongst one part of the population may be shaped by the experience, or changing conditions, of another part of the population. This asymmetry serves as the measure of aggregate level influence. Substantively, I apply this approach to racial stratification in the US, where I use joint changes in educational meaning and attainment to characterize systems of racial stratification from 1940-2000. Using Census data on race and education, I find that the US is characterized by a system of hegemony, where changes in attainment amongst the majority drive the meaning of education for other racial groups.
Contents

Abstract .............................................................................................................................................. iv

List of Tables ........................................................................................................................................ x

List of Figures ....................................................................................................................................... xi

Acknowledgements .................................................................................................................................. xii

1. Introduction ......................................................................................................................................... 1

  1.1 Summary of Chapter 2 ..................................................................................................................... 2

  1.2 Summary of Chapter 3 and Chapter 4 ............................................................................................. 6

  1.3 Summary of Chapter 5 ..................................................................................................................... 9

2. Developing a Model of Categorical Change ....................................................................................... 14

  2.1 Measuring Social Change if Categories are Social Outcomes ....................................................... 17

    2.1.1 Interaction Patterns as Metric ................................................................................................. 19

    2.1.2 Social Space and Social Distance ............................................................................................ 22

    2.1.3 Combining Models ................................................................................................................... 26

  2.2 Methods .......................................................................................................................................... 30

    2.2.1 Background on Distance and Equivalence Models ................................................................. 30

    2.2.2 Model Overview ....................................................................................................................... 34

    2.2.3 Model Details ........................................................................................................................... 36

      2.2.3.1 Calculating Initial Locations .............................................................................................. 37

      2.2.3.2 Equating Categories within Time ....................................................................................... 37

      2.2.3.3 Equating Categories across Time ....................................................................................... 40
2.3 Empirical Example: Education 1940-2000 ................................................................. 45
  2.3.1 Data, Measurement and Models ........................................................................ 46
  2.3.2 Results .............................................................................................................. 48
    2.3.2.1 The Relational Structure of Education: 1940-2000 .................................. 48
    2.3.2.2 The Relational Structure of Education: Theoretical Implications .......... 51
  2.4 Conclusion ........................................................................................................... 52

3. Is Hispanic a Racial Category? ................................................................................. 57
  3.1 Race/Ethnicity as a Social Construction ............................................................... 61
  3.2 Interaction Patterns, Social Locations, and Racial Categories ............................. 65
  3.3 Hispanic as a Racial Category? ............................................................................ 70
  3.4 Data ...................................................................................................................... 75
  3.5 Methods ............................................................................................................... 77
    3.5.1 Review of Chapter 2 ..................................................................................... 77
    3.5.2 Predicting Changes in Locations over Time .................................................. 81
  3.6 Results .................................................................................................................. 83
    3.6.1 Racial Organization ...................................................................................... 83
    3.6.2 Hispanic Change ......................................................................................... 88
  3.7 Conclusion ............................................................................................................ 93

4. The Relational Role of Mixed Race Categories ......................................................... 96
  4.1 Measuring Categories through Interaction Patterns ............................................. 100
  4.2 The Relational Meaning of Mixed Race Categories ............................................. 102
  4.3 Data ...................................................................................................................... 108
4.4 Methods ..................................................................................................................... 109
4.5 Results ...................................................................................................................... 113
4.6 Conclusion .............................................................................................................. 118

5. Using Categorical Fluidity to Test Theories of Stratification................................. 122

5.1 Measuring Stratification Systems............................................................................ 125
5.2 Background on Measuring Social Change as Categorical Change....................... 129
5.3 Using Educational Change to Characterize Racial Systems of Stratification....... 132
  5.3.1 Changes in Categorical Meaning and Underlying Conditions......................... 132
  5.3.2 Education and Race as Nested Social Structure............................................... 135
5.4 Explaining Educational Change by Race ............................................................... 136
  5.4.1 Homogeneity...................................................................................................... 137
  5.4.2 Separate Worlds............................................................................................... 138
  5.4.3 Hegemony ....................................................................................................... 140
  5.4.4 Minority Threat............................................................................................... 142
5.5 Data......................................................................................................................... 144
5.6 Methods .................................................................................................................. 146
  5.6.1 Measuring Educational Attainment and Educational Meaning..................... 146
  5.6.2 Predicting Educational Categories across Time............................................ 149
5.7 Results..................................................................................................................... 153
  5.7.1 Cross Sectional Results................................................................................... 153
  5.7.2 Over Time Results........................................................................................... 157
5.8 Conclusion.............................................................................................................. 161
6. Conclusion ................................................................. 165

6.1 Measuring the Fluidity of Demographic Categories....................... 165

6.1.1 Remaining Methodological Issues ...................................... 167

6.2 Demographic Categories, Social Locations and Macro Level Change .......... 169

6.3 Substantive Conclusions about Education and Race/Ethnicity in the US ........ 171

6.4 Future Work .................................................................... 173

Appendix A ............................................................................ 177

Appendix B ............................................................................... 178

References ............................................................................. 179

Biography ............................................................................... 194
List of Tables

Table 1: Model Fit for Latent Space Solutions by Dimension.................................86
Table 2: Models Predicting Hispanic Change along Hispanic Dimension..................90
Table 3: Summary Statistics for Race and Education over Time.............................155
Table 4: Logistic Regression Results for Educational Change by Racial Group..........157
List of Figures

Figure 1: Example Social Space: Focal Location Age=45, Education=12 ......................... 23
Figure 2: Example of Estimating Social Locations in Social Space over Time ..................... 27
Figure 3: Examples of Locational Maps: Education in One Dimension .............................. 38
Figure 4: Schematic Description of Fluid Category Model ............................................. 44
Figure 5: Map of Social Locations over Time: Educational Categories, 1940-2000 ............... 49
Figure 6: Example Social Locations: Racial/Ethnic Categories at Two Time Points ............ 69
Figure 7: Locations in Racial Space along 5 Dimensions: 1980 ...................................... 87
Figure 8: Racial Locations along Hispanic Dimension: 1980-1990 .......................... 89
Figure 9: Distance between Categories in 1980 and 1990 ........................................... 92
Figure 10: Idealized Types of Mixed Race Categories .................................................... 104
Figure 11: Distances between Mixed Race Categories and Parent Categories ............... 115
Figure 12: Example Social Locations: Educational Categories at Two Time Points ......... 131
Figure 13: Social Locations over Time: Educational Categories, 1940-2000 .................. 135
Figure 14: Map of Social Locations over Time: Educational Categories by Race .......... 154
Figure 15: Cumulative Distribution of Education by Race ............................................ 156
Figure 16: Logistic Regression Results for Educational Change by Racial Group .......... 159
Figure 17: Scatter Plot of Odds Ratios by Relative Distance 1940-2000 ......................... 177
Acknowledgements

This project has evolved considerably over time, and I have benefited from the input and advice of many people along the way. I would like to first thank Miller McPherson for setting me on this intellectual path. Those meetings in my first few years in graduate school introduced me to many of the ideas employed in this dissertation and challenged me to think much harder than I was used to. He involved me in a number of projects that would later inform my thinking here, including my network sampling paper, a paper on social distance and a paper on autocorrelation in social space. I also want to thank him for instilling in me an expectation that a paper, let alone a dissertation, should be ambitious. I hope this project lives up to those expectations.

Jim Moody was instrumental in turning a set of raw ideas into a finished project. He pushed me to reconsider the theoretical/substantive contributions of the project when I was willing to settle for a methods heavy dissertation. His comments on my job talk (based on this dissertation) improved the quality of the entire project. Both the ideas and presentation were greatly enhanced through his patient, but critical, comments. It certainly went beyond the call of duty to hear the same job talk 4 or 5 times. I also want to thank him for reading working drafts of this dissertation as well as other papers. His comments on submitted papers helped me learn the review process, and I appreciate his
investment in me as a scholar. Finally, I want to thank him for guiding me through the murky task of finding a job.

I would also like to acknowledge Lynn Smith-Lovin for her helpful comments on earlier presentations of this material and for drafts of specific chapters. The dissertation is certainly clearer due to her influence. Peter Mucha offered helpful suggestions on the original proposal, and his DNAC presentations have taught me a great deal about the difficulty of clustering networks through time. I want to thank all the committee members for their patience with this project. I am grateful that I was afforded enough intellectual leeway to move from the original ideas to the more developed theories and methods presented here. All things considered, there is surprisingly little political sociology in this dissertation.

Outside of the committee, I would like thank Noah Mark and Joel Levine for their helpful discussion of both methodological and theoretical issues. They helped me clarify what I was trying to say early on in the project. I would also like to thank James Montgomery for introducing me to formal sociology while an undergraduate at Wisconsin. I have benefited from the comments, suggestions and camaraderie of my fellow DNAC colleagues, including Jacob Fischer, (S) Joshua Mendelsohn, David Diehl and Heather Rackin. I thank them for their helpful suggestions and friendship throughout the years. I am also grateful for the fellowship support of the PARISS program at SSRI as well the James B. Duke fellowship earlier in my graduate career.
My family has offered unconditional support throughout this long process I want to thank my parents, Steven and Janis Smith, and my sister, Jenifer McMullen, for always listening to my problems and for having faith that things would work out in the end. I would also like to recognize my grandparents, Joseph and Lorraine Smith, Aaron Schwitzer, who recently passed away, and Iris Schwitzer, for their love and support.

Most importantly, I want to thank my wife, Robin. When I was struggling to come to terms with this project, she always seemed to know what I wanted to say before I was ready to say it. She saw the potential in the project when it was still a raw, not very well articulated idea. I am grateful for her faith in me. Many of the ideas of this project originated in discussion with her, and her conceptual contributions are felt throughout. I can’t overstate how important she is to this dissertation and how important she is to my life.
1. Introduction

Sociology is predicated on the idea that “things are not what they seem” (Berger 1969:23). We cannot rely on commonly held beliefs to interpret social phenomena as commonly held beliefs are likely to be misleading. Explanations should go beyond lay understanding and the biases associated with a particular social/historical moment. Sociologists face another dilemma when studying large, complex social systems. The social organization of large, heterogeneous populations is a core sociological outcome, but such settings are difficult to analyze precisely because they are large and heterogeneous (Blau 1989). In this dissertation, I argue that these goals are intertwined: if we can uncover the meaning of measurable phenomena, we can use that insight to study macro level processes, such as economic inequality or demographic change. The manner in which meaning and appearance are disconnected offer crucial clues to the organizing features of a population (Baily, Loveman and Muniz 2013).

I begin with the core sociological idea that demographic categories are fluid, social constructions (Barth 1969; Wimmer 2008; Zuberi and Bonilla-Silva 2008). An ethnic label (for example) is not an essential characteristic of an individual. The conditions mapped on to an ethnicity may increase or decrease the social distance to other ethnic groups, but ethnicity itself is a not a causal force (Bates and Peacock 1989). As a social construction, a demographic category could change meaning, or reflect different social distinctions, as the conditions surrounding the category change. For
example, the Irish became “White” over the course of the 20th century (Ignatiev 1995; Jacobson 1998). Being Irish used to matter for who you married and who your friends were—essentially who you were. But this is no longer the case as they mix freely in the population, facing the same social limitations as other White ethnic groups (Lieberson and Waters 1988). There is then a difference between the measure of a category (Irish) and the meaning attached to it (non-White versus White) (Loveman 1999). In this sense, demographic categories are not always what they seem. I use this idea to interpret macro level changes in demographic composition, economic inequality and political institutions.

The dissertation is first concerned with how to formally measure demographic categories through time. Chapter 2 describes the details of the approach. Chapters 3-5 apply the framework to substantive problems. I leverage changes in categorical meaning to explore the macro features of a social system. Chapters 3 and 4 interpret the changing racial composition of the US while Chapter 5 tests theories of macro level power and inequality.

1.1 Summary of Chapter 2

Chapter 2 develops a formal framework for measuring categorical fluidity. Sociologists often argue that racial, educational, or age categories are socially constructed, non-essential, and fluid (Bates and Peacock 1989; Burt 1991; Omi and Winant 1994; Anthias 2001). In practice, we tend to do ignore this argument when
measuring social change (Harris and Sim 2002). There are notable exceptions (Brubaker 2004; Wimmer 2008), but the typical approach is still to take fixed demographic categories, e.g. Black/White, and follow their outcomes over time. Here, I develop a framework to bridge the gap between theory and practice, or to make it easier for us to do what we believe.

The overall idea is to measure categories as or through interaction patterns, such as marriage or friendship rates. Practically, we need to define the demographic categories over time and cannot use the labels themselves as the categories are fluid and non-essential. Interaction patterns offer an ideal metric as they concretely capture social distinctions amongst demographic groups (Haller 1981; Bottero and Prandy 2003). Who you spend your time with is clearly important to people. Who you do not spend your time with is then substantively telling. Theoretically, the claim is that a category is not meaningful just because we measured it (White 1966). Muslim and Catholic should be treated as the same if Muslims marry and makes friends like Catholics (see Burt 1991 for a similar argument with age categories or Breiger’s [1981] work on occupations and mobility).

Using interaction patterns as a metric, different categories may come to reflect the same pattern of distinctions over time (in terms of who they are likely to interact with). These changes form the basis of measuring categorical change. Two categories are considered equivalent if the observed behavioral implications of group membership are
the same. The categories are thus equated through time based on observed distinctions, rather than the labels themselves. If College graduates now move in society like High School graduates used to, then College is the new HS.

The framework draws on social space and social distance traditions to formalize this idea (Laumann 1969; Goodman 1981; McPherson 2004). I first place categories into social locations based on observed rates of interaction (e.g. marriage rates); locations reflect the range of possible interactions for an individual with a given demographic profile. Categories are placed close together if they have high frequency of interaction and far apart otherwise (Laumann 1969; Hoff et al. 2002). Similarly, two categories with different patterns of interaction with other categories will be placed far apart (as they have different distances to the same third party category). The question is which categories occupy the same position over time, or which categories have the same range of possible social ties.

A social space approach views the categories as locations in a multidimensional demographic space (Blau 1977). The model assumes that social relationships become less likely as socio-demographic distance increases (McPherson 1983). A social space framework provides the organizing principle of a demographic space (homophily). The distance models, in contrast, place the categories into social locations based on observed rates of interaction (Levine 1990; Rytina 1992; Bottero and Prandy 2003). I thus estimate unknown social locations along known socio-demographic axes (i.e. the demographic
characteristics of interest, such as age or education). Demographically close categories should be close together, but the observed rates of interaction determine the exact distance. A social distance, social space model is tractable because the demographic space is held fixed while the category-location mappings are allowed to change. This makes it possible to trace categories over time as they change social location, but does not assume, a priori, that demographic distance is directly proportional to social distance.

Using this combined approach, there are two basic parts to the model. First, I determine which categories can be collapsed into a single social location. Categories are collapsed if they are relationally equivalent in a given time period. Categories are relationally equivalent if they have the same pattern of social ties and no barriers to interaction (Breiger 1981; Snip 1987). And second, I determine which category at Time 1 occupies the same location as which category at Time 2, and are thus relationally equivalent across time.

As an empirical example, I examine educational categories using Census marriage data from 1940-2000. There are two key findings. First, there are clear signs of inflationary pressure, so that High School in 1940 occupies the same social location as College in 2000. And second, the lower grades are collapsed into one large social category. For example, in 1940 Grade 0 and Grade 1 had very different social outcomes, but by 1980 Grades 0 through 6 are relationally indistinguishable. Whether you have 0
years of education or 6 years of education, you are doing the same things in terms of who you are allowed to interact with. The results suggest, substantively, that the social distinctions attached to an educational category decrease as the proportion earning that degree or more increases. For example, those with a post-college degree used to marry those with HS attainment. In 1940 comparatively few people earned a HS degree. By 1980, however, a HS degree was quite common. What serves as “high” education in 1940 is average education in 1980. Those with the highest education are thus unlikely to marry those with just a HS degree in 1980, while in the past such ties were possible.

1.2 Summary of Chapter 3 and Chapter 4

Chapters 3 and 4 apply the framework of Chapter 2 to the problem of racial change in America. I focus on two kinds of changes, one demographic and one political. Demographically, there have been large increases in the number of people identifying as Hispanic ethnically and Other racially (i.e. not White, Black, Asian, or Native American) (Logan 2003; Frank et al. 2010). Typical surveys ask individuals two racial/ethnic questions, one capturing racial identification and one capturing Hispanic identification. The number of individuals claiming an Other-Hispanic identity represents an increasing proportion of the population for two reasons: first, the number of Hispanics is increasing; and second, within the Hispanic population, individuals are more likely to identify as Other (Saenz 2004; Brown, Hitlin and Elder 2007). In 1980, around 95% of Hispanics identified as White. That numbers drops sharply to 50% in only 10 years (by
the 1990 Census). Chapter 3 asks what that large change meant for the meaning of Hispanic in the population.

I use Census marriage data from 1980-1990 to differentiate between three hypotheses of Hispanic change. First, it is possible that the demographic changes were purely demographic. More people identify as Other-Hispanic but the relational implications of the category are constant across time. Alternatively, the demographic shifts may point to the formation of a unified Hispanic category (Gomez 2007; O’Brien 2008). The racial question may be increasingly irrelevant for Hispanics as many now claim Other racially (Campbell and Rogalin 2006; Hitlin et al. 2007). This would suggest Hispanic itself is emerging as a racial category. Finally, we may be witnessing a schism in Hispanic, where some see Hispanic as a race (and thus identify as Other) while the rest of the Hispanic population does not (and still identifies as White) (Frank et al. 2010).

Formally, if there is no change in the meaning of Hispanic, then the Hispanic categories should occupy the same social locations across time. If there is an overarching Hispanic category, then all of the Hispanic categories should collapse into one social location. If there is a schism in the meaning of Hispanic, then Other-Hispanic should move away, socially, from the other non-Black Hispanic categories (White-Hispanic, Native American-Hispanic, etc.). Empirically, I find that the demographic changes reflected a schism in the meaning of Hispanic. This may indicate a long term trend of Other-Hispanic becoming the umbrella Hispanic category, while most Hispanic
categories are incorporated into White. This would both mimic and diverge from past immigrant experiences (Frank et al. 2010): mimic as Hispanics eventually become White; diverge as there is still a distinct Hispanic racial/ethnic identity.

Chapter 4 examines the relational meaning of mixed race categories. Mixed race categories were officially recognized by the Census in 2000 (Hirschman, Alba and Farley 2000; Farley 2002). Individuals were allowed to identify as Black and White, rather than Black or White. This marked an important departure in what the government viewed as a legitimate racial/ethnic category (Lee and Bean 2007). The question is how this shift in institutional legitimation maps on to the observed racial relational system. And more generally, how do the mixed race categories fit into the existing racial order?

I develop a formal classification scheme to analyze the mixed race categories. The underlying logic is to define the mixed race categories by the positions of its parent categories. Thus we want to define White-Black by the social locations of White and Black. The mixed race categories are not followed over time, but are rather defined relative to more traditional racial/ethnic categories.

Empirically, I use the 2000 Census to map each mixed race category into a structural type. I interpret the mixed race categories as more or less well bounded by the existing racial order. For example, a mixed race category that is distant from both parent categories represents a distinct racial/ethnic category on its own, while a category close to one parent category fits easily into the existing system. If White-Black is socially close
to White, it might as well be White. I find that non-Hispanic mixed race categories fall socially between the parent categories. White-Black, for example, is socially closer to both White and Black than White is to Black. Or, the barriers to interaction are relatively small between White-Black and both parent categories. The non-Hispanic mixed race categories thus fit predictably into the racial system. In contrast, Hispanic mixed race categories are more likely to be distinct racial/ethnic categories. They are more likely to be socially distant from one or both parent categories. For example, White-Black-Hispanic is defined by its stark distance, or lack of social ties, to White. Hispanic thus plays an important role in the creation of emerging racial/ethnic identities. The creation of new racial/ethnic categories is thus rooted in the growing Hispanic population, rather than the intermixing of more established racial groups.

1.3 Summary of Chapter 5

Chapter 5 uses the ideas and models of Chapters 2 and 3 to explain power and inequality in society. I first describe how fluid demographic categories can be used to characterize power/influence in a stratification system. I then develop and test theories of racial stratification using the approach.

The claim is that we can use conditional changes in categorical meaning, here education by race, to determine how much society reflects different parts of the population. I draw on the core ideas used throughout the dissertation: first, that categories represent the aggregate result of the stratification process; and second, that
categories can change meaning over time as the differentiating forces in society change. This chapter adds one important element to these ideas. The key here is that the underlying forces among one set of individuals can drive changes in the relational meaning of categories in another. This asymmetry is a sign of aggregate level influence—where the experience, or changing conditions, among one group is what matters for shaping meaning in a population.

The specific application uses changes in educational meaning and attainment to characterize systems of racial stratification. As we saw in Chapter 2, educational categories move to lower social locations as they become more prevalent in the population. The social distinctions between degrees thus decrease as they become easier to attain. There is then a systematic relationship between changes in educational attainment and changes in social location. I exploit that systematic relationship to study racial stratification. I ask how the meaning of education amongst one racial group is affected by changes in attainment happening across the population. I highlight three hypotheses about racial stratification. Each describes a different way race could shape the relationship between educational attainment and educational locations.

First, we could see a system where educational changes are localized in social space. The relational meaning of education is best predicted by attainment changes happening within that racial group. This means that the reference level of education is local. Individuals are only aware and concerned about those who are socially close. Most
social ties occur within, rather than across, racial groups, making it unlikely that an individual will know what is happening elsewhere in the population (Mark 2003). Even if people did know, they may be unconcerned as they are competing for different stakes in different labor and marriage markets (Bonacich 1972; Kalmijn 1993; Blackwell and Lichter 2004).

Second, the racial system could exhibit hegemony, where the (relational) meaning of education amongst racial minorities is driven by educational attainment changes amongst the majority (Gramsci 1971; Bourdieu 1984; Bourdieu 1990). In a hegemonic system, the racial majority shapes the relationships forming in other parts of the population, even if the level of education is quite different across racial groups. The majority sets the normative level of education, and all relationships are subject to those constraints—as educational categories are considered rare/common based on majority attainment. Similarly, the racial majority sets the constraints (in terms of educational attainment) for how individuals are sorted into organizations and occupations.

Finally, there could be a system of minority threat, where changes in education amongst the racial minority drive changes in educational meaning for the majority (Blalock 1967; Olzak 1992). The majority reacts strongly to changes in minority attainment, which pose a threat to majority control over the resources in society (Lieberson 1980; Taylor 1998; Baumer, Messner, and Rosenfeld. 2003). For example, all minority members may now get a HS degree. This would devalue HS amongst the
majority, even if their rate of HS graduation is constant. Thus, the action/reaction of economic threats are reflected in the changing meaning of the educational categories.

I use Census data from 1940-2000 to differentiate between these three hypotheses. I find that the US is characterized by a system of hegemony. White shifts in attainment drive the meaning of education across the population. The level of attainment in the majority serves as the normative level of “good” education. Those in minority racial groups form relationships based on the distinctions implied by the majority level of attainment. For example, those with high education in the racial minority are unlikely to meet and marry those with <HS. <HS is “low” status amongst the racial majority, making it low status amongst the racial minority, even though a large proportion of the racial minority may still earn only a HS degree. There is then a disconnect between social distinctions and the level of attainment earned.

I end the dissertation by integrating the chapters and discussing future avenues for research. The core things connecting the chapters are: a) a formal model that measures the fluidity of demographic categories; b) a theoretical argument that links categorical fluidity to larger changes in society; and c) a framework that puts a) and b) together to test theories of social change. More generally, I suggest that structural solutions can help answer some of the core questions of sociology, including how power and inequality is maintained and how demographic changes can alter the face of a country.
I discuss future work that will build on the framework developed in the dissertation. I discuss both methodological and substantive projects that will use the relational meaning of demographic categories as the starting point for testing theories. For example, one future project will build on the mixed race category analysis in Chapter 4. The long term goal is to use the classification of mixed race categories to predict changes in racial identity. I will explain why some racial categories have steady populations over time (in terms of who identifies as that category) while others do not. I also discuss possible substantive applications related to period/cohort effects. With these applications in mind, I end the conclusion by discussing the larger contributions of the dissertation.
2. Developing a Model of Categorical Change

Sociologists often describe demographic categories as socially constructed, non-essential, and fluid (Bates and Peacock 1989; Omi and Winant 1994; Anthias 2001). An ethnic label (for example) is not seen as an essential characteristic of an individual (Barth 1969; Zuberi and Bonilla-Silva 2008). The conditions mapped on to an ethnicity may increase or decrease the social distance between an ethnic group and other parts of the population, but ethnicity itself does not cause good or bad outcomes. As a fluid, social construction, a demographic category could change meaning as the conditions surrounding a category change. For example, Italian and Irish immigrants “became White” as they grew less residentially and economically segregated (Ignatiev 1995; Lieberson and Waters 1988; Jacobson 1998).

In practice, we tend to ignore categorical fluidity when measuring social change (Loveman 1999; Harris and Sim 2002; Penner and Saperstein 2008; Brubaker 2009). There are notable exceptions (e.g. Brubaker 2004; Wimmer 2008), but the typical approach is still to take fixed demographic categories, e.g. Black/White, and follow their outcomes over time (Zuberi 2000). We may ask if Blacks have improved economic, health, or residential outcomes (Smelser, Wilson and Mitchell 2001; Downey 2008; Orsi, Margellos-Anast and Whitman 2010). The language of fluid, contingent categories thus does not translate very well into the practice of describing social change.
This paper develops a formal framework for measuring categorical fluidity. The goal is to bridge the gap between theory and practice, and to make it easier for this “radical” sociological notion (that categories are fluid) to be used in everyday research.\(^1\) The core idea is to measure categories as or through interaction patterns—for example marriage or friendship rates. Practically, the demographic categories must be defined across time, and we cannot use the labels themselves as the categories are supposed to be fluid and non-essential (i.e. not defined by the labels). Interaction patterns offer an ideal metric as they concretely capture social distinctions in a population (Haller 1981; Bottero and Prandy 2003). Marriage/cohabitation, for example, is an intimate social tie, meaning demographic groups that do not intermarry or cohabitate have strong barriers to interaction along economic, residential or cultural lines (Qian and Licther 2007).

Different categories may come to reflect the same pattern of relational distinctions over time; one can then use those shifting patterns as a way of defining and equating fluid demographic categories. Theoretically, the claim is that a category is not meaningful just because we measured it. Categories are only meaningful if they map on to observable patterns of behavior (White 1966; Breiger 1981; Carley 1991). Two categories are thus equivalent if the observed behavioral implications of group membership are the same, even if the labels are different (Burt 1991). If College

---

\(^1\) The notion of categorical variability finds expression in classical works of Weber ([1922] 1968), and is core to many current theoretical traditions (Burt 1991; Wimmer 2008; Brubaker 2009).
graduates now marry and make friends like High School graduates used to, then College is the new High School.

The framework combines models of social space and social distance to formalize this idea (Blau 1977; Levine 1990; McPherson, Popielarz, and Drobnic 1992). Working from a social space perspective, I view demographic categories as occupying locations in social space. Locations reflect what interactions are possible for individuals with a given demographic profile (McPherson 2004). From a social distance perspective, categories are placed into locations based on observed rates of interaction (e.g. frequency of marriage between categories) (Bottero and Prandy 2003). Categories are socially close if interaction is likely and socially far if interaction is unlikely (Laumann 1969; Hoff et al. 2002). The question is which demographic categories occupy the same position over time, or which categories have the same range of possible social ties.

There are two basic parts to the model. First, I determine which categories can be collapsed into a single social location (using a Goodman model-Goodman 1981). Categories are collapsed if they have the same pattern of social ties and no barriers to interaction (and are thus relationally equivalent in a given time period) (Breiger 1981; Snip 1987). And second, I determine which category at Time 1 occupies the same location as which category at Time 2. Categories occupying the same social location are relationally equivalent across time. The model is tractable as I hold the structure of social space fixed while allowing the category-location mappings to change. A social space
approach provides the scaffolding for the analysis, while a distance model uses the observed interaction rates to estimate social locations (Goodman 1979; McPherson 1983; Hoff et al. 2002).

After specifying the model, I present an empirical example based on educational categories. There have been large changes in the level of attainment in the population, with the number of people earning a HS or college degree increasing steadily over time (Mare 1995; Snyder, Tan and Hoffman 2006). The question is how the relational meaning of education has changed with the increase in attainment. Empirically, I use information about spouses’ education to capture the changing meaning of education. The data come from the 1940-2000 Censuses. I find two main trends: first, the lower educational categories have grown increasingly less differentiated; and second, the 1940 categories are equated with higher levels of attainment over time. For example, College in 2000 is the social equivalent of High School in 1940. Thus, as a degree becomes more prevalent, or easier to attain, the social distinctions around that degree decrease.

\section*{2.1 Measuring Social Change if Categories are Social Outcomes}

How should one measure aggregate level change if demographic categories are seen as non-essential, fluid social constructions (Bates and Peackock 1989; Loveman 1999; Anthias 2001)? The literature on race/ethnicity offers one of the clearest and most established traditions, although even here a study on categorical fluidity is relatively rare (Barth 1969; Wacquant 1997; Omi and Winant 1994; Zuberi and Bonilla-Silva
Much of this work is in the context of politics or nationalism (Brubaker 2004; Wimmer 2008; Brubaker 2009). During ethnic war or political conflict, there may be moments of high activity and identification centered around racial/ethnic labels, but such moments are often short lived (Brubaker 2004). The “groupness” of a demographic category (or an “us” versus “them” feeling) is thus variable over time and may be politically driven (Brubaker 2009). Lieberson and Waters (1988) looked at more gradual shifts, exploring the decreasing salience of European identities in the US (Lieberson and Waters 1986). Boundaries around European identities became increasingly blurry as everyone was incorporated into a larger White category (see also Ignatiev 1995; Brodkin 1998; Jacobson 1998). It is telling, for example, that Americans have a difficult time describing their ancestral country of origin (Lieberson and Santi 1985; Farley 1991; Lieberson and Waters 1993; Perez and Hirschman 2009).

The goal here is to push the intuition of this work (that demographic categories do not have fixed meanings) to its logical, formal and most general conclusion. If categories are fluid, social constructions, then demographic categories should serve as the outcome in a longitudinal analysis. If the categories have no essential properties,

---

2 Social change is typically measured as the changing differences (e.g. on income, health) between fixed demographic categories (for recent examples, LaFree, Baumer, O’Brien 2010; Miech et al. 2011). Thus, despite recent calls for a more constructionist approach, categories still serve as the basic units in the analysis (Wacquant1997; Loveman 1999; Brubaker 2009). Even studies that do focus on categorical fluidity generally look at individual level changes, asking how racial identification changes over time for respondents (Lieberson 1985; Harris and Sim 2002; Penner and Saperstein 2008). For example, Penner and Saperstein (2008) found that individuals often changed their racial classification and that changes were associated with an individual’s social standing. The question here demands a more macro viewpoint, where the categories themselves are changing over time.
then it is not possible to determine which category at Time 1 is which category at Time 2 simply by the qualities of the categories. It then becomes a theoretical question whether category i is equivalent to category j over time. And more substantively, if the categories are the result of larger social processes, such as physical segregation or economic inequality, then it is natural to treat the categories as the outcome of interest. This pushes the question beyond salience and assimilation (i.e. will minority racial/ethnic groups be incorporated into whiteness?) to something more general: how does the meaning of a demographic category shift over time?3

2.1.1 Interaction Patterns as Metric

A model of social change must first offer a systematic means to equate the categories over time without using the properties of the categories themselves. The first step is to lay out a metric that can properly differentiate and define the categories over time. Practically, interaction patterns, and behavioral measures more generally, offer an ideal starting point for measuring categorical change (Haller 1981; Emibayer 1997; Bottero and Prandy 2003).

Interaction patterns have three key desirable properties as a metric for equating and differentiating demographic categories. First, interaction patterns (marriage, friendship, etc.) describe the rate of contact between categories and are thus defined

3 There is also no substantive restriction to the case of racial/ethnic change.
independently from the categories themselves (Moody and White 2003). This is important as the features of the categories cannot be utilized as the equating metric.

Second, interaction patterns offer a concrete measure of social distinctions in society. Who you spend your time with is clearly important at the individual level. At the macro level, interaction patterns measure the ease of social movement in a population (Qian and Lichter 2007; Diprete et al. 2011). This is a crucial feature of a categorical measure: if two categories are characterized as different based on the metric, then the metric must capture something of clear consequence. There may be many ways to differentiate the categories, but interaction patterns offer a measure of observed distinctions along a metric we know matters for individuals. More substantively, as the categories themselves are not important, we need to determine what a category means socially (Wellman 1988; Emibayer 1997). Measuring a category does not make it empirically relevant or socially real. For example, if Muslims were allowed to marry like Catholics, make friends like Catholics, then Muslims would be Catholics in all ways but name—and the labels themselves cannot be used to define the categories (see Burt 1991 for a similar argument with age categories or the work of Breiger [1981] or Levine [1990] on occupational mobility).

---

4 Two categories may differ on toe nail length but this does not mean that toe nail length will properly capture the implications of a category.
5 It is only clear that Muslim is important (or exists) if the observed behavioral distinctions indicate that Muslim is important, and interaction patterns offer an ideal means to capture such
Finally, interaction patterns are the end result of a multiple stratifying forces, such as economic inequality and physical segregation. Interaction patterns are then the sum of the stratification system, representing an overall measure of distinction. Strong social boundaries exist if repeated, symmetric interactions are unlikely or impossible (Laumann 1965; Chan and Goldthorpe 2007). Categories are socially close if there is no physical segregation and equal status interaction is economically and culturally possible (Torche 2010). By measuring the categories as interaction patterns, one is implicitly taking into account the variety of forces that shape the demographic categories (Bottero and Prandy 2003). And more generally, the specific resources that shape the demographic categories are themselves not particularly important (income, wealth, etc.); what is important is how inequalities translate into limitations of movement in a population, and how this is reflected in the changing meaning of the categories. In this sense, interaction patterns allow the research to sidestep the difficult problem of which metric (wealth, culture?) and simply use a measure that reflects the aggregation of these forces.

In short, interaction patterns capture important, aggregate social distinctions. Interaction patterns can then be used to characterize the demographic categories in distinctions. One can still identify as Muslim, but if being Muslim has no behavioral implications then it is an identity in measurement but not in practice.
terms of social implications—describing what one is allowed to do in society if one belongs to that group. Different categories may reflect the same distinctions, or limitations, at different time points and these changes can be used to measure/equate the categories over time.

2.1.2 Social Space and Social Distance

The key to using interaction patterns to measure categorical change is to combine the imagery of social space with models of social distance. In a social space framework, demographic categories are seen as proxies for social locations, describing where individuals are allowed to go and who they are allowed to interact with (Blau 1977; McPherson 2004). This is an ideal starting point as the measure of social change rests on equating categories with the same behavioral limitations.

Social space models use demographic characteristics (age, education) to proxy the physical dimensions of a social system (McPherson 1983; Mark 1998). The framework builds a multi-dimensional picture of society. Each point in social space corresponds to the intersection of two (or more) demographic variables. For example, consider a space with age on the x-axis and education on the y-axis. One location in the space would correspond to 18 years old, 20 years of education (top left in the space), while another would correspond to 90 years old, 0 years of education (bottom right). I have plotted this space in Figure 1.
The distance between individuals is a function of their location in this multi-dimensional demographic space, and this serves as a starting point for theories of interaction (Blau 1977) and ecological theories of organizations (McPherson 1983). The framework places a structure onto a set of demographic characteristics, organizing the dimensions based on assumed interactional limitations. The assumption is that the probability of a social tie should decrease as socio-demographic distance increases (Blau 1977; Marsden 1987; McPherson, Smith-Lovin and Cook 2001; Smith, McPherson, Smith-
Lovin 2013). High School dropouts are more likely to marry High School graduates than those with a College education (Mare 1991; Smith et al. 2013). Or, from Figure 1, the probability of interaction decreases as we move away from a given point in the education, age space. Individuals with the same education and age are more likely to interact than those who are far apart on education and age. In this sense, homophily organizes the space and all interaction is local and constrained.

A social space framework places the emphasis on locations, rather than categories, and I draw on that analytical distinction to measure categorical change. Locations in social space proxy what is possible, in terms of social relationships, for individuals with a given demographic profile. Categories that occupy the same location at different time points should proxy the same types of social distinctions. The social locations could then be used to equate the categories over time, showing which categories proxy the same types of behavioral limitations across time.

The problem with a social space approach is that the demographic characteristics make up the metric of the space. For example, in Figure 1, the y-axis corresponds to years of education, with each line spaced one year apart. Thus, while the locations are proxies for behavioral patterns, the space itself is not constructed to reflect any observed interaction patterns. This assumption has offered an ideal starting point for past theories (McPherson 1983), but it is insufficient for the purposes here. By definition, social space will be fixed if the axes correspond to the values of education, age, or other demographic
characteristics. Categories with the same values (HS, 40 year old) will occupy the same intersection over time. In this case, the occupied locations cannot be used to equate the categories or to measure social change.

It is thus necessary to estimate the locations in socio-demographic space, and here I draw on models of social distance (Laumann and Guttman 1966; Goodman 1979; Hoff et al. 2002). Distance models use the pattern of social relations (e.g. marriage rates) to place each category into a location in a multidimensional space. Categories serve as the basic units in the analysis, but the metric defining the categories is not based on the categories themselves, but is rather rooted in the pattern of social ties (Levine 1990; Rytina 1992). Demographic categories are “close” in the space if they have high contact frequency and are “far” in the space if interaction between the categories is unlikely. Similarly, two categories with different patterns of interaction with other categories will be far apart (as they will be socially distant and near to the same third party category). The dimensions of the space correspond to latent, unknown dimension that structure interaction patterns—SES, physical segregation, cultural differences, etc. This is in contrast with the social space model where the axes correspond directly to the demographic characteristics. In the social distance model, two categories could be close on one dimension but far in another (e.g. close on language but far on income).
2.1.3 Combining Models

The proposed framework combines models of social space and social distance to measure categories from social locations. Here, the scaffolding of the space, including the number and nature of the dimensions, is known ahead of time when calculating the social locations, but the observed patterns of interaction inform the metric of each dimension. Thus, I still start with a multidimensional space made up of the demographic dimensions. I still assume that the probability of interaction decreases as we move away from a location in the space, and that the categories can be lined up based on decreasing probabilities of interaction. The difference is that I use models of social distance to place the categories into locations (where does HS, 30 year old go?) based on the observed interaction rates. I thus make explicit what is implicit in the social space framework: the categorical locations are not assumed a priori but are rather estimated from observed behavioral limitations.

Figure 2 plots an example of this combined approach with two dimensions and two time points. As in Figure 1, each space has education on the y-axis and age on the x-axis. The metric, however, is now based on (hypothetical) patterns of interaction, where categories are close if interaction is likely and far if interaction is unlikely. I have highlighted three locations in each time point. The green location has high frequency of

---

6 Thus, there is an assumption of transitivity. If category A has a high probability of interaction with B and low probability of interaction with C, then B and C should have a lower probability of interaction than A and B.
marriage (or friendship, etc.) with the red location and fewer ties to the blue location—as the blue and green locations are far apart in the space while the red and green locations are close. This is the same across time. So the green location means having lots of ties to the red location and not so many to the blue. The only thing that changes is the categories “doing” that, or occupying that position. For example Grade 13, Age 50 shifts from the blue location to the red. Similarly, Grades 11-12 and Age 44-45 have collapsed into the green location over time. Thus, Grades 11 and 12 had different interaction patterns in Time 1 but not Time 2.

Figure 2: Example of Estimating Social Locations in Social Space over Time

From the inverse view, we can focus on the occupants of particular locations over time. The blue location, for example, is occupied by Grade 16, Age 60 in Time 2 but Grade 13, Age 50 in Time 1. Age 60 is thus the social equivalent of Age 50. Or, more
colloquially, 60 is the new 50 and College is the new Some College. More formally, Grade 16, Age 50 in Time 2 has the same profile of interaction as Grade 13, Age 50 does in Time 1-low probability of interacting with the red location and even lower probability of interacting with the green location. Two categories are thus equivalent, or occupy the same location over time, if the probability of interaction with the same neighboring and distant locations is equivalent, once we allow for the relabeling of the space (so that the categories can change location).

By using a social distance model, the categories are defined by behavioral information and allowed to vary. By using a social space framework, the categories are more easily equated over time. The dimensions correspond to demographic characteristics in a social space model. This means that only the category-location mappings vary over time, rather than the structure of the entire space. If the categories are fluid, moving targets, then it is necessary to hold something constant when equating the categories over time. For example, if the dimensions corresponded to unknown, latent dimensions, it would be difficult to determine if a category at one time point occupied the same location at another. The dimensions may not represent the same underlying variables over time (e.g. language in time 1 but income in time 2), making comparisons difficult. Similarly, the number of dimensions need not be the same over time. With known dimensions, the locations are directly comparable.
The basic idea is then to leverage the systematic nature of homophily in a society. The salient dimensions are more or less constant across time (so age, education, race, are consistently important in structuring interaction patterns), while the basic organizing principle is always the same—higher demographic distance equals less likely interaction. This offers an empirically robust framework in which to estimate unknown social locations and to track the meaning of categories across time. The locations can be estimated from interaction data, but the basic ordering of the space is known ahead of time.

A social space framework also ensures that the social distance results are theoretically grounded. The goal is to equate categories over time based on inferred social locations. It is important that those locations directly reflect the social meaning of the demographic characteristics. The latent dimensions that structure interaction patterns are less important than the meaning of the categories themselves. The categories, in turn, are the result of multiple stratifying processes. The estimated social locations should reflect the end result of all of these different processes, rather the relative strength of competing variables (e.g. different capitals-Bourdieu 1984). In a social space framework, social locations represent aggregate distinctions, reflecting the range of movement for individuals occupying that location. In a similar manner, a social space framework explicitly treats the demographic categories as an integrated system,
rather than the result of specific (time sensitive) variables. The probability of interaction depends simultaneously on distance in multiple demographic dimensions.7

Formally, the proposed framework builds on log-linear distance and equivalence models (Breiger 1981; Goodman 1981; Levine 1993). These models are used to place the categories into social locations in the demographic space. I first offer a short background section on these models before moving to the framework itself.

2.2 Methods

2.2.1 Background on Distance and Equivalence Models

Distance approaches stretch back to the work of Laumann (1965; 1969), who placed ethno-religious and occupational categories into a multidimensional space based on friendship ties (see also Warner and Lunt 1942; Laumann and Guttman 1966).8 Recent work has typically used the distance-based scores as a measure of occupational status or dominance. The status scores are calculated from mobility or friendship data and then compared to measures of class or SES or used as a predictor of political/cultural behavior (Rytina 1992, 2000; Alderson, Junisbai Heacock 2007; Goldthorpe and Chan 2007).

7 The social space model also captures an intersectional view of social life. The demographic categories are seen as contingent, where the meaning of one category (HS) may depend on the value of another (20 versus 60 year olds). The meaning of the demographic categories can vary not only across time, but across context as well (here regions of social space).

8 Distance models are distinct from Bourdieu’s model of social distance, although the two approaches are conceptually related (Bourdieu 1984). For Bourdieu, the social map is based on the distribution of capitals, rather than the pattern of social relations.
The models employed in a distance approach are generally log-linear models (Levine 1972; Goodman 1985; Levine 1993), although more recent work has built on traditional network models (Hoff, Raftery, and Handcock 2002; Krivitsky et al. 2009). The simplest version of the log-linear distance models are the RC models proposed by Goodman (1979). The model follows the form of:

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta \phi_i \phi_j$$

where $F_{ij}$ is the frequency count in cell $ij$; $\mu$ captures the overall mean, $\mu_i^R$ and $\mu_j^C$ are the row and column means; $\beta$ is a coefficient and $\phi_i$ and $\phi_j$ are the unknown, estimated location scores (which can be constrained to be equal). The model thus produces a latent variable space, conditioned on the marginals. The model is specifically designed to reveal both the scale and ordering of a variable. The distance scale is estimated within the model, and it is unnecessary to know the order of the categories ahead of time. The model can also be extended to multiple dimensions. The latentnet, or network, version of the model can be written as:

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C - \|Z_i - Z_j\|$$

where $F_{ij}$ is the frequency count in cell $ij$; $\mu$ captures the overall mean, $\mu_i^R$ and $\mu_j^C$ are (fixed effects) factors for the row and columns, and $Z$ is the unobserved latent

---

9 There have been a number of alternative specifications proposed in the literature. The goal, generally, is to offer a more flexible functional form for the distance metric, with Levine’s LLD model the most obvious example (Levine 1972). I focus on the RC model in this paper for the sake of simplicity. In principle, however, one could consider both the Levine model and the RC model (as well as a latent network representation) and simply use the model that offers the best fit.
position of i in a multidimensional Euclidean space. As the distance between i and j increases, the frequency of interaction, relative to the marginals, decreases.

Equivalence models use relational data to place categories into larger aggregates.\(^\text{10}\) The relational information determines the categorical groupings, showing which categories clump together and which do not (Burt 1991). These groupings represent the revealed, or observed, social units in the analysis. For example, Burt (1991) placed age categories into structurally equivalent blocks and then looked at the pattern of relations between blocks. Thus, age itself was seen as an arbitrary result of classification, with no necessary social meaning. More formally, in equivalence models, one determines which categories exhibit the same pattern of social ties (Breiger 1981). It is also possible, although less common, to cluster categories based on high frequency of contact (Snip 1987).

Equivalence models use log-linear models on mobility or interaction data to collapse categories into equivalent sets (Breiger 1981; Marsden 1985; Snip 1985). There are a number of approaches, but I draw on the model of Goodman (1981) for the collapsibility of rows and columns in a log-linear model (see also Breiger and Mohr 2004). Formally, the model compares the deviance (or $G^2$) from an independence model on the full table to the deviance from the reduced table (where the specified

\(^{10}\) I denote this set of models as equivalence models, although they really represent a heterogeneous set of models with similar purposes.
row/columns are collapsed). A deviance score close to the deviance of the full table suggests that the rows/columns can be collapsed; if the categories are truly the same, then no information about the relationship between the rows and columns is lost when the two are collapsed. The difference between deviance scores (full and collapsed) follows a chi square distribution and is thus subject to typical statistical tests. Formally:

\[ \log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \mu_{ij}^{RC} \]

where \( \mu_{ij}^{RC} = \mu_{ip}^{RC} \) for \( (i, j) \in \text{ set p in which i and j are treated as equivalents} \)

The model begins with terms for the overall mean and marginals, adding terms for the interaction between all rows and columns. Rows/columns that are collapsed are constrained to have equal interaction terms. The deviance for the model is equivalent to the difference between the full independence model (no collapsing) and the reduced independence model (where the specified rows/columns are collapsed).

The majority of work in the distance and equivalence traditions has utilized cross-sectional data. The few over time studies using distance models have focused on occupations. The goal was often to update scales of occupational status (e.g. Prandy and Lambert 2003), although Rytina (2000) offered a more substantive look at over time change. Rytina (2000) showed that the actual ordering of the occupations was subject to change, making it necessary to consider categorical reordering when analyzing occupational mobility. And more generally, Rytina (2000) suggested that the meaning of the occupations could change over time, and that this was important for understanding
mobility patterns (see also Jencks 1990). Past studies have not, however, made the changing meaning of the categories the focus of the analysis.

2.2.2 Model Overview

Using a combined social distance, social space framework, the model first determines which categories are relationally distinct in each time period. It is necessary to determine which categories are distinct as one cannot assume that a category is meaningful based on the initial measurement alone. The model then equates the categories over time based on inferred social locations. In the cross section, two categories are relationally indistinguishable, and thus occupy the same location, if they have the same pattern of interaction with all other categories and have no barriers to interaction. I denote a set of collapsed categories as an equivalence set. Over time, categories are considered equivalent if they occupy the same social location at different time points, and thus have the same pattern of social distinctions.

I draw on both distance and equivalence models to measure categorical change within a social space framework. An equivalence framework is useful for defining and differentiating categories within time periods. Categories exist if they are relationally distinct from other categories, and equivalence models are designed to answer such questions. A distance framework is ideal for equating categories across time. A distance

---

11 For example, it may not make sense to ask if category i has fewer ties to category j if category i is no longer category i.
model translates the underlying variable (e.g. education) into social locations, making it easier to determine which categories are relationally equivalent over time. Shifts in the collapsed categories signal the emergence or disappearance of categories (e.g. the emergence of Hispanic and the disappearance of Italian) while changes in the category-location mappings highlight the changing meanings of the categories. Together, they reflect the changing boundaries, or social arrangements, for a demographic dimension.

Substantively, the disappearance of a category shows which social boundaries are no longer important. If Italian used to be important for where you lived and whom you married, and now it does not, then Italian no longer socially exists (Lieberson 1985; Lieberson and Waters 1988). You may identify as Italian if pushed, but if it does not matter for any behavioral outcome, then the category is no longer being reproduced in the population (Farley 1991). And given enough time, the label will be dropped altogether from things we identify as (Lieberson and Waters 1986).12

The over time mappings reflect the changing social divides in society. For example, High School and College may represent the same social divide in time 1 as High School and master’s do in time 2. A category may continue over time but change its meaning, where “College” remains distinct, but the behavioral implications of the category changes. Shifts in the meaning of a category reflect changes in the aggregate

[12] It is also possible that labels that are not currently adopted serve as distinct, good proxies for social locations. In this case, an emerging identity may be revealed through changing behavioral distinctions.
allocation of resources in society—the manner in which one set of individuals is allowed to move in society is now equivalent to a different set of individuals in the past. The question of assimilation/incorporation is then only a piece of measuring categorical change.

2.2.3 Model Details

The model begins with a set of disaggregated demographic categories. If there is one demographic dimension (e.g. education), then the categories are simply the measured categories from the raw data. With two or three dimensions, the initial categories are formed from the intersection of the demographic variables (e.g. age X education). I focus on the 1 dimensional case below.

There are three basic steps to the proposed model: 1) calculate the category locations for each year; 2) equate the categories within each period; 3) equate the equivalent sets over time. In placing the categories into locations, the number of dimensions in the space is equivalent to the number of demographic dimensions. An analysis of education would utilize a 1 dimensional space, while an analysis of race and education would utilize a 2 dimensional space. The basic structure of the social space (specifically the number and nature of the dimensions) is thus fixed, while the metric for each dimension is determined by the interaction data.
2.2.3.1 Calculating Initial Locations

The model begins by calculating the latent locations of the categories. For example, one could employ Goodman’s RC model (Goodman 1979). Network approaches or alternative log-linear distance models are also appropriate (Levine 1990; Hoff et al 2002). The input is a contingency table describing the rate of contact between categories. The output is the relational metric of the categories. As a hypothetical example, Figure 3a maps a social space with one dimension and two time points. For concreteness, we can assume that there are two years of educational data for married couples, and that education is measured as <HS, HS, Some College, College, Master’s and PhD. It is clear in Figure 3a that the occupied locations range from low education to high education. Marriages between high school graduates and PhDs are thus less likely than marriages between high school graduates and those with some college (because interaction is denser where distance is low).

2.2.3.2 Equating Categories within Time

Given the initial locations, the model determines which categories (among the full set) are distinct, or socially exist, in a given time period. I use a series of independence tests to determine which, if any, categories should be collapsed for a given year.\(^{13}\) The Goodman model employed here uses two criteria to collapse

\(^{13}\) I assume, for the sake of simplicity, that an initial distance model is run with a disaggregated set of categories, using the full number of available categories.
categories: first, whether the categories have the same pattern of ties to all other categories; and second, whether the categories “interact” with each other as often as themselves (Goodman 1981). The diagonal is thus included in the model, which is traditionally not the case. This model is ideal for my purposes as it tests whether two categories are socially indistinguishable—if they share the same tie pattern and have no barriers to interaction, then the categories are effectively the same.¹⁴

---

¹⁴ It is important to note that these criteria combine the notion of structural equivalence (ties to other categories) with frequency of contact: if two categories have the same pattern of ties to every other category but there exists a distinction between them, then the categories are different. Alternatively, two categories with a high frequency of contact but different tie pattern to other categories are different social entities (as there are different behavioral implications for each).
I start the collapsing process with the two categories closest together in the metric space (based on the distance scores). I thus use the distance scores as a guide for the Goodman model, where I determine which categories should be considered for collapsing based on the inferred distances. Categorical pairs that are far apart have few ties to each other and few ties to the same third parties (as they are not close to the same other categories). They are thus unlikely to be collapsed.

Formally, I take the original table for each year, collapse the two closest categories and calculate the deviance score from an independence model on the reduced table. I then repeat the process on the categories with the next smallest distance between them, keeping the original categories collapsed (see Jackson, Gray and Fienberg 2008 for a similar approach). I repeat the process on the next closest pair, continuing until the table is fully collapsed. For each iteration, I calculate the difference between the full and reduced deviance, as well as the BIC value, based on the difference deviance score. A poor fit will yield a high BIC value, as collapsing the rows/columns results in a large loss of information. At the end of the sequential process, I select the collapsing scheme that yields the best fit based on the lowest BIC value. I take the set of collapsed categories from the best model and reduce the full table accordingly. I end the process

---

15 It is not possible to run independent Goodman tests for each pair of locations. The table, and thus the test, is definitionally dependent on the other categories; thus, any prior collapsing of categories must be taken into account as subsequent Goodman tests are run.

16 The model deviance is the deviance from the full table minus the deviance from the collapsed table and the degrees of freedom are equal to the difference between the full and reduced models (in terms of degrees of freedom). The full table, with no collapsing, has 0 deviance. BIC is equal to Deviance-log(N)*df.
by rerunning the original distance model using the reduced table. I use the distance scores from this model as the final values.\textsuperscript{17} The process is repeated independently for each time point.

2.2.3.3 Equating Categories across Time

After determining which categories can be collapsed for each year, I equate the equivalent sets (or collapsed categories) across time. Formally, for set i in location x in Time 1, we want to identify set j in Time 2 that is in (or close to) location x. For example, Figure 3b offers a straightforward case, where the occupied locations are identical between Time 1 and Time 2. The only thing that has changed is the equivalent sets occupying the locations. Thus Masters/PhD is in the high location (1) in Time 1 but only PhD is in that location in Time 2. Masters/PhD in Time 1 would then be equated with PhD in Time 2. Actual data are, however, likely to be much messier. There are two main complications to tracing equivalent sets over time: first, the total distance in the space may change; and second, the exact locations may not be occupied in every time point. Given these complications, the question is how to systematically determine if two sets are in the same position over time.

Changing Total Distance: I begin with changes to the space itself, where the total distance can shrink or expand, making comparisons difficult. Figure 3c plots the same space as

\textsuperscript{17} The collapsing procedure differs from the blockmodel approach of Burt (1991) as the criteria to collapse categories is more stringent. In this sense, the null hypothesis is to collapse no categories, rather than look for the “best” categories to collapse. A blockmodel approach will tend to yield larger aggregates.
Figure 3b but the range in Time 2 is now wider than in Time 1. The exact numeric locations (on the distance metric) are thus different in the two time periods. For example, it is possible that the income returns to education increased over time, increasing the social divides between educational categories. If, however, one adjusts for total distance, the locations once again become directly analogous across years, and the figures are identical. Masters/PhD in Time 1 is equated with PhD in Time 2 in both Figure 3b and Figure 3c. More formally, I equate equivalent sets based on relative locations, where a set in Time 2 is in the analogous location in Time 1 if it holds the same percentile position in the space at both time points.\textsuperscript{18,19}

\textit{Shifts in Occupied Locations:} Over time comparisons are similarly complicated by shifting locations, where the values on the relational metric need not be identical at different time points. There are two options to equating equivalent sets when the locations are not identical across time: first, using a given year as a baseline and defining equivalent sets relative to the baseline sets; and second, equating equivalent sets based on two years simultaneously.

\textsuperscript{18} The raw locations are scaled, or rounded, so that each location is uniquely identified. The rounding thus does not force two raw locations into the same relative location.

\textsuperscript{19} The relative positions of the categories can only be uniquely compared if the rotation of the locations is made analogous across time. Thus, one must be able to determine what constitutes the “low” end of the space in both years. It is unnecessary to know the exact ordering of all the categories (as this is determined by the model). It is only necessary that a location in one year (e.g. 0) can be directly interpreted as equivalent in another year. Formally, I select the rotation that minimizes the distance between categories over time. The assumption is that slow change is more likely than a complete overhaul of the social system. Substantively, the model assumes that there are a sufficient number of categories ordered in the same manner across years to determine which end is equivalent across time.
With the baseline option, I identify equivalent sets in Time 2 that are closest to the location in Time 1 and consider them equivalent. For example, in Figure 3d, College in Time 2 is the closest set to College in Time 1, even though College has shifted slightly to the right by Time 2.

The second option avoids the use of a baseline year, equating equivalent sets based on shared locations across time. Here, set i in Time 1 is equated with set j in Time 2 if and only if j occupies the closest Time 2 location to i and i occupies the closest Time 1 location to j. The two options are identical in Figure 3d, but this need not be the case. For example, in Figure 3b the baseline option will equate <HS, HS, Some College, College-College and Masters/PhD-Masters/PhD; in contrast, the no baseline option yields <HS, HS, HS-HS, College-College and Masters/PhD-Masters/PhD. For the non-baseline option, an equivalent set at one time point need not have an analogous set at another. The baseline option is an appropriate choice where a clear, substantive baseline exists, while the non-baseline option is the appropriate choice otherwise.

I offer a simple summary of the model in Figure 4. I assume, as in Figure 3, that the raw data measures education as <HS, HS, Some College, College, Master’s and PhD

---

20 The differences arise as the .3 location in Time 2 (Some College) is closest to the .25 location in Time 1, while the .5 location in Time 1 is closest to the .3 location in Time 2. When the location must match for each year (as in option 2), the .3 location in Time and the .5 location in Time 2 are not equated with anything across time. The baseline option will equate .5 to .3 when Time 1 is the baseline but will equate .3 to nothing when Time 2 is used as the baseline (as the .25 location in Time 2 is closer to the .25 location in Time 1 than the .3 location).
and that there are two years of interaction data (e.g. marriage or friendship). Panel 1 maps the initial location estimates for the two time points. In Time 1, for example, <HS is more likely to marry HS than Some College. The second panel takes the initial locations and determines if any categories can be equated for a given year. For example, Master’s and PhD are collapsed in Time 1 but not in Time 2. The process then takes the raw locations and scales them to a 0-1 metric (see Panel 3). From Step 2 to Step 3, the axis shifts from -.65:.65 to 0:1. Substantively, HS and Some College occupy different locations in Step 2 but the same locations in Step 3—as they occupy the same relative position in the space. In Step 4, I determine which categories in Time 2 occupy the analogous locations in Time 1. I have traced out where the categories move from Time 1 to Time 2, and colored the categories the same if they occupy the same location (here using the non-baseline approach). For example, Some College occupies the “median” social location in Time 1, with equal probability of marrying to the very top as the very bottom. Allowing for the possibility that the top and bottom has changed, College now occupies the median social location in Time 2, with the same profile of interaction.
Figure 4: Schematic Description of Fluid Category Model

The results can be used in a primarily descriptive manner (what does race, education, etc. look like over time?) or can be used to test hypotheses. A theory of social change is then a theory of categorical change. The fluid category results can be used as a touchstone, comparing the theory’s predictions to the true categorical shifts. A theory could predict the simple existence/non-existence of the categories or predict the categorical equivalents at two time points. Formally, a test would take the set of disaggregated categories in Time 1 and predict the equivalent category at Time 2. The units are all Time 1-Time 2 categorical pairings. The dependent variable is a 1 if the pair is equated and 0 otherwise. It is also possible to take a more continuous approach. Here,
the dependent variable is the difference in relative location between category i in Time 1 and category j in Time 2 (e.g. the value is zero if they are equated). In either case, a theory is evaluated by how well it predicts changes in the categorical definitions.

2.3 Empirical Example: Education 1940-2000

As an empirical example, I examine the meaning of educational categories from 1940-2000. Using census marriage data, I use the proposed framework to equate educational categories over time. The shifting definitions serve as the key measure of social change. Past work on education has documented a steady increase in attainment over time (Mare 1995; Snyder, Tan and Hoffman 2006). While it is now normative to earn a high school degree, this was not the case in the early part of the 20th century. Additionally, the outlets for attainment past high school have increased dramatically, making some form of college a more realistic goal for a wider portion of the population (Kalogrides and Grodsky 2011). It is now the case that the majority of students expect to go to college (Goyette 2008).

The increase in educational attainment is often viewed in terms of social competition over the resources in society (Collins 1971; 1979). Those with previously high education raise the bar, in whatever way possible, when those with lower education increase their attainment (Lucas 2001). In terms of the current model, the changing categorical definitions can be interpreted as the result of such battles. The push for educational attainment erases, or obscures, previous distinctions while
simultaneously creating new ones. Categories that were relationally distinct but now form a single group are thus the result of past battles. As they are socially identical, there is no longer anything to fight over. And more generally, the over time mappings show which categories represent the comparable pressure over time. The previous tension over High School may be the current tension over College.

The question is how the social meaning of education has changed over time. As a baseline expectation, the educational categories from 1940 should be equated with increasingly higher levels of attainment. It now takes a higher level of education to attain the same range of movement in society. The exact nature and pace of this inflation is, however, of more substantive interest. It may be unsurprising that High School is “worth” less today than in 1940, but it is less clear what categories exist in a given year and how the social boundaries have changed over time. If High School is now normative, is it still distinct from grades 11, 10, etc.? And more generally, how difficult is it to bridge the same social gap in 1940 as in 2000, in terms of educational years?

2.3.1 Data, Measurement and Models

I follow the majority of studies in the educational homogamy literature and use census data over the last 70 years as the basis for analysis. Homogamy studies measure social change as the changing rate of contact between fixed educational categories, and thus use the same basic inputs to measure population dynamics (e.g. Mare 1991; Kalmijn 1991a). Following Rosenfeld (2008), I use 20 year non-overlapping cohorts. I use
the 1940, 1960, 1980, and 2000 censuses and restrict the analysis to married couples between the ages of 20 and 39.¹¹ This parallels the majority of past work as I restrict the population to younger individuals. ¹²

The U.S. Census measures education as years of attainment up to the 1990 survey. From 1990 on, education is measured as the highest degree attained, assuming the individual has more than a high school education. For the 1940, 1960, and 1980 data, I use all 18 years of education as the initial categories in the analysis (the exact number varies slightly by year). For the 2000 data, I use the available grades/degrees as the measure of education. This includes 0-4, 5-6, 7-8, 9, 10, 11, HS, Some College, Associate Degree, Masters, Professional Degree and PhD. The data from 1940-1980 are thus not directly comparable to the 2000 data, and such discrepancies must be kept in mind when interpreting the over time trends. Some of the comparability problems are reduced, however, as the focus is on locations rather than particular categories, which may not exist in different surveys. More generally, one of the side advantages of the approach is that the problem of measurement consistency is reduced—as the method seamlessly handles shifting categorical measurement over time.

¹¹ The non-overlapping cohorts ensure distinct comparisons: where people of the same age in one year are compared to people of the same age in another year, with no overlap in the cohort population.

¹² I include all prevailing marriages in the analysis, following Hou and Myles (2008) (as opposed to including only newly weds, which is often done). The analysis thus captures both the formation and dissolution of marriages, reflecting the aggregate patterns of current interactions (i.e. marriages).
The analysis uses the marriage data to equate the categories over time. I use the RC model as a basis for placing categories into a one dimensional space in each year, collapsing categories where appropriate. I then equate the collapsed categories over time based on shared (relative) locations.23

2.3.2 Results

2.3.2.1 The Relational Structure of Education: 1940-2000

Figure 5 presents a snapshot view of education over time. The figure plots the relative location (on a one dimensional scale) of each category, placing the “high” position at 1 and the “low” position at 0. Overall, the results reflect the ranked nature of educational attainment, where the categories are ordered from lowest to highest grade. The clustering also generally reflects recognizable educational labels. For example, grades 13-15, or “Some College”, tend to cluster together. Similarly, grades 10 and 11 are consistently close, capturing individuals who attended high school but did not graduate.

The categorical locations do not, however, map simply on to grade level. In 1940, for example, grades 1 and 2 are socially indistinguishable, while grades 4 and 5 are not. Thus, a year divide can, but need not, be a meaningful social divide. The exact nature of this can vary over time, where the same educational divide is meaningful in one year but not another. For example, grades 4 and 5 are socially distinct in 1940 but not 1960.

23 I use the non-baseline option for equating equivalent sets.
Additionally, the data resist any simple collapsing into 4 or 5 neat aggregates. For example, grades 7 and 8 have distinct social patterns through the 1980 data. It is difficult to collapse them into a “low” education category, covering grades 0-9 or 0-11. And more substantively, this means that individuals in 1940 with 7 years of education had different social outcomes as individuals with 8 years of education. Such a distinction may feel odd from a contemporary vantage point, but the distinctions were real at the time.

More generally, Figure 5 points to considerable over time change. The categories between 0 and .15 shift from grades 0-3 in 1940 to grades 0-6 in 1980 to grades 0-9 in 2000. The lower categories are not only equated with higher categories over time, they
are generally collapsed into one large social grouping. For example, we see the emergence of grade school, 0-6, as an equivalence set. Categories that existed in 1940, such as grades 3, 4, 5 etc., are no longer distinct, socially salient categories by 1980.

There are similar inflationary shifts in the middle of the distribution, although the initial categories remain largely distinct over time. The “median” social category changes from grade 8 to grade 10 to high school by 1980. The social divide between grade 0 and grade 7 in 1940 is equivalent to the social divide between grades 0-4 and grade 12 in 2000. It thus takes more years of education to bridge the same social divide in 2000 than in 1940. Or, it is as difficult to move socially from 0-4 to 12 in 2000 as it is to move between 0 and 7 in 1940.

We can also follow the path of particular grades over time. For example, grade 12 moves from a relative location of .71 to .61 to .48 and .37 in 2000 (traced out in the figure). This means that there were few marriages between HS and the “bottom” educational category in 1940. Over time, however, there is a clear increase in the number of social ties between HS and the bottom, allowing for the possibility that the bottom category has shifted over time (from no school to Grade School). If most people get a high school degree, high school can no longer serve as a distinguishing marker in society and it moves closer to the bottom of the educational space—at least in social terms. It is important to note, however, that High School remains a distinct social entity, never collapsing with neighboring categories.
2.3.2.2 The Relational Structure of Education: Theoretical Implications

The results point to a general inflationary pressure in the educational categories. This inflationary pressure is not strictly linear, however, and the moments that break with the general trend are empirically and theoretically important. In the case of linear inflation, the categories follow a regular pattern over time. For example, grade 12 is equated with grade 13 between 1940 and 1960, as well as between 1960 and 1980. The same process of inflation is thus repeated between 1940-1960 and 1960-1980. This means that any grade equated with grade 12 between 1940 and 1960, here grade 11, will be equated with grade 13 between 1940 and 1980. The equating structure between 1940 and 1980 is a replicate of the 1940-1960 movement. One can then predict the 1980 equivalent of a 1940 category using the 1940-1960 patterns.

Not all categories exhibit regular over time change, however. Grade 6, for example, is equated with grade 7 between 1940 and 1960 and grade 10 between 1940 and 1980. Grade 6 should, however, be equated with grade 8 based on a linear inflationary pattern (as grade 7 is equated with grade 8 between 1940 and 1960). Non-linear movement occurs when multiple forces push categories in the same direction. The shift of grade 6 to grade 10 offers a clear example. Grade 10 is a near neighbor of grade 12, but the relative distance between grade 10 and grade 12 is large relative to the entire space. Grade 10 is also situated next to grades 0-9. Grades 0-9 are disaggregated in 1940 and 1960 but are largely compressed by 1980. This results in a heavy downward pressure.
The lower educational categories are collapsing but a union between HS and grade 10 is unlikely, leading to a shift towards grades 0-9. More generally, the largest changes in meaning are likely to occur on the edge of strong social boundaries: much of the space is limited, making the occupants more responsive to other shifts from nearby social neighbors.

2.4 Conclusion

This paper offers a new framework for measuring social change in a population. Typical models describe social dynamics as the changing social, economic or political distance between fixed demographic categories (for example, Kalmijn 1993; Orsi et al. 2010; Davenport, Soule, and Armstrong; 2011). Such models are, however, difficult to reconcile with the rhetoric of sociology — where categories are often described as fluid, non-essential and socially constructed. The proposed framework bridges the gap between rhetoric and practice by making the categories themselves the key measure of social change.

Combining models of social space and social distance (Goodman 1979; Goodman 1981; Levine 1990; McPherson 1983), the framework uses interaction data (such as marriage rates) to differentiate and define fluid demographic categories over time. Demographic categories are seen as proxies for social locations, describing which individuals are allowed to interact with which individuals. Different categories may represent the same range of social movement, or occupy the same social locations, at
different time points, and these changes are used to measure social change. Formally, there are two pieces to the approach: first, determining which categories are relationally identical in a given time period; and second, determining which category at Time 1 is equivalent to which category at Time 2, where two categories are equivalent if they occupy the same social location over time.

The framework can be used to test theories of social change or can be used as a primarily descriptive tool (e.g. how has race changed over time?). A theory could predict categorical equivalents (i.e., which category at Time 1 is equivalent to which category at Time 2?) or predict the existence of different categories across time. Categorical shifts reflect the changing distribution of resources in a population, where the movement of one set of individuals is analogous to a different set of individuals in the past. The approach could also be used to test theories of dominance, where power is reflected in the shaping of the demographic categories. For example, a dominant group may shape the meaning of a demographic variable, while a subjugated group could not.

As an empirical example, I examined over time trends in education using Census marriage data from 1940-2000. The proposed model yields three basic findings: first, that there has been inflationary pressure in the educational categories, so that High School in 1940 is the social equivalent of College in 2000; second, that the lower educational categories now represent a recognizable social unit; and third, that the largest changes in categorical meaning occur on the edges of strong social boundaries.
The proposed framework captures social dynamics by tracking demographic categories over time. The shifting categories paint a clear picture of the social divides in a population. The changing categories do not, however, tell us very much about the changing level of inequality, or differentiation.24 As an extension, one could measure system level differentiation through traditional log-linear models (see also Breiger 1981; Burt 1991). Formally, I take the equivalent sets (or the set of collapsed categories) in each year and run a log-linear model. I then compare the log-linear models across years. For example, I present quasi-independence models for the education results (see Figure 5). The models use Census data from 1940-2000. A quasi-independence model includes one dummy term for the cells on the diagonal and captures the overall rate of in-group bias. The results yield log odds of 1.36, 1.07, 1.18, and 1.30. The data thus point to a U-shaped pattern, where there is an initial decrease in differentiation followed by an increase that brings the results back to the 1940 levels (see also Schwartz and Mare 2005). This is consistent with past studies that have found an increase in homogamy since the 1960’s (Kalimijn 1991b; Hou and Myles 2008; Blossfeld 2009).

It is also possible to examine differentiation by using relative distance as the main predictor in the analysis. For each year, the analysis models the frequency of contact between i and j (relative to chance) as a function of the relative distance between

---

24 The categories are specifically equated conditioned on the total distance (on the relational metric) in a given year.
i and j. This yields a curve, or fitted line, which can be compared across years (see also Linkletter 2007). A steeper curve implies more inequality—as the odds of a tie decreases faster for the same unit of distance. For example, Figure 17 in the Appendix presents four plots based on the education results. For each year, I plot the log odds of a marriage as a function of relative distance (between equivalence sets). The odds of a tie predictably decrease as relative distance increases. Of more substantive importance, the rate of decrease is remarkably similar across time. In each block, the decreasing rate of contact is steepest at the lowest distances but flattens out as distance increases. Thus, the overall rate of matching has a U-shaped pattern, but the detailed mixing patterns are quite similar over time.

The method was presented here in terms of over time interaction data, but the approach is clearly appropriate for any contextual study (cross nation for example) with relational data, such as mobility or friendship rates (Moody 2001). Additionally, the empirical results were limited to one demographic dimension, but future work could examine multiple dimensions. Future work could also allow for asymmetry by gender. Perhaps more importantly, one could empirically consider nominal variables, where there is no implicit ordering to the categories (Qian 1997).

Note that a model with no variables offers a simplified view of the space, but may also obscure some of the heterogeneity at a given distance. For example, the model would include the same predictor for all terms on the diagonal, as distance is 0, but there may be stronger in-group biases for some locations than others. The ordering of the categories may, however, be less clear over time, and this would make it more difficult to equate the categories (as the rotation of the space is less obvious).
More substantively, sociologists often describe demographic characteristics in terms of fluidity. A demographic category is not fixed, but fluid, is not essential, but relative, is not “real” but socially constructed. If we truly believe that categories are fluid, arbitrary social constructs, then we need a model that reflects that belief: a fluid category analysis then becomes an ideal, perhaps necessary, part of any over time study of social change.
3. Is Hispanic a Racial Category?

The last 40 years have seen dramatic changes in the racial/ethnic makeup of America. The numerical dominance of the White majority has declined as the Hispanic population has grown (Saenz 2004). Within the Hispanic population, there is a growing number of people claiming “Other” racially, in lieu of White, Black or Native American. In the 1980 Census only 5 percent of Hispanics claimed Other racially, but this increased to 48 percent by 1990 (for those under 40). Such a stark increase has led past work to posit Hispanic as a new racial/ethnic group distinct from the previously dominant Black-White division (Rodriguez and Cordero-Guzman 1992; Gomez 2007). The increase in “Other” Hispanics may reflect a rejection of White or Black as a racial identity. Other work has argued that Hispanics will eventually be incorporated into White, following the experience of past immigrant groups (Gerstle 1999; Bean and Stevens 2003; Lee and Bean 2004). The fact that the Hispanic population is growing so quickly makes it all the more important to understand the role of Hispanic in the racial landscape, and to properly interpret the shift in racial identification (Perez and Hirschman 2009; Frank, Akresh and Lu 2010).

This paper makes sense of Hispanic racial change by applying a new, relational framework to the problem of racial dynamics. I view racial/ethnic categories as fluid, social constructions (Barth 1969; Brubaker 2004; Wimmer 2008). Racial/ethnic categories
can change meaning over time, and these changes are used to interpret the larger
demographic trends—here the shift in racial identification amongst Hispanics.

I measure racial categories through time using the approach of Chapter 2, where
I define racial categories using observed social distinctions, rather than the racial labels
themselves. The framework uses interaction patterns, such as marriage rates, as the
measure of social distinctions. Marriage/cohabitation is an intimate social tie, meaning
demographic groups that do not intermarry or cohabitate have strong barriers to
interaction (along economic, residential or cultural lines). Racial groups allowed to
marry/cohabit in the same manner are treated as the same, even if the labels are
different. If people who identify as White-Hispanic now interact with other categories as
White used to, then those who identify as White-Hispanic are by this logic now White. I
draw on a social space tradition to formally measure the categories through time
(McPherson 1983; Bottero and Prandy 2003). Categories are placed in social locations
based on observed rates of interaction. Categories are socially close if they have high
rates of marriage/cohabitation and socially far if interaction is unlikely (Beshers and
Laumann 1967). I then follow the racial/ethnic categories as they move across social
locations. Categories occupying the same social location over time are treated as
equivalents.

Race/ethnicity is particularly amenable to a constructionist viewpoint as the
categories themselves are historically contingent and serve as points of contention in a
population (Lee 1993; Hitlin et al 2007; Brubaker 2009). Individuals may find the available racial options insufficient (i.e. no Hispanic option), while others may find it offensive to see certain options there at all (being asked to identify as Hispanic). Similarly, the problem is a dual one, with the government defining the range of legitimate categories and the population reacting to the available options (Wimmer 2008; Frank et al. 2010).

I use this tension in racial identification to connect the demographic trends to the changing relational meaning of the racial categories. The increase in Other-Hispanic can be seen, in part, as a reaction to the available racial options. Individuals are generally forced to answer both a racial and Hispanic question. The question is what it meant for so many more individuals to identify as Other racially (rather than White) and Hispanic ethnically. I offer three possibilities. Each formally connects the racial demographic shifts to changes in marriage/cohabitation patterns, and thus social locations.

First, Hispanic could be emerging as a unified racial/ethnic category distinct from Black or White. Here, the racial question is no longer salient for those identifying as Hispanic as Hispanic itself is the dominant identity (Campbell and Rogalin 2006). Marking Other is no different than marking White. Half the Hispanic population thus identifies as Other while the other half identifies as White. This would suggest the existence of an overarching, cohesive Hispanic category, and the dissipation of any social divisions within the Hispanic population. We would then see the same pattern of
ties amongst all racial/Hispanic groups and no barriers to marriage/cohabitation. All the Hispanic categories would occupy the same social location.

Second, we could be witnessing the schism of Hispanic, between those who are racially Hispanic (i.e. claiming Other racially) and those who are racially White (i.e. claiming White racially and Hispanic ethnically) (Frank et al 2010). The increase in Other-Hispanics thus signals the division of a potential Hispanic category. Or, the demographic shifts signal the rejection of a Black/White identification for some of the Hispanic population. Empirically, Other-Hispanic would move socially away from the other non-Black Hispanic categories. Other-Hispanic would thus act like a non-Hispanic category, or occupy a non-Hispanic social location.

Third, the racial change amongst Hispanics could be a purely demographic phenomenon. Here, the relational meaning of Hispanic is constant over time. The Hispanic categories “do” the same things in terms of marriage/cohabitation, even as the population shifts racially from White to Other. In this case, it remains an open question whether Hispanic will eventually be incorporated into an overarching White category.

Overall, the paper will describe the organization of race in America, showing how it evolves in an environment of demographic flux. The paper begins with a background section on measuring categories relationally. I then theoretically link the three hypotheses, cohesive Hispanic category, schism in Hispanic, demographic change only, to shifts in marriage/cohabitation. Empirically, I use Census marriage/cohabitation
data from 1980 and 1990 to describe the structure of race in the US. I then assess changes in Hispanic from 1980-1990 using shifts in social locations as a guide.

3.1 Race/Ethnicity as a Social Construction

Few scholars would claim race itself as a causal explanation (Martin and Yeung 2003). Race/ethnicity is not an essential characteristic of an individual, and racial/ethnic categories, serving as proxies for skin color and nationality, do not cause good or bad outcomes (Barth 1969). Rather, race/ethnicity continues to be important because of the conditions, along economic, political, and social lines, correlated with different racial/ethnic categories (Smelser, Wilson and Mitchell 2001). There is then a distinction between a category and the social position associated with that category. In this sense, race/ethnicity is the result of larger social forces, such as physical and occupational segregation. This opens the possibility that the meaning of a racial category could change as the economic/political conditions mapped onto a category change (Loveman 1999; Brubaker 2009).

Past research has offered two broad approaches for measuring changes to racial categories (Loveman 1999; Wimmer 2008; Brubaker 2009). First, a number of studies have explored the boundaries defining the racial categories. For example, Loveman and Muniz (2007) found that Puerto Rico “whitened” in the early 20th century, in large part because children of cross-race parents were more likely to be classified as white over time. Thus, rather than individuals moving across fixed racial boundaries (e.g. based on
changes in economic status—), the boundaries themselves shifted to a more inclusive definition of “white”. Or, in the US, there may be a consistent set of Black, White, Hispanic, and Asian categories even as the classification rules change. This would change the distribution of people across racial categories, and could, eventually, lead to the disappearance of a category altogether.

Brubaker (2004) offers a more radical approach, viewing ethnicities as temporally situated events (see also Wimmer 2008). Theoretically, he distinguishes between categories, things we measure, and groups, categories that have crystallized into an “Us” versus “Them” distinction. A category could be a group but need not be: there may be moments of high activity and identification followed by periods of more modest ethnic divisions. For example, an ethnic identity may be particularly salient during political conflict. We can then ask whether a category crystallizes into a meaningful social group. The question is what forces or political organizations lead to shifts in the level of “groupness” of a racial category. Here, categorical change can be overtly political, where different factions vie over the ability to define racial categories—even for political gains or to assert the existence of an ethnic movement (Wimmer 2008). Racial categories then serve as a field of contention rather than a causal force in itself. For example, Frank et al. (2010) find that Latino immigrants see an economic advantage in being White but are often excluded from the benefits of this self classification.
The work of Lieberson and Waters falls somewhere between these approaches (Lieberson and Waters 1988; Waters 1990). They empirically examined the decreasing salience of European ethnicities, such as Italian and Irish. Irish and Italian decreased in salience as these groups grew less residentially and economically segregated. Over time, boundaries around European identities became increasingly blurry as everyone became White (see also Ignatiev 1995; Brodkin 1998; Jacobson 1998). It is telling, for example, that Americans have a difficult time describing their ancestral country of origin (Lieberson and Santi 1985; Farley 1991; Lieberson and Waters 1993; Perez and Hirschman 2009). Thus, ethnic categories can have variable salience over time, while the boundaries around a category (e.g. the Irish are now incorporated into White) can expand if the salience around a category is low enough for long enough. If Italian used to be important for where you lived and whom you were allowed to marry, and now it does not, then hyphenated labels such as Italian-American may no longer socially exist (Alba 1985; Alba 1990; Lieberson 1985; Lieberson and Waters 1988).¹

This paper offers an alternative approach to measuring racial/ethnic categories through time. Here, racial/ethnic categories shift along fixed social positions, as opposed to boundaries shifting over fixed individuals (e.g. Puerto Rico whitening) or salience shifting over fixed categories (e.g. Italian becoming less salient). The goal is to describe a

¹ An individual may identify as Italian or Irish if asked, but if this identity does not matter for any behavioral outcome, then the category is no longer being reproduced in the population in any meaningful way (Farley 1991).
racial system by the positions of the racial/ethnic categories over time. Social positions are defined by observed social distinctions between categories, measured by patterns of interaction (such as marriage and cohabitation). The approach draws on aspects of existing models but allows researchers to answer core questions about racial and population change that are difficult to answer currently.

For example, the approach generalizes the question of categorical salience. I ask which category at time 1 occupies the same social position as which category at time 2. The existence, or “groupness”, of a category becomes part of the question. It is important to ask if Italian is a separate category from White (i.e. does it socially exist?), but this ignores the shifting roles that different racial groups play over time. Collapsing is not the only process by which categories can change meaning—especially if one sees the categories themselves as a field of contention. Or, past approaches assume that race/ethnicities can be placed into stable, large categories (i.e. White), without considering that categories can remain distinct but still change meaning. For example, Hispanics may increase their rate of marriage to Asians. This would change the meaning of Hispanic even though Hispanics are not assimilated into Whiteness. It is thus important to capture the changing meaning of the racial/ethnic categories, while recognizing that categories (Irish) can be subsumed into other categories (White).

Additionally, the framework formally connects the meaning of racial/ethnic categories to demographic shifts in racial/ethnic identification. I ask how changes in
classification relate to changes in categorical meaning, rather than take shifts in classification as the measure of racial change itself. In the empirical case here, individuals identifying as Hispanic were increasingly likely to claim “some other race” between 1980 and 1990 (Frank et al. 2010). This change may correspond to the racialization of Hispanic, so that Hispanic becomes a racial category in its own right (Logan 2003; Michael and Timberlake 2008). But the increase in Other-Hispanic is an empirical result, and not evidence of racialization in itself. It is possible that more people now identify as Other-Hispanic but the meaning of that ethnic/racial category is exactly the same. Thus there are individuals moving over fixed categories. The key is having a measure of racial fluidity that will reveal the substantive meaning of shifts in identification. We can then determine if a new category has emerged or if more people are simply claiming an old, existing category.

3.2 Interaction Patterns, Social Locations, and Racial Categories

I draw on the approach presented in Chapter 2 to measure the racial/ethnic categories over time. Demographic categories are measured through interaction patterns, such as marriage or friendship rates. The meaning of race/ethnicity is then a “relational meaning”, where categories are equated based on behavioral indicators. Interaction patterns offer an ideal metric as they concretely capture social distinctions in a population. Who you spend your time with is clearly important to people. It is then telling when there are few observed social ties between demographic groups. Interaction
patterns thus show us where there are observed/revealed distinctions in a population and where there are not.

This follows the literature on racial homogamy, which studies the frequency of cross-race marriages (Blackwell and Licther 2004; Rosenfeld 2008; Qian and Licther 2011). Intimate social ties are used to measure the social distance between racial groups. Close social ties serve as an aggregate measure of distinction, capturing the end result of many forces that divide individuals in population—such as residential segregation and economic inequality (Bottero and Prandy 2003; Torche 2010). Social distance is thus high if there are structural, cultural or status barriers to forming close, reciprocated relationships. For example, the barriers to Black/White marriages were quite strong in the 20th century, with few cross-race marriages (although this weakened somewhat more recently) (Kalmijn 1993; Qian and Licther 2007).

The homogamy literature, however, takes the racial categories as fixed and known (Black, White, Hispanic, etc.), when looking at interracial marriage rates. The goal here is to use intermarriage rates to define the categories themselves (Bottero and Prandy 2003). For example, if Hispanics were allowed to marry like Whites, make friends like Whites, then Hispanics would be White in all ways but name, and the labels themselves cannot uniquely define the categories if they have no essential, fixed meaning (see Burt 1991 for a similar argument with age categories or Breiger’s [1981] work on social class).
The key to extending a traditional approach is to recognize that different categories may come to reflect the same pattern of distinctions over time. One can then use those shifting patterns as a way of defining and equating fluid demographic categories. Categories are thus considered equivalent (across and within time) if the behavioral implications of group membership are the same, even if the racial label is different.

The framework draws on social space and social distance traditions to formalize this idea (Blau 1977; McPherson, Popielarz, and Drobnic 1992; Hoff et al. 2002). I measure categorical change by placing categories into social locations based on observed rates of interaction (e.g. frequency of marriage between racial categories). Locations reflect what interactions are likely for individuals with a given racial/ethnic profile. Categories are close if interaction is likely and far if interaction is unlikely (Laumann 1969). Similarly, two categories with different patterns of interaction with other categories will be far apart (as they have different distances to the same third party category). The racial/ethnic categories are thus defined by their social distance to all other racial/ethnic categories (Levine 1990; Rytina 1992). Categories with the same pattern of interactions will occupy the same social location. The question is which categories occupy the same positions over time, or which categories have the same range of possible marriage partners (friends, etc.).
The framework diverges from models that define racial/ethnic categories by their social distance to Whites (Lee and Bean 2007). Such an approach assumes that assimilation into Whiteness is the key metric in measuring racial categories (Alba 1990; Portes and Zhou 1993; Berry 2006). For example, a typical model may compare the relative distance of Asians and Hispanics to Whites (Qian 1997; Qian 2002; Portes and Rumbaut 2001; Lee and Bean 2004). Here, the question shifts to the overall position in social space (McPherson 1983; McPherson and Ranger-Moore 1991; McPherson 2004). Racial/ethnic categories are simultaneously defined by their social distance to all other categories, not just the White majority. For while it may be the case that the racial order can be described by a Whiteness continuum, this is an empirical question and is not assumed a priori (see also Robnett and Feliciano 2011).

Figure 6 plots an example social space based on 8 racial/ethnic categories. This is a completely hypothetical example used solely for demonstration purposes. I assume, for this simple example, that the racial space can, in fact, be represented by one dimension ranging from White to Black (representing the farthest social divide). In actuality the space may be considerately more complicated, with more than one dimension necessary to fit the data. The figure includes the inferred social locations from two years of hypothetical marriage data. For example, White-Hispanics in Time1 are more likely to marry Whites than Native Americans. The figure also follows the categories as they shift position from Time 1 to Time 2. We can see that the “median”
social category changes from Native American to Asian. Thus, Asians in Time 2 do the same things, or move socially in the same ways, as Native Americans did in Time 1. And more specifically, categories occupying the median social location are equally likely to intermarry with categories at the edges of the social space. Allowing for the possibility that the “bottom” and “top” locations have changed occupants over time (for example, the “bottom” shifting from White to White+White-Hispanic), this profile of interaction, equal chance marrying top/bottom, corresponded to Asian in Time 1 but Native American in Time 2. In this sense, Asian in Time 2 is the Native American of Time 1.

The Figure also makes clear that distinctions can completely disappear. For example, by Time 2 there are no relational differences between White and White-Hispanic, and they are treated as one category, occupying the same position. The boundaries separating White and White-Hispanic have thus weakened over time in this example.

Figure 6: Example Social Locations: Racial/Ethnic Categories at Two Time Points
I use these two ideas, that categories can change social locations and categories can be collapsed into a single location, to test theories of demographic change. Substantively, I use the changes in categorical meaning (i.e. social locations) and changes in social boundaries (i.e. a weakening of social distinctions between demographic categories) to interpret the changes in racial identification amongst Hispanics.

3.3 Hispanic as a Racial Category?

I follow past work that views racial change as a conflict over the meaning and recognition of racial categories (Brubaker 2004; Wimmer 2008). The two main actors are groups in the population, those who are classified and identify racially, and the government, the political organization that sets the bounds for racial identification. Individuals may find the available options incomplete, or include options that do not fully encapsulate the racial identity and experience of a particular group (Lee and Bean 2004). How a group navigates the racial/ethnic question under such conditions can be quite telling. The government, in contrast, serves the role of legitimizing certain racial groups as “real” races while downplaying the experience of others (Frank et al. 2010).

This dual structuring is helpful in interpreting the large demographic shifts seen for Hispanics. Even in the 10 years between 1980 and 1990, there was a substantial increase in the number of people identifying as Hispanic. Surveys such as the Census ask individuals separate Hispanic and race questions. I have included the Census questions from 1980 in the appendix. This forces individuals who identify as Hispanic to
also claim a racial identity (Black, White, Native American, etc.). The increase in the number of people identifying as Hispanic coincided with changes to the internal racial composition of Hispanic. As the number of Hispanics increased, so too did the proportion of Hispanics claiming Other as a race (Rodriguez 2000; Tafoya 2005; Frank et al. 2010). The proportion of Hispanics claiming White as a race has, in contrast, decreased over time (from 94% to 50%). The proportion claiming Black or Native American has been more or less constant. Thus, the growing population of young Hispanics are disproportionally more likely to claim Other racially, while past generations were more likely to see themselves as White—or, perhaps, more accurately, not Black (Logan 2003; Lee and Bean 2004).

The shift towards Other-Hispanic may reflect changes in the meaning of Hispanic in the population. For example, the changing racial distribution could reflect a schism amongst Hispanics. Here, Hispanic is seen as a distinct racial category for some, but not all, of the Hispanic population (Frank et al. 2010). By claiming Other, some Hispanics are rejecting a White, Black distinction, implying that Hispanic itself should be a racial option (Logan 2003; Brown, Hitlin and Elder 2007; Hitlin, Brown and Elder 2007). But it is substantively meaningful that only half of the Hispanic population identities as Other racially, while the other half does not. Or, it is now a meaningful distinction to claim White versus Other after identifying as Hispanic. We may then see a division between those that see Hispanic as a racial category and those that see Hispanic
as White (Dowling 2004). The identification division is likely to map onto different economic and residential outcomes (Penner and Saperstein 2008). For example, one may identify as Other racially if their occupational opportunities were distinct from that of the White majority.

Such processes would be reflected in the changing relational meaning, or social position, of Other-Hispanic (like Asian in Figure 6). We would expect, first, that Other-Hispanic is distinct from White/Black. Other-Hispanic should also grow relationally distant from the other non-black Hispanic categories. Thus, a relational split, or schism, in Hispanic reflects the larger divide in the population. Other-Hispanic will move to a social location away from the non-Black Hispanic categories. And more specifically, Other-Hispanic should shift to a social location previously occupied by a non-Hispanic category. Other-Hispanic represents a racial category on its own and should thus be relationally non-Hispanic.2

Alternatively, the increase in Other-Hispanic may signal the emergence of an overarching Hispanic category (Gomez 2007; O’Brien 2008). The racial question may no longer be meaningful for Hispanics, leading more people to claim Other racially. Individuals who identify as Hispanic do not have a clear racial answer as the answer is really Hispanic (Campbell and Rogalin 2006; Hitlin et al. 2007). For example, Campbell

---

2 It is even possible to see White-Hispanic act increasingly like Whites with Other-Hispanic representing the distinctly Hispanic pattern.
and Rogalin (2006) found that most Hispanics chose a Hispanic only identity when presented with the option. Thus any separate racial identification was less important than the Hispanic identification. In this world, answering White is just as meaningful as answering Other. None of the racial answers are appropriate but individuals are forced to claim a racial identity. Here, this conflict arises for all Hispanic individuals. People may react differently (White or Other) but the underlying problem is the same.\(^3\)

Similarly, the increase in the number of Hispanics (through immigration) increases the pool of possible Hispanic mates. This may lead to an increase in overall cohesion as second generation Hispanics marry newer immigrants (Licther, Carmalt and Qian 2011).

If this is true, then racial distinctions should dissipate, resulting in one large, cohesive Hispanic category distinct from both Black and White. Thus, we would see high rates of within Hispanic interaction and limited social divisions across racial categories (within Hispanic). More formally, all of the non-black Hispanic categories, including Other-Hispanic, should shift to the same social position between Time 1 and Time 2 (like White and White-Hispanic in Figure 6). The social boundaries would exist between Hispanic/non-Hispanic but not between Hispanic racial groups. A coherent Hispanic category would emerge if individuals are sorted into occupations and residential areas based on Hispanic origin, independent of racial identity. The sheer

---

\(^3\) Individuals may even be responding to how they think others see them, but this would not mean that racial identities are salient for the Hispanic population (Rodriguez 2000).
number of Hispanics would also support the claim of a distinct racial/ethnic group. This is also reinforced by a political environment where immigration issues serve to unify a heterogeneous population by mobilizing individuals on shared political issues.

Finally, the increase in Other-Hispanics could be a purely demographic phenomenon. The number of people falling into White-Hispanic and Other-Hispanic may have changed but the meaning of Hispanic is constant across time. Individuals claiming Other racially are no more occupationally or residentially disadvantaged than those claiming Other in the past. They are simply demographically more prevalent. In this case, the social positions occupied by the Hispanic categories would be unchanged and the Hispanic-racial categories would not collapse into a cohesive Hispanic group (like Black or Black-Hispanic in Figure 6). Thus, the Hispanic categories marry and cohabit in the same manner over time. Whether the Hispanic categories are distinct from Black/White remains an open question.

Overall, the paper asks how the meaning of Hispanic has changed in light of larger demographic shifts. Hispanic may have become socially divided, unified or experienced little change over time. In each case, there are specific expectations linking changes in social locations to the changes in racial identification. Demographic change is thus interpreted through the relational implications of the categories, rather than serve as a measure of categorical change in itself.
3.4 Data

The data come from 5% samples of the 1980-1990 Census (Public Use Microdata-Ruggles 2010). Each data set includes information on race and Hispanic identification for all household members. I begin with the 1980 survey as the Hispanic question was not asked in prior years of the Census, and the analysis requires a measure of Hispanic origin.¹ I focus on the change between 1980 and 1990 as this is the time period where the large shift in racial identification (amongst Hispanics) occurs. I thus focus on a short time period but a large demographic change. This restricts the data to the phenomena of interest, and pins the analysis to the empirical change I am trying to explain. The changes in racial identification between 1990 and 2000 are comparatively small. The 2000 data also complicates the comparison as individuals are allowed to mark multiple races. This means that the 1980/1990 data are not directly comparable to the 2000/2010 data. The 1980/1990 data are the only Census years with a consistent measurement of race where the Hispanic question was also asked. It is fortuitous that this time period also coincides with the large shift in racial identification amongst Hispanics. I exploit the consistency of the dataset and the empirical change in racial identification to explore the changing meaning of Hispanic.

¹ Individuals were asked a version of the Hispanic question in 1970 but the Census Bureau determined that the variable was not comparable to 1980-2010 question. The CPS also asked a combined race/Hispanic question prior to 1980 but only includes limited number of racial options. Additionally, the CPS is too small a sample for the analysis.
I restrict the analysis to couples between the ages of 20 and 39. This offers a comparison of similar age couples across time. I use younger couples as they represent the majority of relationships forming between surveys (Rosenfeld 2008). I define a relationship as marriage or cohabitation.  

Race/ethnicity is measured in the 1980 and 1990 Census as White, Black, Japanese, Chinese, Hawaiian, Filipino, Native American, Other, as well other smaller Asian categories. Individuals were also asked if they identified as Hispanic. I constructed a set of racial/ethnic categories from these two questions. The full set of categories includes the cross between race and Hispanic identification. There are categories for White, White-Hispanic, Black, Black-Hispanic, etc. Individuals who identify as White and Hispanic are labeled as White-Hispanic while those that identify as White but not Hispanic are labeled as White. Some racial categories, such as Japanese, had too few individuals to construct a separate Hispanic hyphenated category. Filipino is the only Asian ethnicity with sufficient Hispanic population to sustain a separate Hispanic category.

---

5 Note that the 1980 survey does not explicitly distinguish cohabiting couples from roommates. I treat individuals in 1980 as cohabiting couples if there are only two people in the household who are non-married, non-kin housemates. Thus households with a large number of roommates are not treated as a cohabiting couple.
**3.5 Methods**

**3.5.1 Review of Chapter 2**

I use the framework of Chapter 2 to relationally measure the racial/ethnic categories. The model draws on models of social space and social distance. From a social space tradition, the model characterizes the social world as a multi-dimensional demographic space (McPherson 1983). The probability of interaction is assumed to decrease as demographic distance increases. Homophily thus organizes social interactions (McPherson 2004). With continuous variables, such as age or education, the demographic variables themselves serve as the axes of the social space. As we move away from an age/education location, the probability of a tie forming with individuals in the focal location decreases. The model here differs from traditional social space approaches as the locations are estimated from observed interaction rates, such as marriage or friendship between categories (as opposed to the demographic characteristics themselves determining the metric of the space). The model thus estimates unknown social locations along known axes. Substantively, the categories are defined by interaction rates rather than the demographic values (e.g. years of education).

More formally, the approach has four steps: first, calculate the social locations using the Hoff et al. latentnet model (Hoff et al. 2002); second, collapse categories into the same location if they are relationally identical; third, rotate and scale the locations to
ensure comparability; and fourth determine which categories occupy the analogous locations across time. The latentnet model follows the form of:

\[
\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C - \|Z_i - Z_j\|
\]

where \(F_{ij}\) is the frequency count in cell \(ij\); \(\mu\) captures the overall mean, \(\mu_i^R\) and \(\mu_j^C\) are (fixed effects) factors for the row and columns, and \(Z\) is the unobserved latent position of \(i\) in a multidimensional Euclidean space (Krivitsky et al. 2009). As the distance between \(i\) and \(j\) increases, the frequency of interaction, relative to the marginals, decreases.

After calculating the locations, I collapse categories into a single location if they are relationally identical. Two categories are relational equivalents if there are no barriers to interaction (so within category ties are as likely as ties to the other category) and they have the same pattern of ties to all other categories. If both conditions are met, the categories are doing the same thing relationally and should be treated as one cohesive category. I use the Goodman model for the collapsibility of rows/columns in a log-linear model to test whether categories should be collapsed (Goodman 1981; Breiger 1990).

I start with the closest categories in the space, calculating the deviance from a Goodman model with those categories treated as one category. I then move to next closest pair and so on until are categories are in one large category. I take the set of collapsed categories with the lowest amount of information loss given the number of categories collapsed. See Chapter 1 for more details. Formally,

\[
\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \mu_{ij}^{RC}
\]

where \(\mu_{ij}^{RC} = \mu_p^{RC}\) for \((i, j) \in \text{subtable } p\), such that \(i\) and \(j\) are treated equivalently.

---

6 I start with the closest categories in the space, calculating the deviance from a Goodman model with those categories treated as one category. I then move to next closest pair and so on until are categories are in one large category. I take the set of collapsed categories with the lowest amount of information loss given the number of categories collapsed. See Chapter 1 for more details. Formally,

\[
\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \mu_{ij}^{RC}
\]

where \(\mu_{ij}^{RC} = \mu_p^{RC}\) for \((i, j) \in \text{subtable } p\), such that \(i\) and \(j\) are treated equivalently.
After collapsing relationally identical categories, I scale and rotate the locational scores to ensure comparability across time. The key is making sure that the dimensions correspond over time. If we are going to trace categorical change, it is necessary to know that a location at one time point is analogous to a location at the second time point. The first step is estimating the social locations in 1980. Here, I make no assumptions about the proper number of dimensions, allowing model fit to determine the best fitting space. I then estimate the 1990 social locations conditioned on the number of dimensions in the best fitting model from 1980. I then rotate the 1990 space so that the dimensions (and locations) are lined up, or are directly comparable, to the 1980 space. Formally, I use a Procrustean transformation to rotate and scale the locations.\(^7\) I rotate the 1990 space so that the same category occupies the extreme (right) location for each dimension, relative to the 1980 space. For example, if the second dimension in 1980 is defined by categories’ social distance to Black (so Black occupies the extreme location), then I rotate the 1990 space so that Black occupies the extreme location in the second dimension. This is analogous to giving each (important) racial factor in a regression a distinct dimension in social space.

Substantively, it is more difficult to use race as the dimensions in social space than interval variables like education or age. The axes in the interval case rests on the

---

\(^7\) The Procrustes transformation tries to minimize, through rotation and scaling, the (squared) distance between the estimated matrix and the reference matrix.
idea that demographic distance, and thus social distance, is increasing as we move away from any point along an axis. While the exact locations are estimated from interaction data, the basic ordering (or organizing principle) is known ahead of time. Demographic distance in the categorical case, here race, is less easily translated into the imagery of social space. Categories are demographically far or close based on a binary logic of matching/not matching (e.g. Black ≠ White), rather than a continuous measure of demographic distance (e.g. Grade 6 is farther from Grade 11 than Grade 10). A single axis may be insufficient to capture demographic distance.

As an alternative, we can use binary distinctions to define an entire space of racial demographic differences. Rather than one dimension per demographic characteristic (education, age), the space now has one dimension per salient categorical distinction, determined by the best fitting model in 1980. As we will see in the results section, the dimensions empirically correspond to Black, Native American, Other, non-Black Hispanic and Asian. Thus, just like in the continuous case, the axes are defined by the homophily principle: increasing demographic distance should lead to increasing social distance (Blau 1977; Smith, McPherson, Smith-Lovin 2013). I assume that the structure of social space is constant across this short time period, so that the same basic categorical distinctions hold between 1980 and 1990 (although the exact relations between
categories can change). The social locations thus still come from the observed pattern of relations, but we know what the axes correspond to ahead of time for the 1990 data. The locational values are also scaled to range between 0-1 along each dimension. I thus equate categories across time based on relative locations. A category in Time 2 is in the parallel location in Time 1 if it holds the same percentile position in the space.

3.5.2 Predicting Changes in Locations over Time

The key question is how the meaning of Hispanic has changed between 1980 and 1990, given the increase in Other-Hispanics. I thus focus on the Hispanic categories as they shift social locations across time. There are three parts to the analysis. First, I create a dataset pairing all i categories in 1980 with all j categories in 1990. I then define the dependent variable, $Y_{ij}$, describing the distance between category i in time 1 and category j in time 2. If $Y_{ij}=0$ then the two categories occupy the same location over time. For example, looking at Figure 6, Asian in Time 1 occupies the same location as Native American in Time 2 while Native American occupies a different location over time. Thus, $Y_{A,NA}= 0$ and $Y_{NA,NA}=.25$.

The second part of the analysis constructs the independent variable, $X_{ij}$. $X_{ij}$ is the expected distance between i and j under different hypotheses of Hispanic change. Each hypothesis starts with the baseline expectation that no change occurred. This means that

---

8 More generally, I assume that the structure of the space, or the large distinctions in society, change slower than the category-location mappings.
the Time 1 locations will match the Time 2 locations. I then adjust these expected Time 2 locations based on each particular hypothesis. The predicted distance between i and j is then the observed distance in Time 1 adjusted for the expected shifts in Hispanic.

For example, under the assumption that there are no changes in the meaning of Hispanic, the distance between category i in time 1 and j in time 2 will be the same as the distance between category i in time 1 and j in time 1 (as the categories have not shifted locations).

In contrast, if Hispanic is now an overarching, unified category, then non-Black Hispanic categories should all occupy a similar position in Time 2. The strongest version of this hypothesis would require the Hispanic categories to occupy the same position along all of the dimensions (thus relating to all other categories in the same way). A weaker version would only require the Hispanic categories to occupy the same locations along the non-Black Hispanic dimension (thus relating to White-Hispanic in the same way). If the Hispanic categories have all collapsed into one social location, then they will certainly occupy the same location along the Hispanic dimension. I restrict the analysis to the weaker hypothesis initially. I calculate the distance between i in Time 1

---

9 I restrict the analysis to non-Black Hispanic categories as they cluster together fairly strongly along one dimension in 1980. Including Black-Hispanic would simply make it a harder test for the cohesive category hypothesis (as Black-Hispanic starts so far away).
and \( j \) in Time 2 based on the Time 1 locations (as a baseline expectation) but now the Hispanic categories all occupy the same “edge” position in the Hispanic dimension.\(^{10}\)

Finally, if there is a schism in the meaning of Hispanic, then Other-Hispanic should occupy a position in Time 2 that is distinctly non-Hispanic. I thus calculate the distance between category \( i \) in Time 1 and category \( j \) in Time 2 using the Time 1 locations as the baseline, but now assume that Other-Hispanic has shifted away from the edge of the Hispanic dimension. Specifically, I calculate the distance between category \( i \) and \( j \) assuming that Other-Hispanic has shifted to the mean of the Hispanic dimension (holding all other Time 1 locations the same).

I define \( X_{ij} \) as the predicted distance between \( i \) in Time 1 and \( j \) in Time 2 based on these different coding schemes. I then use the predicted distances, \( X_{ij} \), to predict the true distances between \( i \) and \( j \). \( Y_{ij} \) is modeled using a simple OLS (although I fit the model using the maximum likelihood solution). I restrict the analysis to changes happening along the Hispanic dimension as this is the only place where the models differ.

### 3.6 Results

#### 3.6.1 Racial Organization

Changes in the relational meaning of race/ethnicity can only be understood within the larger structure of race in America. I thus describe empirically what the racial

---

\(^{10}\) If the categories all collapse into the same location, they are most likely to collapse into the extreme position, distinct from all other categories, along the Hispanic dimension.
social space looks like before considering the changing positions, and thus relational meaning, of the Hispanic categories.

The social divide between Blacks and Whites defined the racial order for much of the 20th century (Bobo 1997; Brown et al. 2003; Lee and Bean 2004). This organizing principal may, however, no longer represent a population with an increasing number of Asians and Hispanics. I define social locations based on the patterns of interaction between all racial/ethnic categories. It is then possible to ask if a Black/White divide can properly characterize the social system.

The first question is whether Hispanic (or Asian) can be easily incorporated into the larger White category. There is thus a non-Black/Black divide (Lee and Bean 2007). If Hispanic is not collapsed into White, we can then ask if the Hispanic categories are easily defined by their relative social distance to Whites and Blacks: mixing freely with Whites means not mixing freely with Blacks. This suggests that the system can be characterized in a linear fashion along one dimension, where being “far” from White means being close to Black. If the Black/White divide completely defines the racial space, then categories equally close to White would also have high rates of contact with each other. In that sense, individuals interact if they are they are equally White. Not only is there a hierarchy, the hierarchy characterizes the full pattern of relations between racial groups, such as in Figure 6 (Fu 2001).
Alternatively, the racial order may not be based solely on the social distance to White or Black. There could still be a hierarchy, defined by how close the categories are to White, but the Black/White distance does not completely characterize the organization of race in America. Here, Hispanics and Asians may be socially close to Whites but socially far from each other. Or, categories that are relationally equivalent in terms of “Whiteness” do not necessarily have high rates of intermarriage. There is then a difficulty in placing these other racial/ethnic categories on a linear continuum based on Black/White distinctions.

Table 1 presents the social distance model results for 1980. I make no a priori assumptions about the number and nature of the dimensions when fitting these initial models. Each row in the table presents the fit statistics for a model with different dimensions. The models grow increasingly complex as we move down the rows, going from 1 dimension all the way up to 7 dimensions. I thus allow for the possibility that anywhere between 1 and 7 dimensions is sufficient to the fit the data. The results clearly show that the best fitting model, relative to the number of parameters, is the model with 5 dimensions.

---

11 The 1 dimensional results come from a log-linear distance model rather than the latentnet model of Hoff et al. (2002). The latentnet model assumes a Euclidean space and more than 1 dimension. I thus calculate the 1 dimensional solution using an alternative model, that makes no restrictions on the number of dimensions (Goodman 1979). Formally, the model can be written as: 
\[
\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta \varphi_i \phi_j ,
\]
where \(F_{ij}\) is the frequency count in cell \(ij\); \(\mu\) captures the overall mean, \(\mu_i^R\) and \(\mu_j^C\) are the row and column means; \(\beta\) is a coefficient and \(\varphi_i\) and \(\phi_j\) are the unknown, estimated location scores (which can be constrained to be equal).
Table 1: Model Fit for Latent Space Solutions by Dimension

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>dfs</th>
<th>Deviance</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>137</td>
<td>182930.392</td>
<td>181035.370</td>
</tr>
<tr>
<td>2</td>
<td>123</td>
<td>27349.384</td>
<td>25648.014</td>
</tr>
<tr>
<td>3</td>
<td>109</td>
<td>17626.526</td>
<td>16118.808</td>
</tr>
<tr>
<td>4</td>
<td>95</td>
<td>15830.644</td>
<td>14516.578</td>
</tr>
<tr>
<td>5</td>
<td>81</td>
<td>14625.604</td>
<td>13505.188</td>
</tr>
<tr>
<td>6</td>
<td>67</td>
<td>14618.111</td>
<td>13691.348</td>
</tr>
<tr>
<td>7</td>
<td>53</td>
<td>14613.929</td>
<td>13880.818</td>
</tr>
</tbody>
</table>

Figure 7 describes the 5 dimensional solution. The dimensions correspond to 5 typically used racial/ethnic categories—Black, Hispanic, Asian, Other, Native American. For ease of viewing, I present the space in 4 different subplots. Each subplot has Asian on the y-axis and a different racial/ethnic dimension on the x-axis. For example, the top plot is structured by Asian on the y-axis and non-Black Hispanics on the x-axis (Other-Hispanic, White-Hispanic, Native American-Hispanic and Filipino-Hispanic). Thus, categories are placed on the y-axis by their social distance to the Asian categories and placed on the x-axis by their distance to Hispanics. The second plot reflects the same basic pattern relative to Other, the third plot relative to Black, and the fourth relative to Native American.
It is clear from the figure that the Hispanic categories are not easily collapsed into a larger White category. The Hispanic categories do not uniformly cluster around the locations occupied by White. The racial relational system is not easily reduced to a simple Black/non-Black divide. Similarly, a simple Black-White continuum does not adequately capture the structure in the system. The 1 dimensional model offers the worst fit of any model by far. This does not mean that no Black/White hierarchy exists, but it does mean that Hispanic and Asian categories are not socially close just because they are both closer to Whites than Blacks. Racial/ethnic categories can be more or less Asian, more or less Black, etc., and the occupied location in this multidimensional space defines a racial/ethnic category quite well.
The results also clearly show that Hispanic categories are not collapsed into a single social location. The Hispanic categories may occupy similar locations along one dimension (e.g. along the Hispanic dimension), but the overall locations are distinct, making it difficult to treat the Hispanic categories as relationally identical. More formally, collapsing different racial/Hispanic categories leads to a large loss of information (based on the Goodman test results). The Hispanic categories are thus not only distinct from White/Black, they also exhibit distinct patterns from each other. For example, Native American-Hispanics are socially closer to non-Hispanic Native Americans than White-Hispanics are. This basic result is also reflected in the 1990 data. Hispanic origin does not uniformly supersede racial identity and this is true in both 1980 and 1990. In sum, the racial space is not constituted by a simple Black/White hierarchy, and the Hispanic categories are not easily treated as an overarching category.

3.6.2 Hispanic Change

The Hispanic categories are not aggregated between 1980 and 1990 but they do experience important shifts in social positions over that time period. Figure 8 plots the Hispanic demographic dimension for 1980 and 1990. In 1980, all of the non-Black Hispanic categories occupied positions close to the right edge of the dimension. There are three large clusters. Other-Hispanic, White-Hispanic, Filipino-Hispanic and Native American-Hispanic make up the Hispanic cluster. Categories like White, Japanese and Native American make up the cluster closest to the Hispanic categories.
Figure 8: Racial Locations along Hispanic Dimension: 1980-1990

By 1990, Other-Hispanic has shifted away, relationally, from the other non-Black Hispanic categories. White, Filipino and Native American Hispanics have, in contrast, remained tightly clustered on the boundary of that dimension. Other-Hispanic is thus no longer part of the non-Black Hispanic cluster. I have drawn out the paths of each Hispanic category between 1980 and 1990. It is clear that Other-Hispanic has shifted notably to the middle of the dimension. Thus, rather than occupy positions like White-Hispanic or Native American-Hispanic, Other-Hispanic now occupies locations like Japanese or Native American. It is firmly in the non-Hispanic cluster and is, relationally, not Hispanic. This division is brought into sharper focus by comparing the paths of Other and White Hispanic. While Other-Hispanic occupies a more median position over time, White-Hispanic moves closer to the other non-Black Hispanic categories. The split in Hispanic is thus encapsulated by the inverted paths of White and Other Hispanic,
leading to a more cohesive Hispanic category, but one without Other-Hispanic as a member.

Table 2 presents formal models of Hispanic change. The models predict the distance between categories in 1980 and 1990. The models are restricted to locational changes along the Hispanic dimension. Each model reflects a different hypothesis about the meaning of Hispanic (given the increasing number of people identifying as Other-Hispanic). Model 1 corresponds to no change in the Hispanic categories. Model 2 predicts movement towards a unified Hispanic category, or movement towards the edge of the Hispanic dimension. Model 3 predicts movement towards the mean of the Hispanic dimension for Other-Hispanic.

**Table 2: Models Predicting Hispanic Change along Hispanic Dimension**

<table>
<thead>
<tr>
<th></th>
<th>Demographic Change Only</th>
<th>Cohesive Hispanic Category</th>
<th>Hispanic Schism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Likelihood</td>
<td>132.114</td>
<td>121.647</td>
<td>139.779</td>
</tr>
<tr>
<td>AIC</td>
<td>-258.228</td>
<td>-237.294</td>
<td>-273.557</td>
</tr>
<tr>
<td>BIC</td>
<td>-248.394</td>
<td>-227.46</td>
<td>-263.723</td>
</tr>
<tr>
<td>(based on number of categories)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>-220.697</td>
<td>-199.763</td>
<td>-236.026</td>
</tr>
<tr>
<td>(based on Sample Size)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Categories²</td>
<td>196</td>
<td>196</td>
<td>196</td>
</tr>
<tr>
<td>Sample Size</td>
<td>2003496</td>
<td>2003496</td>
<td>2003496</td>
</tr>
</tbody>
</table>
Model 3 clearly offers the best fit for Hispanic change (based on BIC). Other-Hispanic has moved closer to the mean while the Hispanic categories have not shifted together nor remained fixed over time. The schism model of Hispanic change thus fits best, where Other-Hispanic increasingly represents a distinct racial/ethnic category from the non-Black Hispanic categories. Other-Hispanic is thus becoming less like non-black Hispanics, and more like racial categories with no Hispanic origin. It also means that demographic distance maps less strongly onto social positions. Categories should be close when demographic distance is small (on that dimension), but this is only loosely the case for Other-Hispanic and the other Hispanic categories.

Models 1-3 offer only part of the empirical story, focusing solely on the Hispanic dimension. But it is also important to consider the category-location changes over the entire space. Figure 9 plots the distance, $Y_{ij}$, between categories in 1980 and 1990, using locations in all five dimensions to calculate the distance matrix. The darker the color, the farther away a row category in 1980 is from the column category in 1990. The question is which category in 1990 is closest to which category in 1980. To highlight this, I take the calculated distances for each row and subtract its minimum distance. The white block in each row thus represents the 1990 category (on the column) that is closest to the 1980 category (on the row).

Here, the story shifts from one of change to one of stability. In Figure 9, the lightest block, and thus the lowest distance, falls along the diagonal. Thus, the best
predictor of which category occupies the analogous location across time is the category that occupied that location in Time 1. And more specifically, the location of Other-Hispanic in 1990 is best predicted by the location it occupied in 1980, despite the changes happening in the Hispanic dimension.

Figure 9: Distance between Categories in 1980 and 1990

Thus, the overall relational meaning of Other-Hispanic is similar across time, although there are clear changes to internal structure of Hispanic. As we saw in Figure 8,
the increase in people identifying as Other-Hispanic is reflected in increasing social
differentiation amongst Hispanics. We see Other-Hispanic emerge as a racial/ethnic
category on its own, where it occupies a distinct position in the Hispanic dimension.
Such shifts in meaning are limited, or localized, however, as the other racial groups
relate to Other-Hispanic in much the same manner as before. The distinctions made by
those identifying as Hispanic are thus not matched by the other racial groups. The
increase in Other-Hispanic may reflect a division of Hispanic into those who see
themselves as racially Hispanic and those that do not—but such increasing divides
appear to be restricted to Hispanics themselves.

3.7 Conclusion

This paper examines the changing nature of race in America. I explore racial
change from a constructionist viewpoint, where the racial categories are seen as fluid,
socially constructed and (possibly) contentious (Wimmer 2008). At its crux, the paper
connects large scale demographic change to the relational meaning of the racial/ethnic
categories. I focus on a particularly salient change occurring over the last 40 years: the
increasing number of Hispanics identifying as Other racially in the population.
Formally, I apply a social distance/social space framework to the problem of categorical
meaning (McPherson 1983; Hoff et al. 2002; Smith, McPherson and Smith-Lovin 2013). I
trace racial categories through time as they occupy different social locations, or have
different patterns of interaction. I then interpret the macro level trends in light of the changing relational meaning of the categories.

Using Census data from 1980-1990, I find that the racial system cannot be described by a simple Black-White continuum. Typically used categories (Black, Asian, Native American) serve as dimensions in a racial space, as opposed to categories along a single dimension. Over time, I find that the increase in Other-Hispanic corresponds to two conflicting results. First, Other-Hispanic has shifted, relationally, away from non-Black Hispanics. This suggests a kind of schism in the meaning of Hispanic, between those that see Hispanic as another race and those that do not (Frank et al. 2010). The demographic shifts suggest that Other-Hispanic is emerging as racial category on its own. It now occupies a non-Hispanic social location, acting more like White or Native American than the other Hispanic categories along the Hispanic dimension. In contrast, the relational position of Other-Hispanic is steady over the entire space (save for the Hispanic demographic dimension). This means that Other-Hispanic is relating to most racial categories in much the same way as before, despite the number of people now identifying as Other-Hispanic. The differentiation of Hispanic would appear to be internal to Hispanic itself.

What do the results suggest for the future of racial stratification in the US? First, the differentiation of Other-Hispanic and non-Black Hispanic may indicate a long term trend for non-Black Hispanics. Non-Black Hispanics may be incorporated into the larger
White category, just like European immigrants in the early 20th century. Other-Hispanic would then become the umbrella category for Hispanics who do not fall neatly into the White category (Gomez 2007; O’Brien 2008). In that sense, Hispanic would become a new, distinct category from the Black-White dichotomy, but (some) Hispanics would still be incorporated into the larger White category (Lee and Bean 2004; 2007). Thus the Hispanic experience would both mimic and diverge from that of past immigrant groups (Frank et al. 2010).

More generally, the paper offers a template for studying racial/ethnic systems. The paper reorients typical regression models, placing the constructed nature of racial categories at the center of the analysis (Martin and Yeung 2003). Categorical fluidity is not a convenient metaphor but an empirically tractable phenomenon which can and should be measured (Brubaker 2004). The key is not only measuring the categories through time, but applying this radical sociological idea (categories are not fixed) to important demographic/political questions. The model thus combines aspects of demography, network analysis and racial theory into a uniquely sociological framework—one where demographic change and demographic meaning are essential pieces in solving the puzzle of macro level change.
4. The Relational Role of Mixed Race Categories

The 2000 Census marked a radical change in the measurement of race in America (Hirschman, Alba and Farley 2000; Farley 2002). Individuals for the first time were allowed to mark multiple racial categories. One could conceivably identify as White and Black rather than having to choose between the two. The change in measurement is an important one, as multiracial individuals represent a growing portion of the population, possibly as large as 21 percent by 2050 (Smith and Edmonston 1997; Farley 2002). Multirace identities are particularly common amongst Asians and Hispanics, demographic groups that have growing quickly over time (Lee and Bean 2004). Given these population dynamics, it is important to understand the role of mixed race categories in the racial order. More theoretically, the introduction of new categories offers a unique opportunity to rethink race/ethnicity from a new light (see Baily, Loveman and Muniz 2013 for similar logic).

This paper uses a relational framework to characterize the multirace categories. Past work has generally defined racial/ethnic minorities, including mixed race individuals, by their potential assimilation into society (Alba 1990; Portes and Rumbaut 2001; Lee and Bean 2007). Lower in-group bias and higher contact with the White majority, along friendship or marriage lines, is used as evidence of integration (Qian 2002; Lee and Bean 2004; Berry 2006). This paper diverges from such work as I focus on the position of mixed race categories in the larger system of racial relations. Mixed race
categories are defined by their social distance to all other racial/ethnic categories (Levine 1990; Rytina 1992). The question of incorporation shifts from assimilated/not-assimilated to something more general: how do mixed race categories fit into the existing racial order? Do they relate predictably to other racial/ethnic categories or do they represent something new altogether? And more generally, is interracial marriage a revolutionary force, creating new categories, with new patterns of social limitations and distinctions?

Following Chapter 3, I use interaction patterns as a way of characterizing multirace categories. I assume that a racial label is not an essential, fixed characteristic of an individual (Barth 1969; Anthias 2001; Zuberi and Bonilla-Silva 2008). A category only has meaning through the observed distinctions mapped on to group membership. Racial/ethnic categories can thus be defined, not by their label, but their behavioral implications—and more specifically by whom individuals are allowed to marry, make friends with and share the same social world (Breiger 1981; Burt 1991). I draw on a social space tradition to formalize this idea (Bottero and Prandy 2003; McPherson 2004). Categories are placed into social locations based on observed rates of interaction. Categories are socially far if interaction is unlikely and socially close if they have high rates of marriage/cohabitation (Beshers and Laumann 1967; Laumann 1969). Marriage or cohabitation is an intimate social tie, meaning demographic groups that do not intermarry or cohabitate have strong barriers to interaction; for example, along economic, residential or cultural lines (Torche 2010). The mapping of categories into
social locations thus tells us a great deal about the ordering of the racial system and the boundaries being drawn in society.

Given the observed distances between racial/ethnic categories, I ask how the mixed race categories fit into the racial landscape. I define a mixed race category, e.g. White-Black, by the social distance to its parent categories, White and Black. A mixed race category should, in general, be socially closer to its parent categories than to other racial/ethnic categories. The relational pattern exhibited by the White-Black category is more likely to resemble the relational patterns of White or Black than Chinese. I measure the mixed race categories by their adherence to this baseline expectation. Formally, I develop a classification system based on three social locations: the location of the multirace category (White-Black) and the locations of its parent categories (White, Black).

A mixed race category that is socially close to one parent category falls predictably into the racial space and our view of race is largely unchanged by the change in measurement. For example, there may be high rates of intermarriage and the same patterns of ties between White-Blacks and Whites. Here White-Black is essentially White. This would follow a classic assimilation pattern, where the new multirace categories are incorporated into an existing racial/ethnic category (Alba 1990; Igantiev 1995; Brodkin 1998; Roediger 2005).
A mixed race category could also reflect the relational patterning of both parent categories. White-Black is thus White and Black, and is socially closer to both parent categories than they are to each other. A mixed race category relationally between the parent categories also fits comfortably into the racial system as it represents a combination of existing racial/ethnic categories. A mixed race category of this type offers a unique pattern of incorporation, however, as it does not fall into an existing racial/ethnic category, as in the classic assimilation case.

A category that is socially far from both parent categories, in contrast, represents a distinct racial/ethnic category. White-Black may be simultaneously distant from White and Black, and farther from both than they are to each other. A category that is relationally distinct from both parent categories should reflect a unique set of economic, residential and cultural constraints. White-Black is thus neither White nor Black, and exhibits a relational logic not reflected in the parent categories. In this case, we can ask which non-parent categories the mixed race category is socially close to. If White-Black is not White and not Black, does it relate to other groups like Native American? Or is it not well defined by any existing racial/ethnic category?

In short, a mixed race category may reflect one parent category (White or Black), both parent categories (White and Black) or neither parent categories, and these aggregate differences will be reflected in the occupied social locations of the mixed race and parent categories. I begin the chapter with a quick review of how to measure
categories through interaction patterns. I then discuss the formal conditions used to characteri
Empirically, I use marriage/cohabitation data from the 2000 Census to describe mixed race categories like White-Black and White-
Empirically, I use marriage/cohabitation data from the 2000 Census to describe mixed race categories like White-Black and White-
Other. I ask which mixed race categories are well defined by the parent categories and which are not. I also ask which racial/ethnic category each mixed race category most resembles. I end the chapter by discussing the differences between Hispanic and non-
Hispanic mixed race categories.

4.1 Measuring Categories through Interaction Patterns

The classification of the mixed race categories stands on two ideas. First, following a long literature, I argue that racial/ethnic categories are not fixed, essential characteristics of an individual (Barth 1969; Brubaker 2004; Wimmer 2008). Rather, race/ethnicity is an important social variable because of the economic, political, and social conditions correlated with different racial/ethnic categories. These forces increase or decrease the social distance between racial/ethnic groups, but skin color or ethnicity does not in itself causes good or bad outcomes. The racial/ethnic categories can only be defined relative to each other, and do not exist as essential constructs with their own independent definition (White 1966; Barth 1969).

And second, I suggest that interaction patterns, such as marriage or friendship rates, offer an ideal metric to define the racial/ethnic categories. The meaning of race/ethnicity is then a “relational meaning”, where categories are defined relative to
one another based on behavioral indicators. Interaction patterns offer an ideal measure as they concretely capture social distinctions in a population (Haller 1981; Bottero and Prandy 2003). It is clear that who you spend your time with is important to people. It is then notable when there are few observed social connections between demographic groups. Interaction patterns thus reflect the observed distinctions in a population.

Homogamy scholars, for example, have used marriage patterns to measure social distinctions in a population (Mare 1991; Blackwell and Licther 2004; Rosenfeld 2008; Qian and Lichter 2011). Here, marriage and cohabitation are used to measure the social distance between racial/ethnic groups. Social distance will be high if there are strong economic, residential or cultural barriers to forming close relationships (Laumann 1965; Chan and Goldthorpe 2007; Torche 2010). Social distance is then an aggregate measure of distinction, reflecting the various forces differentiating groups of people in society. A typical study will examine the rate of interracial marriage between a priori categories. For example, we may ask how frequent Black-White marriage ties are relative to chance (Kalmijn 1993; Qian and Lichter 2007). This approach is, however, insufficient for the purposes of this paper. The goal here is to define the racial/ethnic categories by the rates of intermarriage (White 1966; Burt 1991; Bottero and Prandy 2003). For example, if Hispanics were allowed to marry like Whites then Hispanics would be White in all ways but name—and the labels cannot be used, on their own, to differentiate the categories.
The key to extending a traditional approach is to think of the demographic categories as occupying locations in social space (Bottero and Prandy 2003; McPherson 1983; 2004). Formally, I place the racial/ethnic categories into social locations based on observed rates of interaction (e.g. frequency of marriage between racial categories). Social locations reflect what interactions are possible for individuals with a given racial/ethnic profile. Two categories are close in social space if interaction is likely and will be far if interaction is unlikely (Laumann 1969; DiPrete et al. 2011). In a similar manner, two categories with very different rates of contact to other categories will be far apart in the space (as they have different distances to the same third party category). This makes it possible to simultaneously define the racial/ethnic categories by their distance to all other racial/ethnic categories (Levine 1990; Rytina 1992). Categories with the same pattern of interaction will occupy the same social location, and will thus be defined as the same.

4.2 The Relational Meaning of Mixed Race Categories

The manner in which the mixed race categories fit into the racial system reveals a great deal about race in America (Gans 1999; Lee and Bean 2007). A mixed race category that is incorporated into an overarching White category is very different from one that is relationally distinct from existing racial/ethnic categories. Which mixed race categories are incorporated and which are not then becomes an important way to understand the organization of race.
Formally, I use inferred social locations to describe the mixed race categories. I define a mixed category by its social location and the social locations of the parent categories. A category fits easily into the system if it “acts” like its parent categories. For example, if White and Black are both socially far from Native American, then we would expect White-Black to also be far from Native American. White-Black then breaks with the rules of the parent categories if it is relatively close to Native American and far from either White or Black. The baseline assumption is that a mixed race category will resemble its constituent parts, but it would be telling if this was not the case.

Substantively, the mixed race category will follow the relational lead of the parent categories if the economic, residential and cultural experience of mixed race individuals parallels one or both of the parent categories.

Figure 10 presents 4 forms a mixed race category could take. In each panel, the mixed race category occupies a different social location relative to the parent categories. Each panel presents a simplified example racial space in two dimensions. I have also limited the plots to 4 racial/ethnic categories: White-Black (the mixed race category of interest); White (parent category); Black (parent category); and Native American. For example, in Panel 1, White has the fewest social ties to Black. Only the location of White-Black changes across the panels. The 4 panels represent ideal types rather than empirical results. They will be used as a guide in the results section when interpreting the mixed race categories.
Panel 1 presents a world where White-Black falls socially very close to White. White-Black individuals thus have a high rate of interaction with Whites and the same pattern of ties to other racial/ethnic categories. In this sense, White-Black might as well be White (Burt 1991). Substantively, people may identify as White-Black but act like individuals in a parent racial category (i.e. White). White-Black is thus not a “mixed” racial category at all—as individuals who identify as White-Black have the same
limitations in forming relationships as Whites. They are also likely to have the same kinds of aggregate occupational and residential outcomes. More generally, the location of the mixed race category suggests that a multiracial identity is optional or symbolic (Xie and Goyette 1997; Lee and Bean 2004; Lee and Bean 2007). Individuals may mark two races on official forms but their lives are not affected by the secondary identity—they are purely assimilated into an existing racial/ethnic category. There is then a detachment between identification and concrete social outcomes and this is reflected in the relational collapsing into a larger racial/ethnic category (Lee and Bean 2007).

The second question for a setting like Panel 1 is which parent category the mixed race category mimics. Are they always placed farthest, socially, from White? So, for example, White-Black would have the relational implications of being Black and not White. Or are the mixed race categories White, thus reflecting the (sometimes) expansive role of whiteness (Alba 1985; Gerstle 1999; Brodkin 2002; Roediger 2005)?

In Panel 1, White-Black individuals follow the relational lead of Whites. In contrast, in Panel 2, White-Black individuals follow the relational lead of Whites and Blacks. Individuals could conceivably identify as either racial group, if no multirace option existed (Xie and Goyette 1997; Harris and Sim 2002). Here, White-Black is equally close to both parent categories. Additionally, White-Black is closer to White (and Black) than White is to Black. White-Black is thus relationally bounded, or between the parent categories, being closer to either parent category than they are to each other. This
means that White-Black individuals have a high rate of intermarriage/cohabitation with Whites and Blacks, even though there are few ties between Whites and Blacks. Similarly, the relational patterning (relative to other categories) is neither purely White nor Black. White-Black is socially closer to Native American than expected from the social ties of Blacks, but much farther than expected from the social ties of Whites. The mixed race category is then bounded by the parent categories but acts in a distinct manner from both: as it represents a particular combination of the two. White-Black is thus not obviously White or Black but is rather White and Black. It is a mixed race category in the truest sense: incorporated into the racial order but not assimilated into an existing racial/ethnic category.

In Panel 3, White-Black individuals are defined by their stark social distance to Blacks. Like Panel 1, White-Black is socially closer to White than to Black. It is, however, much farther from both parent categories than in Panel 1. Importantly, White-Black is socially farther from Black than White is. Substantively, this means that there are fewer social ties between White-Black individuals and Blacks than Whites and Blacks. Thus, the parent categories are closer together than the mixed race category is to one of them. The mixed race category is defined by this social distance to the parent category: being, in this example, distinctly not-Black. This means that the economic, status, cultural, etc. standing of White-Black individuals is likely to differ starkly from that of Blacks (in this example). Or, children of mixed race couples may systematically distance themselves
from one of the parent categories (e.g. wanting to be White) (Saenz et al. 1995; Qian 2004; Qian and Lichter 2007).

Panel 4 is a combination of Panels 2 and 3: White-Black is equally close to White and Black but is much farther to both of them than in Panel 2. In this case, White-Black is quite far from both parent categories. This means that White-Black individuals have few ties to Whites and few ties to Blacks. It also means that White-Black individuals have a distinct pattern of ties to other racial categories. White-Black is then unbounded, or not well defined, by the parent categories. In this sense, White-Black might as well be called “orange” as opposed to White, Black or White-Black. The occupational, residential and cultural standing of White-Black individuals is likely to differ starkly from that of both Whites and Blacks.

White-Black is not well defined by the parent categories in Panels 3 and 4. White-Black is unbounded (unlike in Panel 2) and does not fall particularly close to White or Black (unlike in Panel 1). For such mixed race categories, the second question is which non-parent category the mixed race category most resembles. Who do White-Black individuals follow relationally, if not White or Black? In Panel 3, it is clear that White-Black is the relational equivalent of Native American. White-Black is the same social distance to White and Black as Native American is to those categories. White-Black is thus not well defined by the parent categories but does fit into the racial order—as it is the social equivalent of an existing racial/ethnic category. In Panel 4, White-Black
occupies a social location distinct from all other racial/ethnic categories. White-Black thus represents something truly different, a category that cannot be derived from the previous racial system. Here, there are clear barriers between the mixed race category and incorporation into existing racial/ethnic categories.

In summary, mixed race categories can be more or less well defined by the parent categories and more or less well incorporated into the existing racial order. The question is which mixed race categories exhibit which type of relational patterning and why.

4.3 Data

The data come from a 5% sample of the 2000 Census (Ruggles 2010). The analysis is restricted to 20-39 year old cohabiting couples. I use younger couples as they represent the majority of relationships forming since the previous census (Rosenfeld 2008). The 2000 Census offers respondents a large number of options when identifying racially. They are allowed to claim more than one racial identity, and there is a separate question concerning Hispanic identification. By crossing the Hispanic and racial answers, the data leave 74 distinct racial/Hispanic combinations.¹ For example, there are traditional categories like White, Black, and Native American. There are also Hispanic versions of these categories, like White-Hispanic, Black-Hispanic, as well mixed racial

¹ This only includes respondents who identified as one or two races.
categories such White-Black and White-Other. Individuals are placed in a unique, non-overlapping racial/ethnic category.

In presenting the mixed race results, I focus on a subset of all possible mixed race categories. I limit the analysis to mixed race categories that involve White and a (numerically) large minority racial group. This serves two purposes. First, it limits the analysis to a set of categories with a common denominator (White). Second, it limits the analysis to mixed race categories large enough to properly analyze on their own. To narrow the analysis further, I only consider mixed race categories with an analogous Hispanic category (so that individuals of that mixed race category sometimes claim a Hispanic identity). This makes it possible to compare Hispanic and non-Hispanic based categories. The mixed race categories of interest include White-Black, White-Other, White-Native American; White-Black-Hispanic, White-Other-Hispanic, and White-Native American-Hispanic.

4.4 Methods

The mixed race categories are characterized through a new, formal classification system. Before classifying the mixed race categories, it is necessary to place all of the racial/ethnic categories into social locations, like in Figure 10. This process follows the methods in Chapter 2 and has 2 basic steps: first, I calculate the social locations using the Hoff et al. latentnet model (2002); and second, I collapse categories into the same social location if they are relationally identical.
The latentnet model follows the form of:

\[ \log(F_{ij}) = \mu + \mu^R_i + \mu^C_j - \|Z_i - Z_j\| \]

where \( F_{ij} \) is the frequency count in cell \( ij \); \( \mu \) captures the overall mean, \( \mu^R_i \) and \( \mu^C_j \) are (fixed effects) factors for the row and columns, and \( Z_i \) is the unobserved latent position of \( i \) in a multidimensional Euclidean space (Krivitsky et al. 2009). As the distance between \( i \) and \( j \) increases, the frequency of interaction, relative to the marginals, decreases.

After calculating the initial social locations, I use Goodman’s model for row/column collapsibility to test if any of the racial categories, mixed or not, can be collapsed into the same location with little loss of information (Goodman 1981, Breiger 1990). Two categories are collapsed if two conditions hold: first, there are no barriers to interaction, so that within category ties are just as likely as ties to the other category; and second, the categories have the same pattern of ties to all other categories. If both conditions are met, the categories are relationally identical and should be treated as one aggregate category. Within this larger process, I ask if the mixed race categories are

---

2 Many of these categories have relatively few respondents, too few to properly estimate the social locations. Thus, prior to the analysis, I collapse categories with very small cells into larger categories. I collapse these small categories by running a series of Goodman models (see Goodman 1981; Breiger and Mohr 2004). I collapse the smaller categories into the larger category that results in the lowest amount of information loss.

3 I start with the closest categories in the space, calculating the deviance from a Goodman model with those categories treated as one category. I then move to next closest pair and so on until are categories are in one large category. I take the set of collapsed categories with the lowest amount of information loss given the number of categories collapsed. See Chapter 1 for more details. Formally,

\[ \log(F_{ij}) = \mu + \mu^R_i + \mu^C_j + \mu^{RC}_{ij} \]

where \( \mu^{RC}_{ij} = \mu^p \) for \( (i, j) \in \text{subtable} p, \text{such that } i \text{ and } j \text{ are treated equivalently} \)
collapsed into a parent category. For example, we can ask if White-Black is sufficiently similar to White or Black to be collapsed into one of those categories.

If the mixed race categories are not collapsed into larger aggregates, I ask if they are bounded, or well defined, by the parent categories. I ask how close the mixed race category is from the parent category locations. I also ask how far the parent categories are from each other. I use this pattern of distances to formally characterize the mixed race categories.

Each mixed race category is described by the four ideal types in Figure 10. A mixed race category could, like in Panel 1, be close to one parent category. It will be socially close to one parent category and far from the other. Specifically, the distance between the mixed race category and the “far” parent category will be equal to the distance between the two parent categories. The distances scores for Panel 1 would look something like: \( \|WB - W\| = .1; \|WB - B\| = 1.9; \|W - B\| = 2.0 \). Mixed race categories will look like Panel 2, or be the intersection of the parent categories, if they are closer to each parent category than the parent categories are to each other. The distances will have the following kind of pattern: \( \|WB - W\| = .75; \|WB - B\| = .75; \|W - B\| = 2.0 \).

Panel 3 is defined by the stark distance to one of the parent categories. The mixed race category is farther from one parent category than the parent categories are to each other. For example: \( \|WB - W\| = .5; \|WB - B\| = 2.5; \|W - B\| = 2.0 \). A mixed race category like Panel 4 will have distances to the parent categories that approach (or are larger
than) the distance between the parent categories. For example: \( \| WB - W \| = 2.5; \| WB - B \| = 2.5; \| W - B \| = 2.0 \). The key is not the specific distances in each case, but the patterning that places the mixed race category into one of the four types. It is also possible that a category will fall somewhere between these ideal types.

Formally:

Define \( d_{mp1} \) as \( \| Z_m - Z_{p1} \| \) where \( m \) = the mixed category of interest and \( p1 \) = the first parent category.

Define \( d_{mp2} \) as \( \| Z_m - Z_{p2} \| \) where \( m \) = the mixed category of interest and \( p2 \) = the second parent category.

Define \( d_{p1p2} \) as \( \| Z_{p1} - Z_{p2} \| \) where \( p1 \) = the first parent category and \( p2 \) = the second parent category.

Assume that \( d_{mp1} \leq d_{mp2} \)

\[ \epsilon = .01 * d_{p1p2} \]

**Type 1: Very Close to1 Parent Category**

**Condition 1:** \( d_{p1p2} - \epsilon \leq d_{mp2} \leq d_{p1p2} + \epsilon \)

**Condition 2:** \( \epsilon \geq d_{mp1} \)

**Type 2: Close to Both Parent Categories**

**Condition 1:** \( \epsilon \leq d_{mp1} \leq d_{p1p2} - \epsilon \)

**Condition 2:** \( \epsilon \leq d_{mp2} \leq d_{p1p2} - \epsilon \)

**Type 3: Far from 1 Parent Category**

**Condition 1:** \( \epsilon \leq d_{mp1} \leq d_{p1p2} - \epsilon \)

**Condition 2:** \( d_{mp2} \geq d_{p1p2} - \epsilon \)
4.5 Results

The overarching question is how the mixed race categories fit into the racial system. Building up to this, we must first ask whether the mixed race categories can be fully collapsed into a parent category, in which case the role of the mixed race categories is quite clear. There would be no relational differences between a mixed race category and a parent category, and the mixed race category would be part of a larger, more established racial/ethnic group. The results of the collapsing process suggest this is not the case. None of the mixed race categories can be easily collapsed into a parent category. The best fitting Goodman model leaves the mixed race categories as distinct racial/ethnic entities, yielding a worse fit when they are collapsed into other categories.

Thus, in the strictest sense, the mixed race categories are not exactly the same as the parent categories. There is some in-group bias and the pattern of social ties does not follow the parent categories precisely.

Relaxing this strict criteria, it also the case that none of the mixed race categories fall very close to one parent category, like in Panel 1 of Figure 10. A mixed race category of this type should be close to one parent category and far from the other (but only as far as the distance between the parent categories). None of the mixed race categories have
this pattern of distances, however. For example, White-Black is closer to Black than
White, but the distance to White is still much closer than the distance between White
and Black. Substantively, this means that White-Black individuals intermarry (or
cohabitate) with both Blacks and Whites at a rate higher than Whites and Blacks do with
each other. Some marriages are to Whites and some are to Blacks. Similarly, the pattern
of marriage/cohabitation does not strictly follow that of Whites or Blacks. This puts
White-Black firmly into the second structural Type in Figure 10: both White and Black,
or relationally between the parent categories. White-Black individuals are not
assimilated into an existing racial/ethnic category, but they do fit predictably into the
racial social space.

I have plotted the distances for White-Black, as well as the other mixed race
categories, in Figure 11. The figure plots 3 distances for each mixed race category: the
distances between the mixed race category and the two parent categories; and the
distance between the two parent categories. For example, the first panel on the left plots
the distance for White-Black. The top row shows the distance between White-Black and
the parent categories (White, Black). The bottom row shows the distance between White
and Black. I use this patterning of distances to place each mixed race category into a
structural type. These structural types are placed on the bottom of the plot for ease of
reference. For White-Black, the lighter colors on the top row relative to the bottom row
places it into the 2nd structural type—White and Black.
Figure 11: Distances between Mixed Race Categories and Parent Categories
Each subplot in Figure 11 represents a different mixed race category. The left hand side represents the non-Hispanic mixed race categories and the right hand side represents the Hispanic mixed race categories.

All of the non-Hispanic mixed race categories fall into the 2nd structural type. The exact distance scores may differ, but the patterning is much the same as we move down the rows on the left. White-Black, White-Other and White-Native American all fall relationally between the two parent categories. Each is socially closer to the parent categories than the parent categories are to each other. Thus, White-Other equals White and Other; White-Native American equals White and Native-American; White-Black equals White and Black.

The non-Hispanic categories do differ in important ways, however. Those who identify as White-Black and White-Other are socially farther from White than Black/Other. Thus, while White-Black and White-Other fall socially between the parent categories, they are still closer to the non-White parent category. This suggests that the economic, residential and cultural conditions for White-Black and White-Other individuals are far from that of the White majority. In that sense, they are really Black-White and Other-White. The opposite holds for White-Native Americans. White-Native Americans have higher rates of intermarriage with Whites than Native Americans (see

---

4 We begin to see some differences in the White-Native American plot. Here, White-Native American is equally far from Native American than White and Native American are from each other. Native American is still not close enough to White to fall into the first structural type (where White-Native American would be very close to White and equally far from Native American as White is).
also Qian and Lichter 2007). Also note that the social distances in the last plot, with Native American, are much lower than with Black-White or Other-White.

The Hispanic mixed race categories offer a very different story, occupying social locations that are often quite distant from the parent categories. White-Black-Hispanic and White-Other-Hispanic fall into the third structural type, defined by their distinct distance to Whites. In both cases, the distance to White is quite far, and farther than the distance between the parent categories. White-Black-Hispanics and White-Other-Hispanics are then defined by their lack of social contact with Whites, and their distinctly non-White pattern of marriage ties. White-Black-Hispanic and White-Other-Hispanics are likely to have very different occupational and residential profiles than Whites, and this is reflected in the social distance between the mixed race categories and Whites.

White-Native American-Hispanic offers a more extreme picture. White-Native American-Hispanics are equally far from both parent categories—White and Native American-Hispanic. The distance to the parent categories is also very similar to the distance between the parent categories. White-Native American-Hispanic thus represents a distinct category on its own, with its own relational patterning and low

\[\text{5 It is, perhaps, not surprising that White and Native American are much closer together than White and Black or White and Other. When White and the non-White parent category are socially close, it is possible for the mixed race category to be quite close to White.}\\
\[\text{6 Technically, the distance to White is slightly less than the distance between the parent categories. This places White-Native American-Hispanic on the cusp of the 3rd and 4th structural types.}\]
rates of contact with the parent categories (which are also distant from each other). If White-Black-Hispanic is not White, then White-Native-American-Hispanic is not White and not Native-American-Hispanic. White-Native American-Hispanics are instead socially closest to White-Native Americans, another mixed race category. White-Native American-Hispanic thus represents a new category on its own: not well defined by the parent categories and relationally closest to another “new” racial/ethnic category.

Overall, the results suggest that non-Hispanic mixed race categories fit predictably into the racial system. They are distinct but close to both racial parents. The Hispanic mixed race categories are less easily situated. They are more likely to be distinct racial/ethnic categories with their own relational patterning. They tend to be far from one or both of the parent categories, with lower rates of intermarriage than expected (based on the social distance between the parent categories). The creation of new racial/ethnic categories is thus rooted in the growing Hispanic population, rather than the intermixing of more established racial groups. The growing Hispanic population will continue to complicate an increasingly complicated racial system (Bonilla-Silva 2004; Rodriguez 2007).

4.6 Conclusion

The 2000 Census marked an important shift in the official measurement of race in America (Farley 2002; Lee and Bean 2004). For the first time, individuals were allowed to claim two racial categories simultaneously. This paper addresses this institutional shift
through a new, relational framework. I ask if the mixed race categories represent truly new, emerging racial/ethnic categories or whether they fit comfortably into the existing racial order. And more generally, did the change in measurement alter our view of race in America?

Using Census marriage data from 2000, I find that Hispanic and non-Hispanic mixed race categories have very different relational implications. Non-Hispanic mixed race categories, such as White-Black, are relationally between the parent categories. They have relatively high rates and similar patterns of intermarriage with both parent categories. White-Black is thus White and Black and fits predictably into the racial system. The Hispanic mixed race categories, such as White-Black-Hispanic, are, in contrast, relationally distinct from the parent categories. They have few marriage ties (and different marriage patterns) with one or both of the parent categories. For example, White-Black-Hispanic is distinctly not White.

The Hispanic mixed race categories thus represent distinct, new racial/ethnic categories. And more generally, Hispanics generate distinct racial/ethnic categories in a way that the mixing of traditional racial/ethnic categories does not. Hispanics may play an increasingly important role in shaping the racial landscape—as it is a growing population and intermixing with different racial groups yields new, not yet named, racial/ethnic categories. This may be another way that the experience of Hispanics differs from that of previous European immigrant groups (Frank et al. 2010).
It is also important that most mixed race categories (except those involving Native American) were socially closer to the non-White parent category. So, White-Black is socially closer to Black than White. This means that in most cases there are still strong barriers between non-White and White individuals, even for those with White parents.

Methodologically, the chapter offers a unique application of the model presented in Chapter 2. Here, I define a category by the social locations of two other categories. In Chapter 2 and 3, I followed the categories as they moved over fixed locations. Here, I hold the locations of the parent categories fixed. I then measure the mixed race categories by their revolution around these two locations. Thus, it is not simply the case that we can use the logic of social distance to trace categories across time. We can also use a social space framework to map out the meaning of new, not yet well understood categories.

Finally, the chapter offers a useful starting point for future work on race. For example, the analysis could be extended to an entire social system. Instead of asking if mixed race categories can be derived from parent categories, we can ask if any category can be “derived” from any other two categories. Or, we can ask what are the minimal categories and generating rules that will properly reproduce an entire racial social space, with the racial categories placed into logically correct social locations (Boorman and White 1976). This would be a formal way of studying racial differentiation and hierarchy. It would the capture the role of particular racial/ethnic categories through
time. More generally, this chapter makes clear that one can effectively study race/ethnicity through the social positions of demographic categories; that one can use racial/ethnic categories to characterize racial organization without assuming that categories are fixed, essential characteristics of an individual.
5. Using Categorical Fluidity to Test Theories of Stratification

How can sociologists test theories of macro-level power and influence? In a theory of cultural hegemony, for example, a dominant demographic/economic group sets the norms for the rest of the population (Gramsci 1971; Bourdieu 1984). In a theory of competition, advantaged groups react to economic or demographic threats from more disadvantaged groups (Blalock 1967; Olzak 1992). It can be difficult, however, to describe power relations at an aggregate level, and to systematically differentiate between competing theoretical predictions. Past work has generally examined one theory at one time point, if there is a formal test at all (although see Jacobs and Carmichael 2002). In this paper, I develop a new approach for testing theories of aggregate level power and apply the framework to the question of racial stratification. I ask how the interrelationships between different racial groups result in a particular type of stratification system—where society reflects some racial groups more than others (McDaniel 1995; Zuberi 2000; Bratter and Zuberi 2001). The focus is thus on the give and take between demographic groups, rather than any particular economic or cultural outcome.

The approach draws on three core sociological ideas: that demographic categories are the results of larger social processes (or are social constructions) (Barth 1969; Burt 1991; Wimmer 2008); that different demographic groups have differential
influence over the institutions in society (Bourdieu 1990; Bonilla-Silva 1997); and that the resources in society (education, income, health) are differentially distributed across demographic groups (Tilly 1998; McCall 2001; Leicht 2008). Put together, the claim is that a stratification system, here along racial lines, can be characterized by how different groups shape meaning in society; and, more specifically, how one racial group can shape the meaning of other demographic dimensions, such as education, in another racial group. I thus apply the “radical” sociological notion that demographic categories are fluid to the particular problem of measuring influence at an aggregate level. The differential distribution of resources makes it possible to see who reacts to whom and how the process of aggregate influence works.

The key to the approach is having a systematic, formal means of measuring the meaning of demographic categories over time—with the focus here on measuring educational categories for the sake of characterizing racial dynamics. I draw on the approach developed in Chapter 2. The basic idea is to define demographic categories by observed social distinctions rather than the labels themselves. Demographic groups allowed to marry/cohabit in the same manner are considered equivalent across time, even if the labels are different. If College graduates now move in society like High School graduates used to, then College is the new HS. Formally, the framework takes interaction patterns, such as marriage or friendship rates, and calculates the social distances between demographic categories. The model then uses these distances to
measure the changing (relational) meaning of the categories. The model tests theories of social dynamics by predicting which categories are relationally equivalent across time.

Using this approach, I first show how the relational meaning of education has changed over time and how these changes correspond to changes in the level of educational attainment. As the level of attainment increases, the educational categories are equated with lower grades from the past. For example, as a high school degree becomes more prevalent, the social distinctions around that degree should decrease. HS in 1980 may be the relational equivalent of Grade 8 from 1940.

I then use those joint movements, educational attainment and educational meaning, to characterize the racial stratification system. The basic question captures the give and take between racial groups: how does the relational meaning of education amongst one racial group respond to changes in educational attainment amongst another racial group? I focus on four idealized racial stratification systems. First, the racial system could be characterized by a system of separate social worlds. Here, the relational meaning of education is best predicted by changes in attainment happening within racial groups. In contrast, a system could be characterized by a system of hegemony. Here, changes in attainment amongst the majority racial group shape the relational meaning of education across the entire population. We could also see a system of minority threat, where increases in attainment amongst the minority drive the relational meaning of
education amongst the majority (by devaluing certain levels of education). Finally, there could be a homogenous system, where the meaning of education is predicted equally well by changes in attainment occurring within or outside a racial group.

Using census data form 1940-2000, I find that racial stratification in the US is characterized by a system of hegemony, although this has weakened somewhat over time. The educational attainment amongst those in the majority shapes what social relationships are likely to form for those in other racial groups. The results suggest that the institutionally reinforced status differences across racial groups creates a system where those in power determine what is good and normal for the rest of the population.

5.1 Measuring Stratification Systems

Models of demographic stratification typically fall into one of three traditions. The first, and by far most common, approach is to characterize differentiation by the fortunes of different demographic groups. For example, a study on racial inequality may ask how wealth or health is distributed across racial groups (Shapiro 2004; Orsi, Margellos-Anast and Whitman 2010). A second tradition explores stratification systems by studying the meaning of the demographic categories themselves (Burt 1991; Brubaker 2004; Wimmer 2008). Being Irish or Italian may have really mattered for where you lived and who you married in the past but this is no longer the case (or at least less so). In that sense, the Irish are now White, and a stratification system can be understood through this emergence and dissipation of salient demographic categories (Lieberson and Waters
1986; Loveman 1999; Brubaker 2009). A third tradition focuses on the role that different
groups play in the formation of a normative culture. Thus the consumption, linguistic
and behavioral tendencies of one group becomes what is “good” and correct while the
practices of another are devalued. For example, studies have explored the consequences
of white beauty norms being the assumed standard in society, while others have
discussed the mapping of cultural tastes by class (Bourdieu 1984; Hill 2002; Alderson,
Junisbai Heacock 2007).

This paper offers an alternative framework for studying stratification, one that
utilizes aspects of all three traditions but does not comfortably fit into any. The approach
is a general one, although the paper is substantively concerned with racial stratification
over time in the US. Like the second tradition, I use the changing meaning of
demographic categories to understand the dynamics of a stratification system. The
question, however, is how different racial groups influence the meaning of a different
demographic dimension, such as education, rather than how the racial categories
themselves are changing. In this sense, the approach is like the third tradition, but
instead of using norm formation to explore racial inequality, the focus is on the shaping
of demographic categories. Similarly, the approach both uses and diverges from the first
tradition, which compares the economic (or health, etc.) fortunes of different
demographic groups over time. The changing conditions, economic for example, are not
used as a metric of increasing or decreasing inequality. Rather, the changing conditions
are used as a tool to determine which racial groups drive the meaning of the demographic categories.

The framework rests on three basic ideas. First, I begin with the idea that demographic categories can be characterized as social constructions (Barth 1969; Anthias 2001; Zuberi and Bonilla-Silva 2008). Few sociologists would think of an ethnic label, for example, as an essential characteristic of an individual; rather, what is important are the conditions mapped on to or correlated with an ethnicity. There is then a distance between a label and a social position, and demographic categories can be seen as the result of larger social processes, such as physical segregation, economic inequality, or even overt classificatory struggles (Frank, Akresh, and Lu 2010). I extend this logic to other demographic dimensions, such as education, age, or occupation (Burt 1991). I thus assume that there is a social meaning/implication mapped on to the categories that is distinct from the label itself.

Second, I assume that the meaning of a demographic category can change over time as the underlying forces structuring the category change. For example, as the economic conditions force later retirement and life expectancy increases, we may expect that the meaning of “old” gets pushed back over time (Sanderson and Scherbov 2005). Or, the meaning of a high school degree may have changed over time as the barriers to HS graduation weaken and the proportion earning that degree increase (Smith 1993; Wilson, Zozula and Gove 2011).
Third, and most importantly, I suggest that the underlying conditions among one set of individuals (“Whites”) can drive the meaning of the categories among another set of individuals (“Blacks”). Thus the experience, or changing conditions, among one group is what matters for shaping meaning in a population. For example, we can imagine a system where the majority racial group is completely dominant—so that the material and cultural conditions amongst whites determine the meaning of categories (e.g. for education) across the entire population. A system of cultural hegemony is only one possible system, and the goal is to test whether such a system (or something else) best fits the data.

I thus characterize the stratification system by the joint changes of the demographic categories and the underlying conditions driving those categories. This is an ideal way of capturing inequality in a large, complex system; societal influence is having the rest of the population react to changes amongst a subset of individuals. We can ask who reacts to whom by following the movement of conditions and categories over time. Thus, different (racial) groups may play different (or no) roles in determining the social meaning of education and that distribution of roles tells us something about stratification system over time.

The framework demands both a measure of changing conditions and the meaning of the demographic categories. The first question is how to measure the changing social meaning of the demographic categories over time. I then develop and
test a theory of racial stratification that utilizes joint changes in educational categories and the distribution of education (the underlying conditions) over time.

**5.2 Background on Measuring Social Change as Categorical Change**

I draw on the approach presented in Chapter 2 to measure the changing meaning of demographic categories over time. The framework offers a formal means of putting a radical sociological idea into practice: that a demographic category is fluid, and need not mean the same thing at one time point as another, just because the label is the same.

The overall idea is to measure categories through interaction patterns, such as marriage or friendship rates. The meaning of a category is then a “relational meaning” through time, where the categories are equated based on behavioral indicators, rather than the categorical labels themselves. One cannot use the labels to define the categories as the categories are supposed to be fluid, non-essential and contingent. Interaction patterns offer an ideal categorical metric as they concretely capture social distinctions in a population (Qian and Lichter 2007; Diprete et al. 2011). Who you spend your time with is clearly important to people. It is then telling when there are few observed social ties between demographic groups. Interaction patterns thus give us a way of describing where there are observed/revealed distinctions in a population and where there are not. For example, if Muslims were allowed to marry like Catholics, make friends like Catholics, then Muslims would be Catholics in all ways but name-and the labels
themselves cannot be used to define the categories (see Burt 1991 for a similar argument with age categories).

Using interaction patterns as a metric, different categories may come to reflect the same pattern of distinctions, or limitations, over time and these changes form the basis of measuring categorical change. Two categories are the same over time if the observed behavioral implications of group membership are the same, even if the label is different. Formally, I measure categorical change by first placing categories into social locations based on observed rates of interaction (e.g. frequency of marriage rate between categories), where locations reflect what interactions are possible or likely for individuals with a given demographic profile. Categories are placed close to each other if interaction is likely and far if interaction is unlikely. I then ask which categories occupy the same positions over time, or which categories have the same range of possible friendships or marriage partners.

Figure 12 plots an example social space based on educational categories. The figure includes the social locations for two years based on hypothetical marriage data. For example, in time 1 those with <HS are more likely to marry those with HS than College. The figure also traces out where the categories move from Time 1 to Time 2. We can see that the “median” social category changes from Some College to College. This means that those with College in Time 2 are allowed to do the same things, or move in society in the same ways, as those with Some College in Time 1.
And more specifically, categories occupying the median social location are equally likely to marry the very top as the very bottom. Allowing for the possibility that the top/bottom has also changed occupants over time (for example, the top shifts from Master/PhD to PhD), this profile of interaction, equal chance marrying top/bottom, corresponded to Some College in Time 1 but College in Time 2. Demographic categories with different labels but the same behavioral profiles are thus considered equivalent, or mean the same thing, over time. We can then use this approach to predict which category at time 1 is the equivalent of which category at time 2. Formally, the approach draws on the strengths of two models, social distance and social space, to estimate the social locations and to compare the categories across time. See Chapter 2 or the methods section in this paper for more technical details.
5.3 Using Educational Change to Characterize Racial Systems of Stratification

Education is an ideal demographic characteristic to characterize racial stratification. The analysis will ultimately take a demographic dimension other than race and use the joint changes in underlying conditions and categorical meaning to characterize racial inequality. This other demographic dimension is then used as a tool to tease out the dynamics of the racial stratification system. An ideal set of demographic dimensions (education by race) will exhibit two key properties.

5.3.1 Changes in Categorical Meaning and Underlying Conditions

First, the “metric” demographic dimension must change meaning over time and react clearly to larger structural forces. It is necessary to have joint changes in the (here educational) categories and the underlying conditions driving those categories. This makes it possible to see which racial group (or really location) is reacting to which racial group. With only changes in the categories, for example, it would be clear that the racial locations were becoming more or less similar, but not why. With both conditions and categories changing, it becomes clear what happens to the categorical meaning among one racial location when the underlying conditions change in another.

Education has exhibited clear, “conducive” changes over time: first, the meaning of the categories have changed over time, measured by the occupied social locations; and second, the categorical changes follow shifts in the underlying conditions, measured
by increasing levels of attainment.\footnote{Changes in attainment are, in turn, proxies for other forces that structure education over time—changes in government policies, credentialism, changes in the occupational structure. Attainment changes thus serve as an ideal summary measure of changing conditions.} There is thus a clear measure of conditions, which change over time, while the meaning of the categories follow those changes quite closely.

Over time, the level of education has increased steadily. While it is now normative to earn a high school degree, this was not the case in the early part of the 20th century (Mare 1995; Snyder, Tan and Hoffman 2006). We can view this increase in educational attainment as a result of institutional, occupational and competitive pressures.\footnote{Institutionally, the educational system has changed the baseline expectations for education (say HS), thus increasing the total number of people attaining a previously unattainable degree. Occupationally, labor market changes have made manual labor a less realistic option, increasing the demand for higher, more formal education (Goyette 2008). Competitively, individuals with (previously) high education may raise the bar when those with lower education increase their attainment. Theories of credentialism (Collins 1979) and maintained inequality (Lucas 2001) show how individuals protect their family’s position by improving their children’s educational profile—either in terms of actual years attained or through more subtle means such as class choice or major (Lucas 1999).}

The general trend towards educational inflation should be matched with changes in the social locations, or relational meaning, of the categories. In general, the social distinctions attached to an educational category will decrease as the proportion earning that degree or more increases. For example, movement between <HS and HS would be difficult if HS represents a rare, distinct badge in society. A degree that is difficult to attain will thus occupy a “high” social location, far from lower educational categories. As HS becomes more common, however, the occupied social location will shift, where
HS in time 2 may occupy the same location as a <HS category from time 1. The educational categories should thus be equated with increasingly higher levels of attainment over time as the level of education increases.

I empirically explore this in Figure 13, which uses census marriage data from 1940-2000 to measure educational categories over time. For each year, I take the demographic information of the couples and use a log-linear distance model to map out the locations, or values, of each category along an education dimension. The locations in the space correspond to social locations calculated from the observed marriage rates. The locations reflect the ease or difficulty in movement across the demographic space, and categories occupying the same social location over time are considered equivalent (and colored the same).

These results are taken from Chapter 2, and clearly show signs of inflation, following the general trend in attainment. The “median” social category changes from grade 8 to grade 10 to high school by 1980. We can also follow the path of particular grades over time. For example, grade 12 moves from a relative location of .71 to .61 to .48 and .37 in 2000. The number of marriage ties between HS and the bottom educational categories thus increases over time. If most people get a high school degree, then high school will no longer serve as a distinguishing marker in society.
Education and Race as Nested Social Structure

The demographic dimensions must also exhibit a nested relational structure, with the “metric” dimension (education) nested within the other demographic dimension (race). Here, race is an ideal conditioning dimension (education by race) as it is strongly structures education relationally. Most marriages (and interaction more generally) takes place within, rather than across, racial lines (Blackwell and Lichter 2004; Qian and Lichter 2007). And while this is slowly changing over time, race is still a stronger determinant of social relationships than education (Marsden 1988; Rosenfeld 2008). For example, Smith, McPherson and Smith-Lovin (2013) found that a distance of 12 years of education is the equivalent of mismatching on race-in terms of decreasing the
probability of forming a strong social relationship. This means that education is likely to change meaning within racial locations.\textsuperscript{3} We can then plausibly ask whether the meaning of education for a racial location was driven by what is happening in another racial location. Otherwise, one is comparing categorical change where change does not take place, making the comparison inconsequential at best and misleading at worst.

5.4 \textit{Explaining Educational Change by Race}

I use the systematic relationship between educational attainment and the educational locations, as well as the nested structure of education/race, to explore racial stratification over time. I can then ask how educational attainment at a given racial location predicts the social locations of education at another racial location. It is useful to define two generic types of change: first, changes taking place within a majority racial location; and second, changes taking place within a minority racial location. For each generic racial location, I assume there is information on educational attainment and educational (social) locations over time. For concreteness, we can think of the majority and minority racial groups as Whites and Blacks in the US over the last 70 years.

Given these distinctions, the question is how the process of educational change varies by minority/majority locations. Do locational shifts in education respond to

\textsuperscript{3} It is, of course, also possible that race itself is not constant over time. The suggestion here is simply that the racial dimension is dominant. The theory and analysis thus assume that educational changes happen within racial locations but not that the racial category-location mapping is fixed. If racial category j in time 2 occupies the same location as racial category i in time 1, then I would compare the educational mapping in racial category i to the educational mapping in racial category j. The basic units where educational change takes place are racial locations, and not the racial categories themselves.
attainment changes taking place within or outside a racial group? And more specifically, does the relational meaning of education amongst the racial minority respond to changes in educational attainment amongst the racial minority or the racial majority?

Race can shape the relationship between educational attainment and educational locations in four main ways. These four broad types characterize the possible racial ordering of a system, where different racial regimes will exhibit different traces, or profiles, of educational change. By having concrete conditions attached to each racial regime, it becomes clear what system best characterizes society over time.

5.4.1 Homogeneity

A homogenous racial system is characterized by parallel educational changes across racial locations. Locational shifts in education are predicted equally well by within race attainment changes as outside race attainment changes (in education). For example, we could ask which educational categories occupy analogous locations across time within the “Black” location. It would then be equally effective to use “White” changes in educational attainment as “Black” changes in educational attainment to predict the equivalent educational categories. Here, the process of educational change is isomorphic across racial locations. This does not mean that the level of educational attainment, or even the social locations, are the same; rather, it means that attainment changes in one racial location do not clearly drive locational changes in another. And more substantively, this suggests that the same forces, such as educational initiatives
and occupational shifts, push educational meaning (through attainment change) across the racial landscape—even if the minority location has lower absolute levels of education. The meaning of the categories is thus driven largely by exogenous forces operating (more or less) uniformly across the population.\(^4\)

### 5.4.2 Separate Worlds

Alternatively, the racial system could be characterized by a system of separate worlds, where the educational changes are localized in social space. In this case, using minority attainment (for education) to predict minority educational locations is better than using majority attainment to predict minority education locations. Similarly, the majority education locations are better predicted by majority attainment than by minority attainment (for education). Intuitively, the best predictor of the meaning of education amongst a racial group is the level of attainment for that group. In a separate worlds system, individuals are only concerned and potentially aware of those who are socially close. If there is little social contact between racial groups, then one may be unaware of what is happening in a distant racial location (Mark 2003).

Even if individuals knew what was happening in a different racial location, it may not matter. This local concern stems from the racial segmentation in labor and marriage markets (Bonacich 1972; Kalmijn 1993; Gullickson 2006).Occupationally,

\(^4\) It is also possible that homogeneity results from credentialing pressures that act uniformly across racial groups.
different racial groups may be systemically filtered into different types of jobs (Huffman and Cohen 2004; Tomaskovic-Devey et al. 2006). Interpersonally, we have already seen that educational differentiation in marriage occurs within race. What does it matter if education increases amongst individuals competing for completely different stakes? Or, alternatively, the normative level of education is determined by what everyone around you is attaining. And given the large racial divides in the population, this means that educational changes in other racial locations are less important than what is happening locally.

Similarly, in a separate worlds system, the opportunity structure for interaction is shaped by the demographic characteristics within a racial location. The opportunity for interaction is shaped by the sorting of individuals into organizations and occupations (McPherson and Smith-Lovin 1987; McPherson 2004). It is difficult to form relationships with people you never meet and you meet people within particular organizational contexts (Feld 1981). In a separate worlds system, the educational composition of organizations is structured primarily by the educational availability within racial locations. Social relationships are thus shaped by the distribution of education within the racial location as it maps onto the foci where social relationships form.
5.4.3 Hegemony

A racial system could also be characterized by hegemony. Here, the meaning of education in the minority racial location is driven by educational attainment changes in the majority location. For example, “Black” educational locations would be better predicted by white attainment changes in education than by Black attainment changes. Thus, what education means is driven by the majority: when educational attainment changes in the majority, there are parallel changes in the educational locations of the minority racial group, even though such changes do not correspond to the attainment changes (in education) occurring there.

Substantively, the racial majority pervades over all parts of society, and the social meaning of education is driven by the tacitly assumed correctness of those in power (Gramsci 1971; Bourdieu 1984; Bourdieu 1990). People look up to those with higher status, where status (at an aggregate level) is derived from and maintained through institutional power. Individuals in dominated groups will then assume the correctness of the existing social order. How those in high status positions “do” education is being what matters for the rest of the population. Thus, those in minority racial locations act as if they had the same educational attainment as the majority, even though this is not the case (or need not be the case). For example, it well established that Black students highly value education even in settings where attaining higher education is unlikely (Ainsworth-Darnell and Downey 1998; Harris 2011). Or, White beauty norms pervade
over the population, where what is beautiful is determined by what Whites do (Hill 2002; Rosado 2004). In this case, those in the minority find a mate in a manner consistent with the valuation of education based on a White or majority metric. One thus interprets the education of a mate (is that “high” education?) by the attainment changes occurring upwards in the system, rather than what is occurring locally, within a racial location. The social meaning of education is thus almost universal, dictated by the majority, but the opportunities for educational change are quite different across racial locations.

A system of hegemony also has structural implications. Here, the sorting process of individuals into foci is determined by the educational attainment of those in the majority. If everyone has a college degree, then those with HS will no longer populate the same organizations as in the past. More importantly, those in the minority racial locations are sorted into organizations and occupations based on changes in attainment happening amongst the majority, even if there are no parallel shifts amongst minority members. The recruiting process still happens locally, by race, but any deviations from the overall educational composition of the racial location will follow the institutional logic (or valuation) of the majority. This means that the minority opportunity structure, at the level of the organization, will parallel that of the majority. Someone with a college degree in the minority location will be surrounded (educationally) by similar people as a majority member with a college degree, even though the overall proportion of people earning a college degree is lower amongst the minority. The social relationships formed
amongst minority members will then look like the relationships formed by the majority members.

5.4.4 Minority Threat

The inverse of hegemony is a system of strong racial competition (Blalock 1967; Olzak 1992): here changes in educational attainment in the minority location drive changes in the social meaning of education in the majority racial location. For example, using “Black” educational attainment to predict White educational locations is better than using White educational attainment to predict White education locations. Here, the majority reacts strongly to changes in minority attainment. Educational increases from the minority location pose a threat to majority control over the resources in society. For example, Lieberson describes white southerners visceral reaction to Black educational attainment in the post civil war south (Lieberson 1980). Similarly, a number of studies have explored the political consequences (on laws or attitudes) of a having large minority population in an area (Taylor 1998; Baumer, Messner, and Rosenfeld 2003). Reacting to potential threats in this context, those in the majority location may make exaggerated social distinctions based on attainment changes in the minority location. For example, increases in HS attainment within the minority location would devalue HS in the majority location, even if the number of HS graduates is constant in the majority location. The devaluation of HS would correspond, socially, with HS shifting to a lower social location and being equated with higher grades over time in the majority location.
These four systems reflect idealized types of racial stratification. Importantly, only some theoretical models will be empirically consistent with the observed data. In this sense, the approach is a formal way of testing difficult to test theories. For example, it can be a difficult task to formally assess a theory of Gramscian order, where those in power set the values and rules of a system (Gramsci 1971). Here we can see these kinds of dynamics in the give and take of different racial groups as the level and meaning of education shifts over time.

There are two key empirical questions in the paper: 1) what racial/education system best characterizes society; 2) how has this changed over time. Empirically, from the early 1900’s to today, there has been a convergence on educational attainment across the racial landscape. For example, the differences in educational attainment for Blacks is closer to Whites today than 70 or so years ago. The convergence on the level of education should, all things being equal, move the racial/education system to one of homogeneity. As the level of educational attainment converges, so too should the educational location. The question is what the system looked like 70 years ago and

Note that the four systems are mutually exclusive. For example, a separate worlds system occurs only if hegemony or minority pressure are not particularly strong forces. This view parallels Mark’s work on cultural taste (Mark 2003). Mark considered theories of cultural distancing and homophily. Under cultural distancing, individuals make an effort to differentiate themselves from those with lower culture; this is similar to the dynamics occurring under minority threat. With homophily, there is localized interaction and a diffusion of culture within bounded regions of social space. This is the dynamics of a spate worlds system: interaction is bounded by racial lines, while the “diffusion” is the appropriate level of educational attainment (given job market conditions and institutional agendas).
whether convergence on attainment did, in fact, push the system towards homogeneity or if other changes disrupted that transition.

5.5 Data

The analysis of educational/racial stratification is based on US census data from 1940-2000. The data come from 1% samples of the Public Use Microdata (Ruggles et al. 2010). The data include information on education and race for both respondents and spouses (as well as other individuals in the household). Additionally, the datasets are large enough to analyze educational mixing within racial groups and cover a long enough time period to examine over time change. Following Rosenfeld (2008), I use 20 year non-overlapping cohorts. The datasets thus include the 1940, 1960, 1980 and 2000 censuses. I restrict the analysis to married couples between the ages of 20 and 39. This offers a comparison of same (or similar) age couples across time. I use younger couples as they represent the bulk of marriages occurring within the 20 year period (e.g. between 1941 and 1960 for the 1960 census).

The U.S. Census measures education as years of attainment up to the 1990 survey. From 1990 on, education is measured as the highest degree attained, assuming the individual has more than a high school education. For the 1940, 1960, and 1980 data, I use all 18 years of education as the initial categories in the analysis (the exact number varies slightly by year). For the 2000 data, I use the available grades/degrees as the
measure of education. This includes 0-4, 5-6, 7-8, 9, 10, 11, HS, Some College, Associate Degree, Masters, Professional Degree and PhD.

Race is measured less consistently than education in the census. Race is measured as 5 broad racial categories from 1940-1960: White, Black, Native American, Japanese, Chinese (there are also residual categories). The 1980 data are consistent with the 1940-1960 data but offer more detailed Asian-Pacific options. Additionally, in 1980 the Census begin asking individuals to identify as Hispanic. Prior to 1980, Hispanic is imputed based on language, origin, name and the like. In 2000, the Census allowed for a much larger number of racial answers, including mixed, or multiple identifications.

The initial analysis is restricted to Black and White racial groups. A simplified White/Black space makes it easier to map the results back to the theoretical discussion of majority/minority racial locations. It is also less clear that the relational meaning of education is consistently determined within the other racial locations. The other minority race boundaries are somewhat more permeable than the Black/White divide. Additionally, Black/White locations remain clearly distinct over time, and I can use those categories as (proxies for) the racial locations throughout the analysis. This is important as the results compare educational shifts occurring within racial locations. It is then necessary to know which racial categories occupy the same location over time. Practically, there is little need to check if the White location in Time 1 corresponds to the White location in Time 2 if the only options are Black and White.
5.6 Methods

5.6.1 Measuring Educational Attainment and Educational Meaning

The analysis begins with a description of education by time and racial location. The two key pieces of information are the level of educational attainment and the shifting relational meaning of the educational categories. I measure educational attainment as the cumulative distribution by race. The cumulative distribution is the proportion of individuals (within that racial location and time) that have the given level of education or lower.

The relational meaning of the educational categories are measured using the framework detailed in Chapter 2, which draws on models of social space and social distance. The categories are first placed into social locations based on observed patterns of interaction. The categories are then equated across time if they occupy the same social location, or have the same range of possible social ties at different time points.

The idea is to characterize the social world as a multi-dimensional demographic space, where the demographic characteristics of interest, here education, make up the dimensions of the space (McPherson 1983). The assumption is that the probability of interaction decreases as demographic distance increases, or that homophily structures social interactions. Locations are thus proxies for what interactions are possible for individuals with a given demographic profile. The approach differs from a traditional social space approach as the locations are estimated from observed interaction rates, as
opposed to having the demographic characteristics themselves make up the metric of the space (i.e. years of education). The model thus estimates latent social locations with known axes. The approach differs from traditional latent variable models as the dimensions of the space are known a priori. This sidesteps the difficult problem of trying to interpret the dimensions in a latent variable model and makes comparing the categories across time tractable.

Formally, the approach has four steps: first, calculate the social locations using an RC Goodman model (Goodman 1979); second, collapse categories if they are relationally identical; third, scale the locations to make them comparable across time; and fourth determine which categories occupy the analogous locations across time.

The RC model follows the form of:

$$
\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta \varphi_i \phi_j
$$

where $F_{ij}$ is the frequency count in cell ij; $\mu$ captures the overall mean, $\mu_i^R$ and $\mu_j^C$ are the row and column means; $\beta$ is a coefficient and $\varphi_i$ and $\phi_j$ are the unknown, estimated location scores (which can be constrained to be equal). The model predicts the frequency of social ties (so marriage or friendship) between categories i and j as a function of the marginals and the estimated locations. Net of chance expectations, the goal is to estimate the latent locations so that demographic categories are “close” if they have high frequency of marriage (or friendship, etc.) and are “far” in the space if marriage between the categories is unlikely.
After calculating the locations, I collapse categories if they are relationally identical, and thus (should) occupy the same social location. Two categories are relationally identical if two conditions are met: a) there are no barriers to interaction (so within category ties are just as likely as ties to the other category); b) they have the same pattern of ties to all other categories. If both conditions are met, the categories are relationally doing the same thing and should be treated as one category. I test whether categories should be collapsed using the Goodman model for the collapsibility of rows/columns in a log-linear model (Goodman 1981).

After collapsing relationally equivalent categories, the locational scores are scaled to ensure comparability across time. The locational values are scaled to range between 0-1. Substantively, this shifts the space to relative locations. Thus, two categories with the same locational score across time hold the same relative social position, or have the same profile of possible social interactions. We can then ask which

---

6 I start with the closest categories in the space, calculating the deviance from a Goodman model with those categories collapsed. I then move to next closest pair and so on until are categories are collapsed into one large aggregate. In the end, I take the collapsing set with the lowest amount of information loss given the number of categories collapsed. See Chapter 1 for more details. Formally,

$$\log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \mu_{ij}^{RC}$$

where $\mu_{ij}^{RC} = \mu_p^{RC}$ for $(i, j) \in \text{subtable } p$, such that $i$ and $j$ are treated equivalently

I compare the BIC from this model under different sets of collapsed categories. This also equivalent to comparing the deviance from a full model, with all interactions, to the deviance with the specified categories collapsed. If the categories are relationally equivalent, then little information will be lost when the two categories are collapsed (so the deviance will be close to the deviance under the full model).

7 I also restrict the calculation to a one dimensional space. This means that the locations across year correspond directly to an educational dimension in social space (as opposed to a location in a multidimensional space where the meaning of the dimensions are unknown). This ensures that the comparisons across years are tractable (see chapter 1 for more details).
category at Time 2 occupies the closest location to a given category at T1. This process is repeated for each racial group.\(^8\)

### 5.6.2 Predicting Educational Categories across Time

Given the measures of educational locations and attainment, the analysis predicts changes in locations as a function of changes in educational attainment. The goal is to predict which categories across time are social equivalents, or relationally mean the same thing. Changes in the cumulative distribution of education are used to predict the equivalent educational categories. The basic question is whether attainment changes in one racial location can predict the changing meaning of education in another.

There are three parts to the analysis. First, I create the dependent variable, defined as \(Y_{ij}\). \(Y_{ij}\) is a 0/1 variable describing if categories occupy the same location over time, or are relationally equivalent. I create a dataset pairing all i categories in time 1 with all j categories in time 2. \(Y_{ij}=1\) if category i and category j occupy the same location in time 1 and time 2.\(^9\) \(Y_{ij}=0\) otherwise. For example, looking at Figure 13, Grade 13 in 1960 occupies the same location as Grade 12 in 1940 while Grade 12 occupies a different

\(^8\) It can be difficult to properly estimate the educational locations in a racial location with few sampled individuals. This makes it necessary to collapse educational categories prior to calculating the educational locations. For each year and racial location, I employ a series of Goodman tests to determine which, if any, of the educational categories can be collapsed with little loss of information (Goodman 1981). I first form a table describing the rates of marriage between educational categories. Male education is placed on the rows, while female education is placed on the columns. Two categories are collapsed if treating them as one category does not statistically alter the relationship between the rows and columns in the table.

\(^9\) Technically, \(Y_{ij}=1\) if category j in Time 2 is in the closest location (compared with all occupied locations in Time 2) to the location occupied by category i in Time 1.
location over time. Thus, \( Y_{12,13} = 1 \) while \( Y_{12,12} = 0 \). I define \( Y_{ij}^{maj} \) and \( Y_{ij}^{min} \) as \( Y_{ij} \) for the majority and minority racial locations. I calculate \( Y_{ij} \) separately for each racial location.

The second part of the analysis constructs the independent variable, \( X_{ij} \), predicting \( Y_{ij} \) based on changes in educational attainment. I define \( X_{ij} \) as the predicted distance between category \( i \) in time 1 and category \( j \) in time 2, given the time 1 locations.

Formally:

\[
X_{ij} = |l_{t1i} - (l_{t1j} + (a_{t2j} - a_{t1j}))|
\]

Where, \( a_{tj} \) is the proportion of the population earning degree \( j \) or lower in time \( t \). \( X_{ij} \) compares the Time 1 location of category \( i \), \( l_{t1i} \), to the predicted Time 2 location of \( j \), \( l_{t1j} + (a_{t2j} - a_{t1j}) \). Specifically, I predict the Time 2 location of category \( j \) by adding or subtracting shifts in attainment to the Time 1 locations. We can define these predicted values as \( pl_{t2j} \). I take the Time 1 locations as given so that only the shifts in attainment drive the predicted changes in which category should be equivalent to which category. The baseline prediction, with no shifts in attainment, is that category \( i \) should be equated with category \( i \) over time.

The basic idea is that educational categories that are becoming more prevalent should move to lower social locations. By moving to lower social locations, we would predict that they should be equated with lower educational categories from the past. For example, assume HS is in location .5 in time 1 and Grade 11 is in location .4. Also assume that the proportion of the population earning a HS degree or lower moves from
.7 to .6 between time 1 and time 2. Thus, HS is become more common in the population. The predicted location for HS in time 2 would be .4, following the .1 shift downward in the cumulative distribution and the .5 starting location \((p_{t2j}=l_{t1j} + (a_{t2j} - a_{t1j}) = .5 - (0.7 - 0.6) = .4\). \(X_{\text{Grade11,HS}}\) would then equal \(|l_{t1j} - p_{t2j}|=.4-.4=0.\) We would predict that HS in Time 2 should occupy the location previously occupied by Grade 11.

Finally, I regress \(Y_{ij}\) on \(X_{ij}\) using logistic regression. I thus predict which category will be equivalent to which category based on changes in the cumulative distribution of education. If the predictions are accurate, then categorical pairs with low values in \(X_{ij}\) should be more likely to be equated over time-as they occupy similar positions at different time points.

The analysis is repeated twice for each racial location. For a minority racial location, the analysis first predicts the educational equivalents, \(Y_{\text{min}ij}\), as a function of minority location predictions \((X_{\text{min}ij})\). The analysis then predicts \(Y_{\text{min}ij}\) as a function of majority location predictions \((X_{\text{maj}ij})\).\(^{10}\) The only difference between the two regressions is the observed changes in the cumulative distribution. If the changes in attainment were the same across racial locations, then the predicted locations would be the same. The analysis is then reversed. The majority equivalents, \(Y_{\text{maj}ij}\), are regressed on majority and minority predictions.

\(^{10}\) The majority predictions start with the minority locations in time 1; they then add the appropriate value based on changes in attainment in the majority racial location. The minority predictions similarly start with the minority locations in time 1, but add the appropriate attainment shifts from the minority racial location.
It is important to map the theoretical conditions for each theory back to the regression models. The question is what regression results, based on Ymaj, Ymin, Xmaj, and Xmin, would correspond to each racial system. For example, under a homogenous system, it is equally effective to use majority or minority attainment to predict the meaning of educational categories in the majority racial location. Similarly, it is equally effective to use majority or minority attainment to predict educational categories in the minority racial location. The conditions consistent with that world can be written as:

**Homogeneity:**

\[ Y_{maj_i} - X_{maj_i} = Y_{maj_i} - X_{min_i} \]
\[ Y_{min_i} - X_{min_i} = Y_{min_i} - X_{maj_i} \]

Where “~” means regress Y on X and the “=” signs implies the fit of the regression is the same.

We can map similar conditions for the other racial systems. Under a system of separate worlds, it is better to predict the educational categories by attainment changes within a racial group.

**Separate Worlds:**

\[ Y_{maj_i} - X_{maj_i} > Y_{maj_i} - X_{min_i} \]
\[ Y_{min_i} - X_{min_i} > Y_{min_i} - X_{maj_i} \]

Under hegemony, majority attainment is a better predictor of educational categories in both the majority and minority locations. Thus:

**Hegemony:**

\[ Y_{maj_i} - X_{maj_i} > Y_{maj_i} - X_{min_i} \]
\[ Y_{min_i} - X_{min_i} < Y_{min_i} - X_{maj_i} \]
In a system of minority threat, the meaning of education in the majority location is best predicted by changes in attainment in the minority location:

Minority Threat: \( \text{Ymaj}_i \sim \text{Xmaj}_i < \text{Ymaj}_i \sim \text{Xmin}_i \).

\( \text{Ymin}_i \sim \text{Xmin}_i > \text{Ymin}_i \sim \text{Xmaj}_i \).

There are thus clear, testable conditions linking systems of racial stratification to the data. The conditions clearly delineate the systems and are derived directly from the underlying theoretical arguments. We can then use these conditions to decide what system best fits the data and whether this has changed over time.

**5.7 Results**

**5.7.1 Cross Sectional Results**

I begin the results section with a description of education across time and race. Figure 14 maps the inferred social locations for education by race. Each panel has four rows, one for each year in the analysis. The locations range from 0-1 across all panels, running (roughly) from low education to high education. In both the black and white racial locations, there is a general trend of inflation as “high” categories in 1940 occupy lower locations by 2000. As with Chapter 2, there is a compression at the bottom of the distribution. Grades 0-9 occupy a much smaller portion of the space over time in both racial locations.
This broad similarity across racial locations is matched in the quantitative comparison in Table 3, which presents a snapshot of the key variables. For each year, I compare the educational locations across racial locations using a standard mean square error score. This captures how different, or how far apart, the educational locations in one racial location is to the educational locations in another. Lower values mean that the locations in one racial location are closely approximated by the locations in another. I calculate the MSE for each year.

In 1940, for example, the average distance (or MSE) between educational locations in the black location and white location was .043. The locations grow apart...
between 1940 and 2000, with MSE values of .11 and .095 in 1980 and 2000. Thus, in 1980, educational categories in different racial locations are more likely to occupy different social locations. The overall picture is, however, still one of similarity across racial locations.

Table 3: Summary Statistics for Race and Education over Time

<table>
<thead>
<tr>
<th>Year</th>
<th>1940</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance between Black and White Education Locations (MSE)</td>
<td>.060</td>
<td>.073</td>
<td>.080</td>
<td>.085</td>
</tr>
<tr>
<td>Distance between Black and White Educational Cumulative Distributions (MSE)</td>
<td>.256</td>
<td>.133</td>
<td>.047</td>
<td>.041</td>
</tr>
</tbody>
</table>

The cumulative distribution of education similarly exhibits a strong inflationary trend. The proportion earning higher degrees has increased steadily in both racial locations. This is clear in Figure 15 which plots the cumulative distribution of education by race for 1940-2000. Across each plot is it clear that the slopes get flatter over time, so that the proportion earning a HS degree or more (for example) has increased. In every other respect, the attainment trends offer an inverted picture to the locational results. Here, there is a large, initial difference by race that dissipates over time—as opposed to a small difference that increases. In 1940, the MSE between educational distributions was .25, nearly 6 times the locational value. This initial difference between distributions decreases from .25 to .133 to .046 from 1940 to 1980. Or, looking at Figure 15, the Black
and White lines converge over time. This is not to say that Blacks and White have the same level of attainment, but the differences clearly decrease over time (Kao and Thompson 2003).

Figure 15: Cumulative Distribution of Education by Race

There is then no simple relationship between educational attainment, educational locations and race. The 1940 results are particularly telling. The racial locations have very similar social meanings of education but very different levels of
attainment. Similarly, the attainment levels have grown closer over time, while the educational locations have not. The racial system in 1940 may be quite different than the system in 2000.

5.7.2 Over Time Results

I explore the racial systems more formally in Table 4. Here I predict which educational category in Time 1 is equivalent to which educational category in Time 2, by racial location. There are two regressions for each racial location and year, one using the within racial location predictors and one using the predictors from the other racial location.

Table 4: Logistic Regression Results for Educational Change by Racial Group

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>175.156</td>
<td>226.627</td>
<td>277.523</td>
</tr>
<tr>
<td>Area Under ROC Curve</td>
<td>.984</td>
<td>.920</td>
<td>.917</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>218.705</td>
<td>255.705</td>
<td>276.443</td>
</tr>
<tr>
<td>Area Under ROC Curve</td>
<td>.942</td>
<td>.947</td>
<td>.969</td>
</tr>
</tbody>
</table>
I use fit statistics as a means of testing the theoretical conditions corresponding to each racial system. For each regression, I have included the log-likelihood and BIC. The coefficients themselves are not of interest and are not reported in the table. It is only necessary to know the relative fit of each model for each racial location. Note that these models are non-nested models. They have the same number of parameters but the independent variables are different. We can use BIC as a general means of comparing the models. If the BIC values are within 2 of each other the models are considered substantively the same (Raftery 1995).

For ease of reference, I have also plotted the likelihoods for each year and regression in Figure 16. The bars represent the fit of the model predicting categorical equivalents, first using minority and then majority changes in attainment. The black bars correspond to the models with minority attainment changes, while the white bars correspond to the models with majority attainment. The top of the figure includes the expected pattern of results based on each possible system of stratification. For example, in a hegemonic world, it is better to use majority attainment in both the majority and minority groups to predict categories over time. This is reflected in the white bars being longer than the black bars in both the majority and minority case.
Starting with the 1940-1960 results, it is clear that changes in majority educational attainment are better predictors of educational equivalents than minority attainment changes. This is the case within both the majority and minority racial locations. For the minority location, BIC is lower (or the likelihood is higher) when using majority educational attainment change than minority educational attainment change. Or, in Figure 16, the likelihoods are larger using majority attainment, matching the expected pattern under the hegemony hypothesis.

The results for 1960-1980 and 1980-2000 parallel the 1940-1960 results. Majority attainment changes are still the dominant predictor of the educational categories over
time (see BIC in Columns 4-7). The predictions have, however, converged for the minority racial location, where the differences between the minority attainment and majority attainment regressions are much closer than in 1940-1960. In Figure 16, the white and black bars are much closer together in 1980-2000 than 1940-1960. The overall picture remains the same, however: changes in minority attainment predict which educational category is equivalent to which educational category in the minority racial location better than minority attainment changes.

The results suggest that the meaning of education in the minority location is best predicted by attainment changes happening in another racial location. Any baseline expectation that the best predictor would be the within location attainment change is not supported by the data. Or, returning to our theoretical conditions, it is clear that the other racial systems do not map well onto the empirical results. There is little evidence for minority competition or systems of homogeneity. A homogenous system does not fit because there are clear differences across the regressions. In Figure 16, the black and white bars indicate different fits using Black/White attainment. It is clear that a separate worlds hypothesis does not fit as the minority attainment ($X_{\text{min}_i}$) does not predict the minority categories ($Y_{\text{min}_i}$) better than majority attainment ($Y_{\text{maj}_i}$). Finally, it is clear that a system of minority threat does not fit the data: the patterning of fits statistics implied by a system of minority threat (see Figure 16) does not match the observed models. The results thus suggest that the fluid, contingent meaning of education is
determined by attainment changes happening amongst the majority. What is “good” or “normal” education is not set by those around you (if you are in the minority racial location) but by the culturally and institutionally dominant group.

5.8 Conclusion

This paper uses a new analytical approach to characterize racial stratification in the US. I measure racial stratification by describing who sets the norms and expectations in society. The approach draws on the idea that demographic categories are fluid and socially constructed. The variable nature of the demographic categories makes it possible to see who is driving meaning across the population. The key to approach is having: a) a clear measure of the meaning of demographic categories across time; and b) a clear model describing what underlying forces affect the (relational) meaning of the demographic categories. Thus, the question is whether the experience, or changing conditions, among one group is what matters for shaping meaning for the rest of the population.

Substantively, I ask how the educational attainment in one racial group affects the meaning of education in another. I laid out four possible systems: homogeneity, separate worlds, hegemony, and minority threat. Using census data, I found that a system of hegemony fits best, although there were some signs of convergence over time. By 2000, there was little difference between black and white predictions of the meaning of white educational categories. Theoretically, the results suggest two asymmetries are
important in the racial stratification system: first, the status and institutional power
differences across racial groups; and second, the sheer numerical differences between
majority and minority groups.

First, differences in status, reproduced and legitimized through institutional
power, make it likely/possible for the minority racial groups to take on the values and
norms of the higher status, majority racial group. Thus, the level of attainment in the
majority serves as the normative level of “good” education throughout the population.
In a similar way, those in minority may be sorted into organizations and occupations
based on the educational attainment of the majority. This has clear implications for how
those in the minority form relationships along education lines. Those in the minority
form relationships based on the distinctions implied by the majority level of attainment.
So, for example, those with high education in the minority are unlikely to meet and
marry those with <HS school, because <HS is “low” education in the racial majority,
even if a larger number of people in that (minority) racial location earn it. There is then a
disconnect between the social distinctions mapped onto the educational categories and
the educational attainment of that racial group.

Second, part of the majority influence resides in the power of large numbers.
There is a necessary asymmetry in majority/minority relations: a minority member is
more likely to know a majority member than a majority member is to know a minority
member (Blau 1977; Blau and Schwartz 1984). The minority is thus much more likely to
be aware of majority changes than the majority is of the minority; here in terms of educational attainment changes. This does not mean that such changes would necessarily drive educational meaning in the minority location, but it does set the stage for majority influence. Given the economic and political disparities across racial locations, these numerical asymmetries do, in fact, translate into a system of hegemony—where what is “good” is defined by those in power.

The convergence in majority/minority predictions suggests that the system may be growing less hegemonic over time. Substantively, the majority racial group represents a smaller portion of the population while the status differentials by race have decreased. For example, the correlation between race and income, education and occupation is weaker than in the early part of the 20th century. The norms of the majority, while still generally dominant, may be less obviously the standard than in the past. It is perhaps telling that the White, majority group itself seems to respond more strongly to shifts in attainment happening amongst other racial groups over time. Looking forward, it is possible that that the racial system will eventually reflect a homogenous world, or perhaps more likely, a system of separate worlds.

This paper focused on racial stratification, but the approach offers a general way of capturing power relations at an aggregate level. Given the right type of data, we can ask who sets the normative expectations in society. The framework thus provides the basic metric to differentiate between systems of stratification. One simply needs a
relational measure of the demographic categories and a set of conditions that drive those
categories. For example, an analysis of religion as the organizing dimension (rather than
race) may yield very different results as religion is less bound up in legal institutions.
We may then expect to see a system of separate worlds, or perhaps a homogenous
system, where everything moves in parallel over time. It is thus possible that race looks
increasingly like religion over time.

More theoretically, the approach serves as a formal means of testing theories of
hegemony and cultural dominance. By mapping specific, empirically testable conditions
onto different theories of stratification, it is possible to see if hegemony, for example, fits
a society and whether this has changed across time. Crucially, the approach allows a
researcher to compare systems of stratification across time and context. This
theoretically shifts the onus of sociological explanation: from the outcomes of
individuals to the large scale relations between demographic groups.
6. Conclusion

There are three broad goals to this dissertation: first, to develop a methodological framework that measures the fluidity of demographic categories; second, to interpret macro level forces through the lens of demographic, categorical change; and third, to contribute substantively to debates about racial stratification, educational change and the maintenance of power/inequality in America. In this conclusion, I broadly describe how each goal was accomplished. I also discuss some unresolved technical and substantive issues. I end the conclusion by discussing some future projects that will push this work in new directions.

6.1 Measuring the Fluidity of Demographic Categories

Chapter 2 described the methodological details of the dissertation. I began with the argument that demographic categories are fluid, non-essential social outcomes (Bates and Peackock 1989; Loveman 1999; Anthias 2001). I then developed a framework that takes this idea seriously when measuring social change. The measure of interest becomes the categories themselves, rather than the changing fortunes of fixed demographic categories. The model defines demographic categories using social distance as the underlying metric; I measure social distance through interaction rates (Laumann 1969; Bottero and Prandy 2003). Demographic categories are defined by their pattern of marriage, friendship, etc., and categories with the same interaction patterns are considered equivalent across time, even if the labels are different.
Formally, I combine models of social space and social distance to measure the categories across time (McPherson 1983; Levine 1993; Hoff et al. 2002). I place categories into social locations based on observed rates of interaction. Social locations represent the range of possible interactions for individuals with a given set of demographic characteristics (Bottero and Prandy 2003; McPherson 2004). The model determines which categories occupy which social locations and are thus relationally equivalent across time.

The larger goal of Chapter 2 is to take the radical sociological notion that categories are fluid and make it part of everyday research. I highlight two ways that could happen. First, it could become common practice to ask which category at Time 1 is relationally equivalent to which category at Time 2. This type of analysis would likely be descriptive, and serve as an initial exploration of the data. The analysis would tell a researcher how their basic demographic units have changed across time. This would not only serve as a useful way to understand a social setting, it could also shape the comparisons used in the analysis.

Pushing this a bit further, it is also possible to ask how income, health, wealth, and other outcomes of interest map on to social locations over time. The units of comparison would shift from fixed categories to social locations, but one would still maintain a substantive interest in health (for example). Substantively, a researcher could differentiate between categories moving over social locations and outcomes moving over demographic categories. It may be the case that education is now more highly
correlated with good health, but this is because lower educational categories occupy lower social locations, and not because the “returns” to knowledge/skill increased (McPherson and Smith 2009). It is also possible that health outcomes map onto social locations differently over time. Differentiating between these possibilities would open up new avenues for research, and bring current findings into sharper focus.

6.1.1 Remaining Methodological Issues

Despite the promise of the approach, there are a number of methodological issues that must be resolved before it can be applied as a general means of analysis. For example, the analyses presented in Chapters 2-5 sidestepped the problem of uncertainty due to sampling error. Sampling error is small with the US Census (as the sample is quite large) but this may not be the case in other analyses. It is thus important to consider sampling error when testing hypotheses about the changing relational meaning of demographic categories. The suggestion is that one can use a bootstrap approach to capture all of the sources of uncertainty (see also Smith et al. 2013). One would take samples from the original sample, rerun the analysis (including calculating the social locations) and summarize the results over the samples. This would incorporate uncertainty in the underlying social locations, as well as any uncertainty due to sampling.

In the future, it is also important to consider other types of social ties when defining the social locations. Here, I focused entirely on marriage and cohabitating
relationships. This is a particularly strong social relationship (DiPrete et al. 2011; Smith et al. 2013). It may be useful to reconsider the results using a weaker relationship, although it is unlikely that the main conclusion would change (as the basic patterning is the same across different types of ties-Blackwell and Licther 2004). Additionally, I took the marriage/cohabitation data and analyzed the contingency tables as if they were symmetric. This simplifying assumption was appropriate for this analysis but future work could explore this in more depth. For example, there may be strong asymmetries by education and gender (Hou and Myles 2008). In the past, high education males married lower education females (Schwartz and Mare 2005). Such asymmetries are ignored in the current analysis but could offer important insights into the changing meaning of the demographic categories. Methodologically, the key is estimating educational locations by gender (for example), so that Male-HS and Female-HS would occupy similar but distinct locations in the space.

In a similar manner, the current analysis does not fully exploit the multidimensional nature of social space. Multiple demographic dimensions may structure interaction (Blau 1977; Kalmijn and Vermunt 2005; Smith et al. 2013). These dimensions may not act independently but rather jointly create barriers to social ties. The analyses here used 1 dimension at a time, conditioned on another. For example, the analysis in Chapter 2 had 1 dimension for education but limited the analysis to couples under the age of 40. I could, alternatively, have looked at the intersection of age and
education, using age/education as the dimensions of social space (McPherson 1983; Mark 1998). This would have simultaneously captured social distance along age and education lines, as well as any interactions/non-linearities occurring in the space. It would thus recognize that homophily on one dimension is dependent on the properties of another (Kalmijn and Vermunt 2005; Smith et al. 2013). See the example social space discussed in Chapter 2.

Future work could also explore the assumptions of the model more thoroughly. For example, in the Hoff model distance is based on a linear Euclidean logic (Hoff et al. 2002). More nuanced work could make the functional form of distance a substantive question in its own right (Levine, Klein and Matthews 2001). It is possible that the distance calculation in a Euclidean space should really be raised to a power (now implicitly set to 1). And more generally, we can ask whether Euclidean distance, city block or something more complicated (Minkowski distance, for example) best captures the relational patterning in the space (Levine 1993; Levine et al. 2001).

6.2 Demographic Categories, Social Locations and Macro Level Change

The second goal of the dissertation is to connect categorical change to the larger changes in society. I offered three different ways that categorical change could be used to understand macro level properties. First, as in Chapters 2 and 3, I use changes in relational meaning to directly interpret demographic changes. In Chapter 2, I asked how changes in educational attainment affected the social locations of education. In Chapter
3, I asked how the meaning of Hispanic changed as the number of people identifying as Other-Hispanic increased. In both cases, I interpreted demographic changes through the changing social locations of the demographic categories. Here, a theory of social change is one that connects demographic change to the changing meaning, and thus social positions, of the demographic categories.

The second approach focuses less on the changing locations of the demographic categories and more on the locations of the categories in a given time period. In Chapter 4, I defined mixed race categories by their social location and the locations occupied by the parent categories. I thus developed a classification scheme where mixed race categories are defined as types of structural positions. These structural types are based on the relative distance between the mixed race categories and the parent categories. Over time, one could ask if the mixed race categories fall into different structural types, but this is still very different than asking if the categories occupy the same location across time.

The third approach uses demographic/locational changes in one demographic dimension to understand stratification in another. This is the approach developed in Chapter 5, which examines educational changes by racial groups. Chapter 2 mapped changes in educational attainment to the locations of the educational categories. Here, the shifts in education are part of a larger theoretical framework. I ask how educational attainment in different racial groups affects the meaning of education across the
population. Thus, the demographic/locational shifts in education are used to capture the larger features of the racial stratification system.

6.3 Substantive Conclusions about Education and Race/Ethnicity in the US

The dissertation substantively contributes to the literature on race and education in America. Chapter 2 examines the relational implications of increasing educational attainment (Goyette 2008). There are clear signs of inflationary pressure. Educational categories shifted to lower social locations as the level of attainment in the population increased. High School in 1940 is the social equivalent of College in 2000. This generally supports theories of credentialing, where individuals attain increasingly higher levels of education as other groups increase their level of education (Collins 1971; 1979). HS used to be sufficient to attain a privileged position in society, but this is no longer the case (Mare 1995; Snyder, Tan and Hoffman 2006). Such dynamics are reflected in the social distinctions mapped on to the educational categories.

Chapters 3 and 4 looked at two questions about race/ethnicity. Chapter 3 examined the changing meaning of Hispanic in the population. Over time, there has been a large increase in the number of people identifying as Other racially and Hispanic ethnically (Rodriguez 2000; Tafoya 2005; Frank et al. 2010). I interpret this change through shifts in social locations. I find that the increase in Other-Hispanic corresponds to a growing divide in the Hispanic population—between those that see Hispanic as a distinct racial category and those that do not. This finding adds to the growing body of
work trying to place Hispanic in the racial landscape (Campbell and Rogalin 2006; Gomez 2007; Lee and Bean 2004; Frank et al. 2010).

Chapter 4 explored the meaning of mixed race categories. Starting with the 2000 Census, individuals were allowed to identify with more than one racial group (Hirschman, Alba and Farley 2000; Farley 2002; Lee and Bean 2007). I analyzed these “new” racial categories by asking how they fit into the old racial order. I find that non-Hispanic mixed race categories (like White-Black) fit easily into the racial system—falling somewhere, relationally, between the parent categories (e.g. White, Black). In contrast, Hispanic mixed race categories (e.g. White-Black-Hispanic) are more likely to occupy distinct social locations. Hispanic thus plays an important role in the creation of new, not yet recognized racial/ethnic categories (in a way that mixing traditional racial groups does not). The current analysis was primarily descriptive, but future work could explain why some categories fall easily into the racial relational system, while others do not.

Finally, Chapter 5 tested theories of racial stratification using changes in education. I ask how educational attainment and educational locations vary across racial groups. I find that the system is characterized by racial hegemony, where the normative level of education is set by the White, racial majority. Minority racial/ethnic groups form relationships based on this White standard. This hegemonic tendency has, however, dissipated as the absolute level of inequality has decreased between racial groups.
Moving forward, I will consider other (non-White, non-Black) racial/ethnic minorities when testing theories of racial stratification.

**6.4 Future Work**

The dissertation represents a first step in developing a general analytical framework, and there are a number of ways this work could be extended in the future. I quickly highlight three example projects that draw on the ideas presented here. Some push the analysis methodologically and theoretically, while others apply the framework to new substantive settings.

The first project is primarily methodological and compares different ways of measuring social distance from relational data. I will compare log-linear distance models to more traditional network approaches. In the dissertation, social distance is inferred from the frequency of interaction between demographic categories. Categories are placed into locations in social space based on their pattern of interactions (Bottero and Prandy 2003). Social distance is then the distance between occupied locations in the space (Krivistky et al. 2009). In contrast, a traditional network approach defines distance as the path lengths between nodes in an observed network (Wasserman and Faust 1994). Demographic groups that are socially distant will have high path lengths between their members (on average). Substantively, a network approach more explicitly accounts for network topology when calculating distance. The question is whether these different approaches offer distinct information about the social structure of a population. I would
start with typical ego network data as inputs to both models. I would use the ego-alter
demographic information and log-linear distance models to calculate social distance. I
would then calculate network-based distance using the simulation approach in Smith
(2012). The simulation approach generates a full network consistent with the local
information in the ego networks (including homophily, degree distribution, differential
degree and local clustering). After generating the network, I will calculate the average
network distance between demographic groups. In the end, I will compare the two
approaches, asking how our conclusions about social distance vary across the two
methods (Beshers and Laumann 1967)

The second project uses the methods of the dissertation to answer a substantive
question about period/cohort effects and homogamy. A typical study of remarriage will
focus on the characteristics of those who divorce and remarry later in life (Schwartz and
Mare 2012). They will then ask how such characteristics affect homogamy rates. Are
second marriages more or less homogenous than first marriages? (Dean and Gurak 1978;
Jacobs and Furstenberg 1986). This approach implicitly ignores the fact that individuals
are forming relationships in another time period. It is not simply that different types of
people divorce and get remarried. Rather, the very act of remarriage occurs later than
the original marriages of their cohort. I leverage this disconnect in timing to
simultaneously compare marriage patterns across time and cohorts. Specifically, I will
calculate the social locations of educational categories (like Chapter 2) for one cohort in
the present. I will then compare those locations to the locations of the same cohort in the past. I will also compare the current educational locations to the locations of a younger cohort in the present. Do older individuals who form relationships now resemble the period, what the younger cohort is doing, or the cohort, what their cohort did in the past? Empirically, I will use Census data from 1980-2010 to compare across cohorts and periods.

The third example project will build on the mixed race category analysis in Chapter 4. The current analysis defines the mixed race categories from three pieces of information: the distances to the parent categories and the distance between the parent categories. I use these distances to empirically place the mixed race categories into ideal structural types. In the future, I will use this classification scheme to predict changes in racial identity. There is a growing literature interested in individual level changes in race/ethnicity (Lieberson 1985; Harris and Sim 2002; Penner and Saperstein 2008). Penner and Saperstein (2008), for example, found that individuals often switched racial classification and that these changes were associated with changes in social standing. An individual is more likely to identify as Black if they lose a job, go on welfare, etc. The goal here is to use the classification of the mixed race categories to think through the process of racial switching. Different structural types should yield different probabilities of moving between parent categories. For example, if White-Black is collapsed into White, then we should see few switches between White and Black (as mixed race
individuals are firmly White). If White-Black is relationally between White and Black, then we should see more back and forth identification (as mixed race individuals are both White and Black socially).

Overall, the dissertation offers an analytical framework that will be useful in many substantive settings. Future applications may be straightforward (how has the relational meaning of Republican/Democrat changed?) or require additional theoretical work (i.e. how can we use the mixed race classification to predict changes in racial identity?). What will connect these projects is not a particular substantive problem or even a specific theory, but rather a way of thinking through empirical evidence. The overarching assumption is that methods and theory should be isomorphic (White 2000). If we believe that categories are fluid and socially constructed, then demographic categories should be allowed to change meaning over time. Such a matching of methods to theory opens up new avenues for empirical research. In the analysis here, I explained demographic change and the maintenance of macro level power. Both analyses relied on the model of categorical change specified earlier—but this was not developed to test theories of demographic change or power per se. Future applications will hopefully yield unexpected substantive and theoretical insights. The larger hope is that this work will spur researchers to think formally, not only about the meaning of demographic categories, but about all of our most basic sociological assumptions.
Appendix A

This appendix presents Figure 17 in Chapter 2. It plots the log odds of a tie between educational categories the same distance apart.

Each figure plots the log odds of a marriage relative to chance as a function of distance between relative locations. Values at 0 distance represent the diagonal. The plots include N*N points, where N is the number of occupied locations for that year.

Figure 17: Scatter Plot of Odds Ratios by Relative Distance 1940-2000
Appendix B

This appendix presents the racial and Hispanic question used in the 1980 Census.

See Chapter 3.

Race: “Fill the circle for the category with which the person most closely identifies”

O White
O Black or Negro
O Japanese
O Chinese
O Filipino
O Korean
O Vietnamese
O Indian (Amer.)

Print tribe ______________

O Asian Indian
O Hawaiian
O Guamanian
O Samoan
O Eskimo
O Aleut
O Other - Specify __________

Hispanic:
“Is this person of Spanish/Hispanic origin or descent?”

O No (not Spanish/Hispanic)
O Yes, Mexican, Mexican-Amer., Chicano
O Yes, Puerto Rican
O Yes, Cuban
O Yes, other Spanish/Hispanic

"A person is of Spanish/Hispanic origin or descent if the person identifies his or her ancestry with one of the listed groups, that is, Mexican, Puerto Rican, etc. Origin or descent (ancestry) may be viewed as the nationality group, the lineage, or country in which the person or the person's parents or ancestors were born."
References


Jencks, Christopher. 1990. "What is the True Rate of Social Mobility?" in *Social Mobility and Social Structure*, edited by R. L. Breiger. Cambridge: Press Syndicate of the University of Cambridge.


Logan, John 2003. "How Race Counts for Hispanic Americans." Lewis Mumford Center for Comparative and Urban Regional Research, University of Albany, SUNY.


Biography

Jeffrey Smith was born on January 12, 1983 in Wheat Ridge, Colorado. He received his B.S. in economics and English from the University of Wisconsin-Madison, graduating with distinction. He recently published an article on network sampling, “Macrostructure from Survey Data: Generating Whole Systems from Ego Networks”, in *Sociological Methodology*. An early version of this dissertation won the ASA Mathematical Sociology Section Outstanding Dissertation in Progress Award. He was previously awarded a James B. Duke fellowship as well as the Vorsanger-Smith Award for “overall excellence in the department’s graduate program”. He was also a PARISS (Program for Advanced Research in the Social Sciences) fellow last year. Next fall, Jeffrey will be an assistant professor in the Department of Sociology at the University of Nebraska-Lincoln.