Definitional Tension:

The Construction of Race In and Through Evaluation

A Dissertation

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Srividhya Shanker

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Jean A. King, Adviser

December 2019

ACKNOWLEDGMENTS

If you regard things as existent by virtue of their intrinsic reality,
you thereby regard them as bereft of causes and conditions.
And thereby you are condemning effects, causes, agents, actions,
activities, originations, cessations, and even fruitional goals…
Nothing whatsoever is found which is not relativistically originated....

Nagarjuna, Mūlamadhyamakakārikā, XXIV, 16-19, 150-250 CE

In the USA, we often hear that knowledge is power. In India, we often hear that knowledge confers humility. Power—whether wielded justly or unjustly—is what makes knowledge. The power of generations of family and community made this dissertation what it is, by making me who I am and making it possible for me to feel and think and act and write. I would like to acknowledge as many of them as I can. I seek forgiveness in advance for those I have inevitably forgotten or run out of time or room to name.

First, I would like to acknowledge Meena and Raman Shanker for engendering both interest in and access to a formal education that allowed me to look critically—thanks to the arts—and write critically. I would also like to thank you for the informal education at home that allowed me to think critically. From Mom I learned to question words (like "tolerance!") and gained a living intersectional analysis. I also got grounding in an ancestral value system that prioritizes the means relative to the ends. From Dad I learned the power of math and science in everyday life, that within-group differences are typically greater than between-group differences, about vicious and virtuous cycles, and about the fallacy of race and caste as biological categories.

I would like to acknowledge Priya Shanker for practicing critical literacy and for analyzing and acting against patriarchy and white supremacy with me since middle school. I thank you for expanding my understanding of oppression to include all the earth's beings. I also thank you for caring for Mom, Sarojini's dear Thathi, in ways that I could not.

I would like to acknowledge Vijay Shanker and Dee Martin for support and encouragement as well as engagement in legal reasoning, politics, feminism, and family

fun. I thank sweet Anjali Martin Shanker for being a great big sister to Sarojini and day-brightener to everyone who interacts with you.

I would like to acknowledge Josephine Mansaray for the role you have played in our lives, especially but not exclusively with Sarojini. I thank you for feeding Sarojini, dancing her to sleep, and teaching her first words. I also thank Jeneba, Doris, Michaela, Che, Bridget, Samara, and Zanita for being Sarojini's brothers and sisters. Along with Sahr and Tamba Brima, you all have broadened my understanding of family beyond capitalist, casteist, patriarchal definitions.

I would like to acknowledge Joseph and Talita Pokawa and Abu and Gladys Brima for allowing Hindolo to come into my life. Thank you for nurturing critical praxis in us both.

I would like to acknowledge the memory of Larry Olds for creating a path for Hindolo that allowed us to meet and for acting like a father to him and thus to us. I thank you for allowing me to engage in community with thoughtful people and to experience the participatory research conference in Vancouver.

I would like to acknowledge Hindolo Pokawa for living this work with me every day. I thank you for supporting and encouraging and stretching my integration of what we call mind, body, spirit; theory and practice; domestic and international. I have unlearned even more than I have learned from your visceral experience with—and analysis of—poverty, war, migration, and status as a black man in the USA, as well as the underbelly of international aid and NGO work. You showed me the regenerative power of being with the earth and the redemptive power of fulfilling our duty to family and community. You have made me a better person, organizer, researcher, and mother.

I would like to acknowledge Sarojini Tehyeh Pokawa—my bright-eyed and beautiful bodhisattva—for the healing that comes from fully focusing on the present without losing sight of the past or the future and the restoration that comes from immersing oneself in play. You showed me how much we can learn in just a day. You allowed me to strengthen my ability to exercise patience and compassion, and to cultivate interactions and relationships based on love and hope rather than power and control. Beyond eating frozen meals and listening to my presentations for the last year, you taught

me that critical literacy is an innate gift that only has to be nourished. You taught me that artificial differences like gender can be unlearned in just one generation, giving me the impetus to think even more deeply about prevailing messages regarding otherized groups. And you taught me to feel and name and respond to my feelings as well as those of others.

I would like to acknowledge Jean King for seeing something of value in me. I thank you for encouraging and supporting me intellectually, emotionally, and in some ways financially throughout the duration of my PhD—especially generous and courageous considering the topic and perspective relative to your own experience. Thank you also for building community and making everything family friendly so Sarojini always felt invited and not just accommodated.

I would like to acknowledge Rainbow Research for accommodating my writing as well as experimenting with, expanding, deepening, and critiquing my ideas. I thank Mary McEathron, Beki Saito, Katie Fogel, Iván López, Dane Verret, Al Kuenzli, and Lateesha Coleman for taking a risk in me, taking on my share of work, and taking Sarojini into your lives.

I would like to acknowledge the nearly exclusive access to literacy—among other even more basic rights—at the expense of the vast majority of Indians that my ancestors unjustly arranged, the oppression and material effects of which can be felt in my lifetime. I thank contemporary Dalit scholars and activists like Thenmozhi Soundararajan and those I have read through Velivada and Ambedkar King Study Circle for their work, which has helped me see and act with more clarity; I also thank Ambedkar himself for developing India's system of reservation, which facilitates the expression of such analysis. And I thank so many Scheduled Caste and Scheduled Tribe workers whose labor allows all work in India, and beyond, to happen. I specifically thank the drivers, watchmen, and domestic workers—most of whose names I never knew—who made my own experiences in India, and subsequent opportunities, possible. At the same time, I am grateful for my Brahminism because it has given me insight into how whiteness operates in relation to class and cis-hetero-patriarchy through opportunity and kinship structures. I

am learning to atone for, resist, and repair the harm, which is ongoing, during the second half of my life.

I would like to acknowledge the stewardship, displacement, and dispossession of the indigenous peoples of this continent, to which many Brahmin families like my own migrated and on which we settled upon the institution of the system of reservation in India. I would like to acknowledge the African American traditions of resistance—through the arts, academia, and collective action—that gave us the civil rights that we enjoy as well as many of the tools of critique that we employ.

I would like to acknowledge Westerville Public Schools for providing me the childhood experience of racial otherness, exclusion, and violence during the era of busing. From it, I gained some—limited—insight into the experience of African Americans and indigenous folks and, more importantly, I developed a sense of kinship and solidarity with them.

I would like to acknowledge my schooling at CSG, which afforded me the training to read, think, and write critically; the art and art history training to see critically; as well as the experience of distinctly institutional rather than interpersonal racism in relation to the matrix that includes class, gender, and ability status. I thank the school for illustrating how increases in socio-economic class and educational level do not correlate with decreases in racial bias—or if anything, they do so inversely. I thank the Walker, the county, and Catholic Charities for reinforcing this lesson.

I would like to acknowledge the University of Michigan for exposing me to formal study of and engagement with the history of Asian American resistance movements and intersectional feminism. During this time, I read Gloria Anzaldúa, Cherríe Moraga, Chandra Mohanty, Mitsuye Yamada, and had a chance to meet Grace Lee Boggs. I thank you all for giving me tools to understand and communicate my experience and ideas, as well as those of others, in this dissertation and in my research and evaluation work. At the same time, I'm forced to acknowledge the reduction of the Asian American Studies department to a mere shell of what it was even when I was there, before it had reached its pinnacle.

I would like to acknowledge Deidre Hamlar for recognizing something of value in me and for her pivotal role in providing me the opportunity to move to the Twin Cities of Minnesota, where I am simultaneously surrounded by the largest urban indigenous population and the largest refugee populations. I thank both for helping me think more deeply through forced assimilation as well as alternative conceptualizations of justice and reparations. We are fortunate to live at the birthplace of the American Indian Movement and are especially grateful to the Dakota and Ojibwe peoples, whose arts and spirituality envelop us daily. I thank Pow Wow Grounds and the coops for allowing me to write there all day.

I would like to acknowledge Asian American Renaissance and the Twin Cities community of Asian American artists and activists. I thank Chamindika and Sumitra Wanduragala for nourishing me in multiple ways during my earliest years here; Marlina Gonzalez for showing me how an activist parents; Dipankar Mukherjee and Meena Natarajan for proving another (nonprofit) world is possible; Bao Phi for transparently engaging in individual and collective critical reflection as well as for your encouragement, analysis, validation, and refusal to accept simplistic answers. Mikari Suzuki, Sherry Lee, and Charissa Uemura—I have spent the last 15 years coming full circle in hopes of experiencing once again that spirit of creativity and camaraderie that we shared.

I would like to acknowledge Samuel Myers' scholarship and teaching. Thank you for being my sole source of encouragement and validation at the Humphrey and ever since.

I would like to acknowledge my experience at OAP, now Voices for Racial Justice, for the vicarious training in and subsequent experience of community organizing that I received. I thank Jackie Byers and Sarah Agaton Howes for the insights and validation you gave me at that time and for continuing to live your dreams and values.

I would like to acknowledge Joan Dejaeghere for inviting me to join that first trip, and subsequent trips, to Bangladesh and for initially thinking through with me how to incorporate my experience into my scholarship. I thank Andrea Smith not only for your scholarship and for the opportunities to train with INCITE!, but also for personally

calling me in response to my questions about using my research to address the imperialist current running through such trips as well as through AEA's Feminist Issues in Evaluation TIG. I similarly thank Bessa Whitmore for encouraging me to start my dissertation research from my own experience. I thank Donna Mertens and Karen Kirkhart for reading my first dissertation proposal, which was already aimed at language and honoring the experience of racially otherized evaluators. I thank Tamara Bertrand Jones for creating structured opportunities for novice evaluators like myself at the time to interact with senior evaluators representing racially otherized groups through the conference session format. I thank the MIE TIG for underwriting my attendance at AEA that year through its essay contest.

I would like to acknowledge several other scholars for their generosity with their time and knowledge: I thank Michele Goodwind for sending me your syllabus on intersectionality; Joe Soss for meeting with me and asking me what my dissertation story was about; William Darity for sending me copies of your articles that I had trouble finding; and Stuart Yeh for assisting me in articulating my fundamental problem statement in ways that the evaluation field could understand it.

I would like to acknowledge everyone who took a risk in mounting The Revolution Will Not Be Culturally Competent with me at AEA's 2012 annual conference: Dipankar Mukherjee, Rebecca Frost, David Nicholson, Sharon Day, Cindy Reich, Diana Yefanova, and Meena Natarajan. I thank Pamela Fletcher for analyzing the response to the play's reprise at the 2013 MESI, and for exposing me to concepts from critical race theory that provided me insight into the origins of scientific research and language as currency. I also thank Michael Quinn Patton for sharing the floor with me and encouraging me to fully develop and publish the play.

I would like to acknowledge assistance I received that directly touched the dissertation. I thank Laura Pejsa and Chris Moore for writing together, Julie Nielsen for sharing your dissertation and time with me, Eryn Lee and Alyssa Belcher for administrative support managing data for more than 5,000 articles. I thank Jeanne Zimmer for practice interviewing and Jasmine Tang for integrating writing support with critical analysis and therapy that effectively undoes decades of racialized, gendered,

classed, and ableized strains of imposter syndrome. I thank Betsy Barnum for being able to see both forest and trees as you edit. I thank Rabi Michael Crushshon for babysitting Sarojini over the summer and the entire family—Marie, Dion, and Naima—for welcoming her, Hindolo, SLFND, and me into your lives.

I would like to acknowledge Louis Alemayehu, Sam Grant, and Kim Kokett for supporting our family and for deepening our understanding of alternative structures and processes. I thank Alyssa Erickson, Marsha Williams, Kirsten Johnson/REASN, and Chris Opsal for maintaining friendships and creating community around children/families, organizing/activism, and our day jobs.

I would like to acknowledge my local group of nonprofit and evaluation critics—Leigh Combs, Ariella Tilsen, Sida Ly, Patricia Nelson, Antonia Apolinario-Wilcoxon, Cynthia Mathias—and my national and international group of critics: Lisa Jackson, Nicky Bowman-Farrell, Andrea Guerrero Guajardo, Aditi Sushama Ashok, Prachee Mukherjee, and Nan Wehipeihana. I thank you for both validating and stretching my thinking.

I would like to acknowledge the energy invested into making AEA and evaluation an emancipatory force for otherized groups everywhere. I thank Stafford Hood for your scholarship on African American evaluators who would otherwise be anonymous and invisible to evaluation students; Joan LaFrance for advancing the IPE TIG, which has expanded evaluators' thinking about indigeneity and evaluation—certainly my own; Bob Covert for thinking critically about advancing representation through the MIE TIG and for consistently encouraging me to Phinish and be Done; Ricardo Millett for putting money where your mouth was twenty years ago and for continuing to engage with me around systemic oppression; Kien Lee/Community Science and Geri Lynn Peaks/Two Gems Consulting—not only for making recommendations that were apparently ahead of their time, but also for your responsiveness to all my questions; Anna Madison and Hazel Symonette for decades of work at multiple levels, from which I and many others have benefited. I hope my research begins the process of documenting all your work.

I would like to acknowledge my committee members and back-ups. I thank Peter Demerath for seeing the proposal I made in his class through to fruition ten years later,

Muhammad Khalifa for first meeting around CRT and indigenous methodologies, and Tania Mitchell for trying to work around my delays. I thank John LaVelle and Jeremy Hernandez for ensuring I finish one way or the other in the event of anyone's absence.

Finally, I thank Rodney Hopson for all of the above: thank you for working at the systems level while providing individual-level support to so many of us.

This research would not be what it is without the power of each of you behind it.

DEDICATION

To Honeybee and Kunj.

ABSTRACT

Despite the centrality of racialized difference to evaluation, the field has yet to develop a body of literature or guidelines for practice that advance understanding of difference and inequality, including its own role therein. The purpose of this study was to broaden understanding of observed differences and inequality in evaluation beyond individuals and individual lifetimes. Drawing from critical theories of systemic oppression and system dynamics, it used a discourse-historical approach to answer three questions: How has the U.S. scholarly evaluation literature constructed racialized difference? How has that construction changed since the field began formalizing? How is that trajectory related to surrounding systems? Results showed four discursive patterns: (1) minoritization and ambivalence toward whiteness; (2) the invocation of diversity and inclusion; (3) the replacement of race with culture; and (4) the rise of and decoupled relationship between indigeneity and colonization. All four patterns were tied to meso-level dynamics. In the second two, existing recruitment and training efforts initiated and led by and for evaluators representing racially otherized groups at lower levels of the American Evaluation Association were elevated to the association's board-level, where leadership and language were broadened to represent dimensions of difference beyond race. Analysis of archival documents and interviews tied this meso-level pivot away from race to macro-level discourse and policies associated with racialized neoliberalization, which attributes inequality to individual as opposed to structural deficits. Unlike "Equal Opportunity" or "Affirmative Action," "diversity," "inclusion," and "culture" depoliticize difference and privatize the responsibility for—and benefits of—desegregation. In fourth pattern, literature that authors who identified as indigenous published, which explicitly complicated the relationship between indigeneity and colonization, increased sharply and remained higher following the organizing efforts led by evaluation scholars and practitioners who identify as indigenous. Their efforts remained in their hands rather than being elevated or broadened. Variation among the patterns suggests that the American Evaluation Association's relations with its racially otherized members and with educational institutions, large firms, philanthropy, and government are linked to the field's construction of racialized difference through existing institutional mechanisms. Whether the mechanisms counteract or amplify racialized neoliberalization depends on whether they circulate capital in ways that enable otherized groups to exercise collective agency and produce knowledge for structural change.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AEA	American Evaluation Association
APC	Advocacy and Policy Change
BDI	Building Diversity Initiative
BIPOC	Black, Indigenous, People of Color
CDA	Critical Discourse Analysis
CEO	Chief Executive Officer
CGO	Chief Grantmaking Officer
CREA	Culturally Responsive Evaluation & Assessment
CRT	Critical Race Theory
CFASR	Council for Applied Social Research
CST	Critical Systems Thinking
ENet	Evaluation Network
ERS	Evaluation Research Society
GEDI	Graduate Education Diversity Internship
IEF	Indigenous Evaluation Framework
INGO	International Nongovernmental Organization
IPE	Indigenous Peoples in Evaluation
ISEE	International Society for Evaluation Education
JCSEE	Joint Committee on Standards for Educational Evaluation
LaRED	Latinx Responsive Evaluation Discourse
LGBTQ	Lesbian, Gay, Bisexual, Trans, Queer
MIE	Minority Issues in Evaluation/Multi-ethnic Issues in Evaluation
NGO	Nongovernmental Organization
OECD	Organization for Economic Co-operation and Development
TIG	Topical Interest Group
UK	United Kingdom
UN	United Nations
U.S.	United States
USA	United States of America

PROLOGUE

Early in my doctoral program in Evaluation Studies, before persecution of the Rohingya people put Bangladesh on U.S. news channels, I took an opportunity to serve as part of a team of U.S. citizens from the University of Minnesota contracted to evaluate the girls' education program of a leading international nongovernmental organization (INGO) working in Bangladesh. Aside from me, the team consisted of one professor of European ancestry not much older than I, but farther along in an academic career that included considerable international research related to girls' education; she identifies as a White woman. It also included another student of European ancestry somewhat younger than I and one year behind me in the pursuit of his doctorate in Comparative and International Development Education; he identifies as a White man. As a U.S.-born student of South Asian ancestry who has invested several years working with others who identify as women and as Black, indigenous, and of color (BIPOC) toward gender justice, both domestically and internationally, I was eager to learn from and contribute to this experience. Also eager, my fellow student was keenly aware that he had never traveled outside North America and was not especially familiar with issues of gender, let alone in the South Asian context.

After a two-day flight, we landed in the capital city as the sun rose. Following a quick shower and hour-long commute, we arrived at the INGO's country office to meet the manager of its empowerment division that housed the girls' education program. My teammates and I seated ourselves around a large rectangular table in the only air-conditioned room of the office—the professor and I on the side closest to the door and the other student on the opposite side.

When the empowerment division manager—a Bangladeshi woman not much older than our professor—entered through the door, we all stood. She immediately shook our professor's hand, walked to the other side of the room to shake the other student's hand, and then walked down to the far end of the table to seat herself. She never shook

hands with me—the person in the room who looked the most like her—nor did she acknowledge my presence in the room.

Why?

I never asked her.

I have shared this story several times, however, and nearly all listeners characterize the behavior of the empowerment division manager as an interpersonal snub. Some attribute it to self-hate or internalized oppression on her part. From listeners racialized as White—to whom the story is shocking—and from listeners who are racially otherized—to whom it is painfully familiar—the story typically elicits a reaction focused on the division manager as an individual and her treatment of me as an individual. Others suggest that our professor should have intervened, as part of her institutional role, to make me visible to the empowerment division manager. Not surprisingly, no one's reaction has focused attention on the larger system within which we were all operating.

Expanding the Boundaries of Analysis

How can an understanding of the dynamics that may have been taking place *within* the empowerment division manager and *between* her and myself as individuals shift if we were to expand the boundaries of our unit of analysis by posing questions about the institution surrounding us? For example, what does the organizational structure look like? Who are the super-ordinates and sub-ordinates of the empowerment division manager? Answers to these questions—increasingly recognized as pivotal to effective evaluation practice—provide more than contextual information about the power dynamics among individual or institutional players in this program and its evaluation specifically. The answers are also important because they point to the tension between individual agency and structural conditions underlying persistent inequality.

Structurally Mediated Relations

The structure of the INGO operating in Bangladesh was hierarchical, and the hierarchy was racialized as follows. The manager of the empowerment division reported to the assistant director of the Bangladesh country office—a woman of European ancestry and Australian nationality, racialized as White. That woman reported to the director of

the Bangladesh country office—a man of European ancestry and Australian nationality, racialized as White. In contrast, the members of the country office staff who reported to the division manager were nearly all of South Asian ancestry and Bangladeshi nationality—racialized not as White…perhaps as "native"—as were nearly all the program participants.

How can our understanding of the interaction that occurred between the empowerment division manager and myself further shift if we were to expand the boundaries of our unit of analysis still farther, by posing questions about structural arrangements not just *within* the institution surrounding her and me, but also *among* it and other institutions? For example, where is the larger INGO headquartered? Who governs its decisions? Who funds it? Beyond Bangladeshis, who participates in its programs? Answers to these questions provide insight into what determines the way in which resources flow through the arrangement and how this affects performance.

In the interest of space, answers are provided for the INGO alone, although the same questions must be asked of the University of Minnesota and the conditions surrounding our team's contract with the INGO. The INGO is headquartered in the USA. Its decision making rests with a 13-member board representing Australia, Austria, Canada, Denmark, France, Germany, India, Japan, Netherlands, Norway, Thailand, the UK, and the USA. Eleven of these are advanced capitalist countries. Seven are former colonial powers in Western Europe, and three are settler-colonies dominated by descendants from Western Europe.[1] In addition to receiving funding from several multi-lateral agencies such as the European Union, World Bank, and UN, all of which are dominated by many of the same OECD countries represented on its board, the INGO receives funding from the governments of all eleven of the advanced capitalist countries on its board (Australia, Austria, Canada, Denmark, France, Germany, Japan, Netherlands, Norway, UK, and USA). Smaller funders include corporations, private foundations, and

[1] According to its website, the INGO was founded more than 50 years ago. Representation on the board grew to include non-European countries for the first time when members from Japan joined in 1987. The "third world" (non-aligned countries) was first represented when members from Thailand joined in 2003. More recently, affiliate members from India and Peru joined with no voting rights in 2011 and 2012, respectively. Since then, the members from India have become full members.

individuals who are disproportionately based in the same advanced capitalist countries, which utilize labor and resources from the countries receiving the INGO's programming. Decision making, governance, funding, and employment of labor are examples of the mechanisms that link actors within the structure of international development,[2] establishing and reproducing relations of super-ordination and sub-ordination by determining how resources such as funding and information flow among the actors.

In contrast to the countries represented on its governing board and among its funders, the countries receiving the INGO's programming and providing its direct-service labor are located in Africa, Asia, Central and South America, and Eastern Europe. Only two of the 87 countries in which it conducts programming have representatives with voting rights on its governing board—one has had them since 2003 and the second since 2013. *None of the countries in which it conducts programming is represented among its funders*. The disproportionate representation of former colonial powers in Western Europe and countries settler-colonized by Western Europe among the decision makers of the INGO and the nearly full representation of their former franchise colonies among the countries receiving programs map nearly perfectly onto the racially salient phenotypic differences among senior management, program staff, and program participants within the INGO itself, illustrating the racialized nature of the institution and larger system's stratification (Kothari, 2005, 2006a, 2006b, 2006c).

Structural racialization. Considering the structural arrangements within the INGO and its Bangladesh country office in particular, as well as those surrounding the INGO and field of international development, the Bangladeshi manager of the empowerment division had probably never received experts contracted and sent by headquarters—certainly from OECD countries—who looked anything like her. How might the racially hierarchical structure of reporting or feedback have affected the

---

[2] Kothari discussed the international development field's origins in the imperialist and colonialist efforts of Europe and the USA (Kothari, 2005, 2006c) in relation to its current discourse, which she described as highly racialized and yet, simultaneously, silent about race (Kothari, 2006a, 2006b). For example, the scholarly journal known as *Foreign Affairs* from 1922 until its demise in 2012 had previously been called the *Journal of International Relations* (from 1919 to 1922) and originally called the *Journal of Race Development* (from 1910-1919).

division manager's behavior toward me in this situation? In accordance with what is expected of rational economic actors, she was likely orienting herself toward her source of sustenance—people who appear to be from institutions and countries that control the distribution of and access to human and natural resources more generally. Understanding what I interpreted as her negation of me—the only other person of South Asian ancestry in the room—solely as an interpersonal dynamic rooted in internalized racism and thus a personal failing on her part alone diverts necessary attention from the structural arrangements and mechanisms that shaped those dynamics. Moreover, responding to the likelihood that she had never received experts who were not of European ancestry by simply recruiting evaluators such as myself—who are considered "diverse,"[3] who can claim some level of competence in the local culture, and who practice stakeholder-inclusive evaluation—leaves the racially stratified structural arrangements in which evaluators play a material and discursive role unexamined and therefore intact (House, 2017).

Structural racialization is the process through which race-based inequality reproduces itself, regardless of the intentions or behaviors of individual actors involved in any particular event (powell, 2013). The arrangement of actors within structures and the mechanisms by which they depend on and interact with each other shape the nature of the dynamics, the way that resources are exchanged or flow, among them. Hierarchical arrangements produce relations of super-ordination and sub-ordination. In situations in which the hierarchy or stratification is racially differentiated—or racialized—the nature of the super-ordination and sub-ordination is also racialized. Importantly, however, the racialized relations of super-ordination and sub-ordination and their subsequently

---

[3] Rather than its dictionary definition of "multiform, varied," or "with its varieties" (OED Online, 2016), "diverse" has come to connote those who are racially or linguistically *marked as differing from a norm that is encoded as White and English-speaking*. This is exemplified in the U.S. evaluation literature and other disciplines by recurring phrases such as "racially," "ethnically," "culturally," or "linguistically" "diverse" (e.g., Hood, 2001; Hopson, 2009; Maack & Upton, 2006; Schlueter, 2011; SenGupta, Hopson, & Thompson-Robinson, 2004), which are commonly used to refer to students, program participants, and evaluators who are racially otherized or who speak languages that are otherized rather than to an entire pool that consists of *assorted* racial, ethnic, or cultural backgrounds—one of which is called White—and *assorted* languages—one of which is English.

racialized effects, compounded over generations, also produce and normalize racial categories—including the equation of whiteness[4] with ownership, decision making, and other forms of power and the equation of racially otherized[5] groups with deficits in resources, knowledge, and agency (Doane, 2003).

Differences among the parent INGO's decision makers, its program administrators, and the participants in its programs are thus not merely lateral differences in racial classification, nationality, or culture (various combinations of which are often fused together [e.g., Mertens & Russon, 2000]) that can be addressed through increased diversity or cultural competence. Those with greater decision-making and economic power within the INGO's structure represent countries with greater decision-making and economic power than Bangladesh more generally, both historically—through colonization—and currently—through international agreements (Shiva, 2007). Furthermore, those with greater decision-making and economic power within the INGO's structure are disproportionately of European ancestry and racialized as White while those with less decision-making and economic power are disproportionately of South Asian ancestry and racialized not as White. As such, the nature of the stratification is racialized. Moreover, the racial stratification described of this particular INGO is so pervasive within the nongovernmental and nonprofit industries that everyone involved tends to perceive it as natural, without remembering that it is artificially (re)produced, and

---

[4] "The 'hidden nature' of whiteness is grounded in the dynamics of dominant group status…. [W]hites in the United States have used their political and cultural hegemony to shape the racial order and racial understandings of American society…. Historically, white-dominated racial understandings have generally focused upon the characteristics (i.e., 'differences') of subordinate groups rather than the nature of whiteness. This emphasis by whites upon the racial 'other' has gone hand in hand with the politically constructed role of whiteness as the 'unexamined center' of American society…. Consequently, in a discourse that focuses upon differences and the racialized 'other,' white becomes a default category…." (Doane, 2003, p. 7).

[5] Racially otherized groups are those whose observable phenotypes continue to be constructed and encoded through the language of race as "other" relative to those normatized as "White"…. [R]acial meaning is extended to "a previously racially unclassified relationship, social practice, or group…[n]ot because of any biologically based or essential difference among human beings…but because such sociohistorical practices as conquest and enslavement classified human bodies for purposes of domination—and because these same distinctions therefore became important for resistance to domination as well" (Omi & Winant, 2014, p. 13).

consider it "normal" rather than racist—a label that is reserved only for the most blatant and intentional acts of discrimination (Delgado & Stefancic, 2001).

Social Group Interests

Stratified differences in classifications of "race" and nationality are significant to the evaluation of nongovernmental and nonprofit organizations in that the nature of organizational actors' experience with existing structures (likely related to their structural location, which is demonstrated above as associated with their racial phenotype and nationality) potentially influences how their organizations understand situations, conceptualize programmatic solutions to problems, and evaluate those solutions. The understanding of poverty and inequality underlying the conceptualization of programmatic solutions can differ specifically in terms of their emphasis on individual decisions as opposed to policy-level decisions, which are important from the perspective of measuring their effectiveness. Actors also differ—to varying degrees resulting from their locations within the structure of the nonprofit and nongovernmental industry—in terms of the extent to which they are members of social groups that benefit from individualized as opposed to structural understandings of poverty and inequality as well as from the proposed solutions that stem from those understandings.

Depending on their emphasis, understandings of and solutions to poverty and inequality in Bangladesh can serve the interests of poor Bangladeshis who participate in the INGO's programs, middle class Bangladeshi NGO workers who administer the programs, and nationals of assorted socio-economic classes from advanced capitalist countries who plan and evaluate the programs very differently (Karim, 2008). Prevailing solutions, rooted in neoliberalism,[6] minimize government spending by deregulating trade, labor, and environmental policies and simultaneously privatizing resources and services to individuals through the above industries. Those experiencing poverty under deregulated conditions are treated as individual consumers in need of self-improvement (Karim, 2008; Soss, Fording, & Schram, 2011). Some argue that such solutions

[6] Neoliberalism extols the virtues of market forces as natural and fair, and emphasizes individual agency (Roberts & Mahtani, 2010).

ultimately serve corporate shareholders and, to a lesser extent, consumers in advanced capitalist countries (Karim, 2008; Shiva, 2007).

One widely accepted understanding of poverty and inequality in Bangladesh points to environmental destruction related to the deregulation, privatization, and consumption of natural elements in former colonial powers and settler colonies as well as exploitative labor conditions within the garment industry upon which the country has pinned many of its hopes for economic development (Khosla, 2009; Shiva, 2007). The INGO featured in this account receives funding from the retail clothing giant Walmart, which utilizes labor from Bangladeshi garment factories and has played an influential role in deregulating labor through international agreements, to run a program that builds what it calls "life skills" among individual garment factory workers. It does so instead of seeking funding to address factory workers' poverty and poor health at a structural level by working with several existing Bangladeshi people's movements organizing to increase regulation through international environmental, labor, and trade policies. Similar strategies by philanthropic foundations to placate particular social groups have been documented (Greene, Millett, & Hopson, 2004). For example, during the Civil Rights era in the USA, the Ford Foundation funded social services for individual African American residents of inner cities in what some consider a deliberate effort to steer them away from movement organizing directed at redistributive policy-level changes that would increase African American communities' opportunities for self-determination and instead cultivate compliance with capitalism, upon which it depended financially (Delgado, 2002; Greene et al., 2004 citing Stanfield, 1985).

Knowledge Production and Prevailing Narratives

Disproportionately from Europe, the USA, and other states settler-colonized by Europe, evaluators—including myself—benefit materially from and risk contributing to individual-level conceptualizations of and programmatic solutions to poverty. We benefit not only as producers of knowledge, from whatever fees, publications, presentations, and other perquisites that may be associated with contracts to evaluate such programs, but also as consumers, from the retention of unregulated markets that allow us to continue

over-consuming elements of nature by purchasing goods produced cheaply under exploitative conditions. Employed by the INGO sector, members of the Bangladeshi middle class who work as program administrators similarly benefit, albeit to a lesser degree, from individually-focused conceptualizations of and programmatic solutions to poverty (Karim, 2008).

Regardless of our geographic ancestry, knowledge about cultures, or inclusion of program participants; regardless of our explanation of the contextual factors contributing to participants' situations; and regardless even of the results of our evaluations, evaluators also legitimize the program theories underlying programs simply by accumulating evidence about them (Weiss, 1993). Moreover, we potentially reify the individually-focused conceptualization of poverty underlying them, leaving questions about racial stratification unasked and thus unanswered.

Counter-Narrative

Throughout this dissertation, the phrases "racialized as," "racialized difference," and "racially otherized" are intentionally used to convey the process of differentiation along a hodge-podge of attributes—phenotypic, cultural, geographic, genetic, legal, etc.—that have become conflated and naturalized. Racialized difference is inherently gendered, sexualized, classed, and coded with respect to ability status. Such phrases call attention to the production of difference and the dialectical, processual nature of differentiation.

Use of the word "industry" is also deliberate. According to the U.S. Bureau of Labor Statistics (n.d.), workplaces are classified into industries based on their principal activity, even if they individually and collectively include employees representing several different occupations. "Industry" refers to a group of workplaces or establishments that provide similar services or products. Social and human service providers[7] compete with each other for funding from individual donors and foundations as well as for market share of services they provide. They depend on philanthropic organizations and governments to

[7] Such organizations are distinguished from government in most of the world and from for-profit business in the United States.

finance an increasingly professional class of staff who may specialize in grant-writing and marketing and who may have no direct experience of the issue the organization was established to address. Philanthropic organizations depend on nonprofit organizations to disburse grant funding and governments to maintain their tax-exempt status. Governments increasingly depend on foundations and nonprofit organizations to address their citizenry's basic needs. Both utilize evaluation to hold programs accountable, make programmatic and funding decisions, and attract additional resources. As a network of actors that depend on each other's work and financial and informational resources, social and human service providers, philanthropic foundations, and evaluation service providers thus compose an industry closely tied to both the for-profit sector and government.

Being a presumably "diverse," culturally competent, and stakeholder-inclusive evaluator who was ultimately rendered "illegible" by those accustomed to seeing people who look like me—who look like themselves—as recipients of expertise and services rather than producers of knowledge and agents of change illustrates the limitations of a focus on individuals. My personal experience is situated within larger structures, including institutions within a nonprofit/nongovernmental industry that is predicated on economic inequalities that remain sharply racialized. Increased awareness of the racialized pattern in which information and other resources flow within these industries led me to focus this dissertation on examining the process of racialization in relation to evaluation, which plays a pivotal role in the flow of information and funding within and among the industries' institutions. In doing so, I draw from critical theories of systemic oppression—critical race theory, postcolonial theory, and feminist theory—all of which recognize experiential knowledge and advance the scholarly application of storytelling, particularly testimony and counter-narrative (Delgado & Stefancic, 2001; Lynn & Parker, 2006; Parker, 2004). Asian American critical race theory in particular "uses the power of narrative voice" and "stresses the importance of narrative and storytelling to use in a critical reading and tracing of the use of language and discourse" (Parker, 2004, p. 87). I offer this prologue in that spirit.

CHAPTER 1: INTRODUCTION

> When the racist chant began Wednesday night—"Send her back! Send her back!"—President Trump paused to let the white-supremacist anger he had stoked wash over him. George Wallace would have been so proud.
>
> That moment at a Trump campaign rally in North Carolina was the most chilling I've seen in American politics since the days of Wallace and the other die-hard segregationists. Egged on by the president of the United States, the crowd was calling for a duly elected member of Congress—Rep. Ilhan Omar (D-Minn.), a black [*sic*] woman born in Somalia—to be banished from the country because Trump disapproves of her views.
>
> This hideous display followed Trump's weekend call for Omar and three other House Democrats, all of them women of color, to "go back" to the "totally broken and crime infested places from which they came." All of this is an unmistakable echo of the racist taunts that used to be leveled at minority groups that had the temerity to demand civil rights and the gall to achieve political and economic success—go back to Africa, go back to Mexico, go back to China.
>
> After the election and reelection of the first African American president, one might have thought we were beyond such ugly, desperate racism. To the contrary, perhaps Barack Obama's tenure surfaced long-buried fear and loathing that made Trump's ascension possible.
>
> We can leave that for the political scientists to figure out in the fullness of time. Right now, we have an emergency to deal with…. (Robinson, 2019)

More than a half century has passed since video footage broadcast to the world Alabama Governor George Wallace's inaugural speech, during which he declared "segregation now, segregation tomorrow, and segregation forever" to a cheering crowd—many wearing white flowers that symbolized their commitment to white supremacy. Two generations later, mainstream and social media are once again awash with images of what many consider state-provoked violence against members of racially otherized[8] groups.

---

[8] While "people of color" is often distinguished from "indigenous peoples," in an attempt to recognize sovereignty under conditions of settler colonization, the phrase "racially otherized" in this

Despite their painful familiarity to many members of those groups, the images seem surprising and unbelievable to many Americans racialized as White, especially, who largely see the hateful rhetoric as coming out of nowhere by extremist individuals rather than as part of a larger structure and pattern of behavior. Mainstream reaction to presidential comments made earlier regarding a preference for immigration from Norway as opposed to "sh*t hole countries" reveals a lack of awareness that such comments describe what has essentially been U.S. policy since its inception as a settler colonial state. The nature of this difference in perception and awareness—a racialized difference not unlike the difference between "sh*t hole countries" and Norway—is the topic of this dissertation. Chapter One briefly reviews alternative constructions regarding the nature of difference, evaluation's acknowledgment of and ethical responses to difference, and the field's need to develop a critical body of literature that could provide decision-makers with the theoretical and analytical specificity, or racial literacy (Guinier, 2004), necessary to understand and disrupt the reproduction of racially differentiated outcomes.

Organization of Chapter One

This chapter begins by distinguishing prevailing, liberal constructions of difference from critical constructions of difference. The chapter then details evaluation's five ethical responses to difference (propriety, common good and equity, cultural competence, social betterment, and justice). Upon situating this dissertation research within critical theories of systemic oppression and as an extension of the field's justice-oriented response to difference, the problem statement, research purpose and questions, and scope and significance of the study. Definitions of key terms are presented at the end of Chapter One.

Context: Evaluation's Recognition of Difference

It is within the societal context described earlier—of racially differential, but not disconnected, experiences and perceptions—that government, philanthropic, and

---

dissertation includes groups that are indigenous to settler colonial states (including those colonized by countries in the Iberian Peninsula). This is not because such groups constitute a "race"—considering that race is an artificial category—but because such groups *are racialized* as "other."

nonprofit institutions develop and manage programs, enlist the services of program evaluators, and utilize evaluation results. The field of evaluation, to the extent that it is represented by the American Evaluation Association (AEA) and its two flagship journals,[9] has increasingly recognized difference. AEA's 2004 edition of the Guiding Principles provide the following examples of "difference," per se: "culture, religion, gender, disability, age, sexual orientation, and ethnicity" (AEA, 2004b). The most recent edition lists "race, ethnicity, religion, gender, income, status, health, ability, power, under-representation, and/or disenfranchisement" (AEA, 2018b).

The evaluation field's professional association and scholars have further recognized a difference—an incongruence—between program participants and program evaluators (e.g., Kirkhart, 1995; Kirkhart, 2010; Kirkhart & Hopson, 2008; LaFrance, Nichols, & Kirkhart, 2012; Lincoln, 1991; Madison, 1992b; Symonette, 2004). The incongruence is framed frequently in terms of culture (e.g., Chouinard & Cousins, 2009; Chouinard & Hopson, 2016; Hood, 2004; Hopson, 2001; Mertens, Farley, Madison, & Singleton, 1994). The professional association's documents describe culture as broadly including not just race and ethnicity, but also other dimensions of "diversity" (AEA, 2004c) including "gender, age, sexual orientation, social class, disability, language, and educational level or disciplinary background" (AEA, 2004a).

Constructions of difference. Prevailing constructions, rooted in liberalism, treat difference as pre-social, intrinsic, and lateral: Dimensions of difference are typically described as intrinsic to or attached to individuals or aggregates of individuals. These individual attributes are described as dimensions of identity around which social groups form and shared experience arises (Young, 2011). In contrast, critical constructions—rooted in critical theories of systemic oppression—treat difference as produced and reproduced over time and asymmetrical: Dimensions of difference are described as tied to systems of oppression—specifically, to structurally mediated socio-economic relations and processes. Critical theories of systemic oppression describe shared social group

[9] The *American Journal of Evaluation* and *New Directions for Evaluation*

interests and cultures as *constituting* rather than *deriving* or *arising from* individual identity.

Critical constructions of difference. In particular, critical theories of systemic oppression conceptualize perceived differences in race, gender, sexuality, ability status, educational level, and certain (but not all) other attributes as continually (re)produced through "the *networks of (interactional) relationships* among actors as well as the distributions of *socially meaningful characteristics* of actors and aggregates of actors" (Whitmeyer, 1994, p. 154, emphasis added). Moreover, because it is produced and reproduced cumulatively and interactively through historical and ongoing relations of super-ordination and sub-ordination—asymmetrically arranged exchanges of capital—difference according to critical theories of systemic oppression is stratified rather than lateral.

The juxtaposition of phenotypic characteristics like eye shape, hair texture, and skin color against eye color and hair color illustrates the production of difference. Differences in eye shape, hair texture, and skin color are attached to racial and cultural differences, whereas differences in eye color and hair color are typically not given the same type of meaning, nor does shared eye color or hair color suggest a shared lived experience. This is because socio-economic relations and decision making in the 21st-century USA are not organized by eye color or hair color in the way that they are organized by eye shape, hair texture, and skin color, among certain other phenotypic markers. (When German society was organized by eye color, eye color *was* associated with racial and cultural meaning [Suedfeld, Paterson, Soriano, & Zuvic, 2002].) The salience of these phenotypic characteristics as indicators of an important classification—namely race—was thus produced and continues to be reproduced through the ongoing stratification of groups in ways that are tied to the phenotypic characteristics, which have become considered markers of race.

At the same time, markers of race and racialized distinctions among social groups are not always phenotypically based (Bonilla-Silva, 1997). The ongoing condition of indigeneity under settler colonization and the legacy of colonization specifically by Spain

and Portugal has become racialized. “American Indian” or “Native American” and “Latinx” are each considered racial classifications despite the vast array of phenotypic (among other) variation among the peoples of the Americas. Despite phenotypic and other variation, peoples indigenous to the Americas are racially and ethnically differentiated according to whether they were colonized by France and Britain or Spain and Portugal. The difference among these groups was artificially produced, and is continually reproduced, by policy.

This illustrates the second point of contrast between liberal notions of identity that locate difference in individuals and those of critical theories of systemic oppression, which focus on relations among social groups. According to the latter, individuals are constituted by their identification with particular social groups (Young, 2011). The shared experience of sub-ordination and the shared experience of super-ordination each gives rise to specific sets of shared interests, around which groups form, with which individuals identify (Bonilla-Silva, 1997). Social groups form to protect differential sets of interests that arise from differential experiences and produce differential perspectives. It is individuals’ identification with particular social groups and not others that constitutes their identity. Shared experience under U.S. immigration policies ensures that those from preferred countries like Norway develop a corresponding set of shared interests and perspectives that necessarily differs from that of those from “sh*t hole countries.” In the industries surrounding evaluation, evaluators, program staff, organizational leaders, and funders overwhelmingly represent an overlapping set of interests and perspectives arising from the preferred status of their groups while program participants represent the interests and perspectives of excluded groups (House, 1990, 1995a). The field of evaluation has yet to tie the incongruence observed between individual program participants and individual program evaluators to the interests of their respective social groups.

Third, critical theories of systemic oppression conceptualize categories of difference as produced through a dialectical process, “that is, the creation of a category of ‘other’ involves the creation of a category of ‘same’” (Bonilla-Silva, 1997, p. 471). For example, in an effort to increase cultural competence in evaluation, SenGupta, Hopson,

and Thompson-Robinson refer to Lee (1997) as they contrast "East and West" (SenGupta et al., 2004, p. 7). Equating notions of geographic region, economic system, and time (all from the perspective of colonial Europe and capitalist progress), they distinguish "Eastern," "Agricultural," and "Traditional" from "Western," "Industrialized," and "Modern" (p. 7) in much the same way that indigenous savagery and extinction continue to be produced in contrast to European civilization and presence or Black servitude and labor continue to be produced in contrast to White management and ownership.

Fourth, critical theories recognize difference as the structurally mediated legacy of accumulated advantage and disadvantage. The processual nature of difference points not only to the importance of history, but also to the ongoing role that time plays in reproducing difference. Differences observed today are the distal, compounded effect of numerous interactions over time. Under such a construction of difference, identity is not static, but fluid—continuously re-constituting at both individual and collective levels. Not only do the social groups and associated power and privilege that individuals identify with—and with which they are associated by others—change over time and by situational context, but societal categories of difference and the meanings society attaches to such categories change as well.

Finally, liberal narratives tend to describe power and privilege, not just difference, as essential attributes attached to individuals. When acknowledged, they are generally inhered to phenotypic markers as opposed to causally tied to structural mechanisms of oppression. Such conceptualizations of power contrasts with those rooted in critical theories of systemic oppression, which consider the power that evaluation as a field wields within the industries immediately surrounding it, whose organization of labor and decision-making processes are stratified by race, class, gender, and ability status (Sturges, 2015). It contrasts further with conceptualizations of power that contextualize the ways in which the social groups associated with program participants, staff, and evaluators each benefit from that structure and the larger socio-economic structure, which is also stratified. Indeed, according to critical theories of systemic oppression, difference is continually (re)produced and identity is subsequently (re)constituted through the ongoing

asymmetry of channels through which groups can exercise power. The U.S. evaluation literature and the most recent AEA Guiding Principles acknowledge difference not just in terms of identity, but also in terms of power—specifically, the power that evaluators hold as individuals (AEA, 2018b; Fein, Staff, & Kobylenski, 1993; Gong & Wright, 2007) relative to groups that are characterized as inherently less powerful in the arena of evaluation practice (e.g., Mertens, 1999, 2007). This construction of difference in identity and conceptualization of power within the field of evaluation has implications for the field's response (House, 1983) to the incongruence observed between program participants and program evaluators, five of which are discussed in the next section.

Evaluation's responsibilities with respect to difference. Correlated with the evaluation field's construction of the difference between program participants and program evaluators is the field's conceptualization of its ethical responsibilities as a profession, if any, regarding the ongoing reproduction of racialized difference through surrounding structures and systems. Liberal conceptualizations of identity, as arising from differences in atomized attributes affixed to individuals, yield responses that correspondingly focus on individuals. In contrast, critical conceptualizations of identity—as socially constituted around shared interests arising from shared experiences of super-ordination or sub-ordination—yield responses focused on the structural arrangements underlying those asymmetrical relations. Five responses by the U.S. field of evaluation are discussed below. The first three represent official documents approved by AEA membership. The last two responses—social betterment and justice—come largely from the literature.

Propriety. The current (third) edition of the Standards describes the Standard of Propriety as supporting "what is proper, fair, legal, right, acceptable, and just in evaluations" (Yarbrough et al., 2011, p. 106) in terms that attend to culture and inclusion,[10] encompassing a responsive and inclusive orientation and formal agreements

[10] The standards for propriety also include protection of human and legal rights and maintenance of the dignity of participants and other stakeholders; clarity and fairness; transparency and complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations; identification and resolution of real or perceived conflicts of interests that may compromise the evaluation; and fiscal responsibility.

that take into account the needs, expectations, and cultural contexts of clients and other stakeholders (Yarbrough et al., 2011). Despite the opportunity presented by the addition of justice among the descriptors of the standard domain Propriety—a revision from the second edition, which described the standard in terms of legality, ethics, and due regard for the welfare of those involved in the evaluation (Sanders, JCSEE, & AASA, 1994)—the Program Evaluation Standards make no reference to structural dynamics,[11] namely, structural oppression.[12]

Common good and equity. AEA's 2018 edition of the Guiding Principles for Evaluators (hereafter, Guiding Principles) describes the principle of Common Good and Equity as "striv[ing] to contribute to the common good and advancement of an equitable and just society" (AEA, 2018b). It defines the common good as:

> … the shared benefit for all or most members of society including equitable opportunities and outcomes that are achieved through citizenship and collective action. The common good includes cultural, social, economic, and political resources as well as natural resources involving shared materials such as air, water and a habitable earth. (AEA, 2018b)

It defines equity[13] as "the condition of fair and just opportunities for all people to participate and thrive in society regardless of individual or group identity or difference," the quest toward which includes mitigating historic disadvantage and existing structural inequalities.

Specifically, the Common Good and Equity involve recognizing and balancing the interests of the client, other stakeholders, and the common good while also protecting the integrity of the evaluation; identifying and making efforts to address the evaluation's potential threats to the common good especially when specific stakeholder interests conflict with the goals of a democratic, equitable, and just society; identifying and

---

[11] The content of the Cultural Reading of the Program Evaluation Standards, analyzed in greater depth in Chapter Four, focuses considerable attention on power differentials and dynamics within evaluation practice and in society at large, although not on the role that evaluation as an enterprise plays therein.

[12] Under Young's enabling conceptualization of justice, "injustice refers primarily to two forms of disabling constraints, oppression and domination" (2011, p. 39).

[13] Other concepts that the 2018 edition defines include contextual factors, culturally competent evaluator, environment, people or groups, professional judgment, and stakeholders.

making efforts to address the evaluation's potential risks of exacerbating historic disadvantage or inequity; promoting transparency and active sharing of data and findings with the goal of equitable access to information in forms that respect people and honor promises of confidentiality; mitigating the bias and potential power imbalances that can occur as a result of the evaluation's context; self-assessing one's own privilege and positioning within that context. This description is elaborated upon in ways that differ markedly from previous editions of the Guiding Principles in its explicit recognition of the power dynamics that are present in evaluation—certainly in the evaluation of social programs—almost by definition, their acknowledgment of historic disadvantage, and their naming of both structural inequalities as well as self-reflexivity with regard to power and privilege. The 2018 edition's divergence from previous editions of the Guiding Principles—in which its potential as a response to the incongruence between program participants and program evaluators lies—as well as its limitations—which lie in its failure to define "justice"—are discussed in Chapter Six.

Cultural competence. Cultural competence features prominently in AEA documents as well as in the U.S. evaluation literature. Both the Guiding Principles and the Standards, as well as the Evaluator Competencies approved by AEA membership in 2018, recognize the importance of culture. AEA's Statement on Cultural Competence arose from its Building Diversity Initiative (BDI) study recommendation (#10) that the professional association "[e]ngage in a public education campaign to emphasize the importance of cultural context and diversity in evaluation for evaluation seeking institutions" (Association for the Study and Development of Community, 2002, p. 2). The statement is explicitly tied to the Guiding Principle of Competence, which currently specifies that evaluators "[e]nsure that the evaluation team collectively possesses or seeks out the competencies necessary to work in the cultural context of the evaluation" (AEA, 2018b).

The Statement defines cultural competence as "a stance taken toward culture, not a discrete status or simple mastery of particular knowledge and skills[14]…. Culturally competent evaluators respect the cultures represented in the evaluation" (AEA, 2011). Thus, while often applied to tools or processes, cultural competence is a characteristic that can only be borne by individual evaluators (as opposed to non-living tools or processes). It represents a way that people approach culture: respecting the cultures within the evaluation context. As john powell explains, however, "[i]f depressed life outcomes are produced by structures, then ending conscious discrimination is of little consequence and might actually exacerbate the negative impact of these structural dynamics by insulating the status quo from intervention" (powell, 2013, p. 13).

Importantly—unlike the current JCEE Standards, but like AEA's current edition of the Guiding Principles—AEA's Statement on Cultural Competence names asymmetries with respect to power: "Cultural groupings are ascribed differential status and power, with some holding privilege that they may not be aware of and some being relegated to the status of 'other'" (AEA, 2011). This phrasing suggests that differences in status, power, and privilege are ascribed to groups organized around an essentialized notion of culture as opposed to recognizing culture as arising largely from groups organized around shared experiences and interests. Moreover, it treats status, power, and privilege as static attributes "held" by groups as opposed to being exercised by groups through structurally mediated mechanisms. According to critical theories of systemic oppression, it is the asymmetrical flow of capital that reproduces difference (Young, 2011). While the Statement is elaborated upon in Chapter Six and its limited potential as a response by the field of evaluation to the incongruence between program participants and program evaluators are analyzed in greater depth in Chapter Eight, the static understanding of power that it conveys corresponds with the field's conceptualization of

[14] The *Oxford English Dictionary* defines competence as "I. In sense of compete: To enter into or be put in rivalry with, to vie with another in any respect. 1. Rivalry in dignity or relative position, vying. II. In sense of compete: To be suitable, applicable, or 'competent'. 2. An adequate supply, a sufficiency of; 4a. Sufficiency of qualification; capacity to deal adequately with a subject" and rooted in the Latin for "to strive after (something) in company or together" (OED Online, 2016).

justice (discussed at the end of this segment), with implications for the field's role in the systems surrounding it.

Social betterment. The "special relationship" (as the 2004 edition of AEA's Guiding Principles referred to it) between evaluators and clients constitutes the basis for critiques proffered against the primacy of use. These critiques gave rise to the notion of social betterment by Henry (2000), Henry and Julnes (1998), Henry and Mark (2003), Julnes, Mark, and Henry (1998), and Mark, Henry, and Julnes (2000) as well as to the notion of justice by House (e.g., 1997), which is discussed later in this section. The former have directly challenged the risks—which include incrementalism—involved in the relationship between evaluators and clients that an exclusive focus on evaluation use may obscure:

> The evaluation enterprise is no less vulnerable to goal displacement than the interventions that evaluators study…. Use, once injected as a goal or guidepost for planning an evaluation, can begin to take on a life of its own, rather than serving as a means to an end. (Henry & Mark, 2003, p. 86)

Contrasted with use, "social betterment is defined as the extent to which public policies and programs meet the ever-emerging complex of human needs" (Henry & Julnes, 1998, abstract) or more specifically as "improved social conditions, the reduction of social problems, or the alleviation of human distress" (Henry, 2000, p. 86). It is illustrated through a litany of "commonsense" examples (Henry & Julnes, 1998, p. 53).[15]

---

[15] "Less homelessness is better than more. Higher levels of reading comprehension are preferred to lower levels. Work is preferable to welfare" (Henry & Julnes, 1998, p. 53). Given the first two examples, one might expect the third to be phrased as "More work available is preferable to less," "More jobs are preferable to less," or even "Higher rates of employment are preferable to lower rates." Instead, the last example is the only one made in qualitative rather than quantitative terms, wherein the notion of work is contrasted with the highly racialized, gendered, and classed notion of welfare. See Omi and Winant (1993) for a discussion of racism as simply a crystallization of the dogma that underlies "common sense" regarding social groups in a racialized society and (Roberts & Mahtani, 2010) for a discussion of the "seductive, common-sense logic to neoliberalism that reproduces racist ideologies" (Roberts & Mahtani, 2010, p. 255). Written soon after passage of the Personal Responsibility and Work Opportunity Reconciliation Act, otherwise known as "Welfare to Work," Henry and Julnes' (1998) subtle deviation from parallel structure raises the following questions: Within whose 'common sense' is welfare the scalar opposite of work? Under what conditions is welfare less preferable than work? For whom is movement from welfare to work social betterment?

The ideas of social betterment, improvement, and progress (used interchangeably by Henry and Julnes [e.g., 1998]) deviate somewhat from the ideas of propriety, fairness, legality, righteousness, and justice in the Joint Committee on Standards for Educational Evaluation's (JCSEE) Standards for Program Evaluation as well as from the ideas of general and public welfare or common good and equity in AEA's 2004 and 2018 Guiding Principles, respectively. Proponents of social betterment as the appropriate goal of evaluation engage with notions of justice and even cite House (1995b) as foreshadowing a summary of their position. They do so, however, without adopting the language of justice or "more direct political purposes, such as social critique and social change" (Greene & Walker, 2001, pp. 369-370):

> Inability to pose a set of universal values, such as equality or justice, as moral guideposts has given rise to the belief that progress is an illusion…. [W]e admit that social progress is notoriously difficult to define in that it is not inevitable nor is it likely to be completely comprehensible in terms of any current value system. Nonetheless, we reject claims that improvement is not possible or that it is impossible to intentionally support this sort of progress…. [W]e cling to the necessity of the possibility of social betterment as the preeminent rationale for evaluation…. (Henry & Julnes, 1998, p. 57)

Despite the phrase's reverberation of the explicit purpose of the eugenics' movement (Wikler, 1999), some evaluation scholars continue to declare (Henry, 2000; Henry & Julnes, 1998; Henry & Mark, 2003; Julnes, Mark, & Henry, 1998; Mark et al., 2000) and others accept (e.g., Chen & Turner, 2012; DeGroff & Cargo, 2009; Letichevsky & Penna Firme, 2012; Orwin, Campbell, Campbell, & Krupski, 2004) the notion of social betterment as evaluation's ultimate goal. While the notion of social betterment has weathered critique (e.g., Greene & Walker, 2001; King, 2016), its potential as a response by the field of evaluation to the incongruence between program participants and program evaluators is limited by its failure to address whose idea of betterment prevails and its failure to justify how or why. Greene and Walker summarize these limitations as follows:

> First,… our claims to know about a given social program or policy are made from a particular value stance—one that privileges participatory democracy, liberal pluralism, technical progress, utilitarianism, or social betterment. Our claims to

> know are unavoidably *interested* claims. Second, the social practice of evaluation is not separable from the socio-political practices and institutions to which it is designed to contribute or in which it is embedded (House & Howe, 1999). There are no viable sidelines in democratic political discourse. The very activity of generating evaluative knowledge about social programs and policies helps to constitute the form and function of this discourse. (2001, p. 371, emphasis in original)

Justice. As much as constructions of difference are tied to power, conceptualizations of power are tied to justice. While the JCSEE has refrained from elevating the notion of justice to a standard (the third edition of the Standards includes it among the descriptors of the standard domain of Propriety), AEA's 2018 edition of the Guiding Principles includes justice in the principle of the common good and equity. However, unlike the common good and equity—both of which it defines—it does not elaborate on the meaning of justice.

Some evaluation scholars have deepened the discourse around evaluators' ethical responsibilities beyond JCSEE's Program Evaluation Standards and AEA's Guiding Principles by exploring the concept of justice. In introducing the notion of justice into the field of evaluation, House refers to values. Not only are values not necessarily entirely subjective (a realization House connects to Scriven's later scholarship on logic), but values are also not necessarily entirely new to the fact-focused endeavors of science and research. They are already implicit in evaluation, and by articulating a place for justice in evaluation, House says he is simply making his values explicit.

The general notion of justice has appeared in the U.S. evaluation literature since the publication of *Justice in Evaluation* (House, 1976)—sporadically, initially, with a proliferation of articles published in the issue of *New Directions for Program Evaluation* (now *New Directions for Evaluation*) devoted to the intersection between evaluation and social justice in education (Sirotnik, 1990). While "justice"—and "social justice" in particular—have featured consistently in the evaluation literature every year since 2009 (e.g., Datta, 2011; Hesse-Biber, 2013; Thomas & Madison, 2010; Woelders & Abma, 2015)—most prominently written about by House (e.g., 1976, 1990, 1991, 1995b), Kirkhart (Collins, Kirkhart, & Brown, 2014; Kirkhart, 1995), and Mertens (2007, 2013,

2016)—only six articles in the U.S. peer-reviewed evaluation literature as of December 2018 contain the phrase "economic justice" and another six contain the phrase "racial justice."[16] This is despite the field's beginnings in educational research and government research on Great Society programming such as the War on Poverty, both of which explicitly engaged with larger processes of racialization.[17]

Social justice. Although frequently invoked, social justice is infrequently defined in the U.S. evaluation literature. Heavily influenced by Rawls' *A Theory of Justice* (1971), House's understanding of social justice in educational evaluation follows:

> Inequality, if it's allowed, should be to the benefit of those least advantaged in society. I saw that as being an entrée into programs for the poor, the impoverished. The other principle is a rights issue. You shouldn't violate student rights. (The Oral History Team, 2015, p. 273)

In her Presidential Address at AEA's 1994 Annual Conference, whose theme was Evaluation and Social Justice, Kirkhart defined social justice as "fundamentally about equity" (1995, p. 2). Citing House (1976, 1991), she explicated the underlying issue in social justice as one of "balancing the interests of different segments of our society" (Kirkhart, 1995, p. 2) and, in ways that foreshadow AEA's Statement on Cultural Competence that would be approved 15 years later, she advanced the notion of multicultural validity as necessary to first understand those different segments' interests. Kirkhart defined multicultural validity as the "ability to capture these multiple cultural perspectives accurately, soundly, and appropriately" (p. 2) and as including

---

[16] The phrase "racial equity" has appeared eight in the peer-reviewed evaluation literature from its inception through December 2018. This does not include literature in educational research and other related fields.

[17] Educational research originated in the eugenics movement (Besag, 1981) and the War on Poverty was based on *The Negro Family: The Case for National Action Report* (Moynihan, 1965). Both the eugenics movement and the Moynihan report's language of pathology reflect and reinforce processes of otherizing beyond racialization and systems of oppression beyond white supremacy. The first does so on the basis of heredity and the second does so on the basis of culture. Specifically, they reflect and reinforce ableism, capitalism, and cis-hetero-patriarchy. Notably, while "the culture of poverty" is often associated with Moynihan and the supposedly pathological matriarchal culture among urban residents classified as Black that his report described, the phrase originates in a structural understanding of poverty that was derived from studies of those classified as both Black and White in both urban and rural settings (Harrington, 1962).

methodological, interpersonal, and consequential validity. Methodological validity "concerns the soundness or trustworthiness of understanding warranted by our methods of inquiry" (p. 4). Recalling Stake's work, interpersonal validity "refers to the soundness or trustworthiness of understandings emanating from personal interactions" (p. 4). Consequential validity refers in Kirkhart's conceptualization to "the soundness of change exerted on systems by evaluation and the extent to which those changes are just" (p. 6).

This understanding of social justice and interest in validity[18] also underlies Mertens' (2007) description of social justice as a rights-based theory of ethics applied to the group or societal level, which she describes as equal weight if not precedence being given to the least advantaged groups in society, both to generate an accurate representation of their viewpoints and to provide a form of redress and empowerment for those considered without sufficient power to take an active agent role in social change. She places evaluative work to address injustice in the lineage of scholars such as Robert Stake (1974), whose responsive model of evaluation emphasized the importance of context and relationships with the stakeholders, including their experiences, beliefs, and values (Mertens, 2013). Stake himself, however, rejects the idea that evaluation should play any role in social justice-oriented movement organizing (House, 2001).

In contrast to the primary purpose of evaluation proffered by Mark, Henry, and Julnes (2000), Mertens names her assumption of evaluation's primary purpose as the promotion of human rights and social justice and the concomitant need to recognize power and cultural differences (Mertens, 2013; also see Bledsoe, 2014). She recognizes the need for evaluators to "identify those cultural norms, beliefs, and practices that support human rights and social justice and those that sustain an oppressive status quo" (Mertens, 2013, p. 29) and calls for evaluators to arrange "culturally appropriate opportunities to address the norms, beliefs, and practices that support or conflict with the pursuit of social justice" (p. 29). Mertens' summary of the social justice approach, however—as requiring that evaluators "design their evaluation in ways that facilitate

[18] These distinct but related interests parallel the distinction between practical participatory evaluation and transformative participatory evaluation (King, Cousins, & Whitmore, 2007).

leaving the community better off than it was before the evaluator departs" (2013, p. 29)—differs minimally from the social betterment approach of Mark, Henry, and Julnes (2000): "It is hard to imagine anyone laboring in the field of evaluation who does not think in terms of making the world a better place through his or her work…" (Mark et al., 2000, p. 19).

Referring to community-based social work, Todd explains the problem with "such a hopeful heroic stance" (2011) that applies also to evaluation:

> … [I]t tends to obscure the multiple effects of [the] work, some facilitating progressive change and others re-inscribing oppressive relations (or drawing on practices or perspectives that rely upon and thus sustain existing practices of inequality). These multiple effects happen despite the good intentions and skills of academics and community practitioners. They reflect the social fabric into which community work is woven. (Todd, 2011, p. 118)

This dissertation is rooted in the tension described above, i.e., between individual evaluators' intentions and the racially differentiated "economic, social, political, or ideological rewards or penalties received by [individuals and institutions] for their participation (whether willing, unwilling, or indifferent)" (Bonilla-Silva, 1997, p. 469) in the social fabric within which evaluation is woven.

Expanding boundaries beyond individual practitioners. Drawing from House's conceptualizations of validity, both Kirkhart and Mertens place primacy on evaluation approaches and methods that allow the evaluation to represent the perspectives of those whose perspectives may not necessarily be sufficiently represented otherwise. Both also focus on difference as cultural; in their conceptualization, evaluators—as individual actors—are responsible for attending to the cultures represented in the evaluation, including their own. Finally, both Kirkhart and Mertens consider the effects of the evaluation on the participants. Kirkhart's concept of multicultural validity includes consequential validity and Mertens' transformative mixed methods approach explicitly calls for evaluators to provide some compensatory act of advocacy or material support, beyond individual incentives that would serve the long-term interests of "the community."

Neither considers the material and discursive role that evaluation plays within the industry surrounding it or in larger society (House, 2017). Like much of the evaluation literature that has referenced House's writing on social justice, both Kirkhart and Mertens stop short of addressing the larger context[19]—the "social fabric" of racialized neoliberal public policy into which evaluation as a field is "woven"—that House himself emphasizes on more than one occasion (e.g., House, 1985, 1999a; 2017). In fact, House calls for a philosophy of evaluation as a field (House, 1983b), identifying the increasing need for the field of evaluation to consider its role as a participant in public policy (e.g., House, 1987a; 2014; House & Howe, 1998) and to hold itself accountable from the perspective of justice (House, 1987b, 2014; Oral History Team, 2015).[20]

The move away from House's explicit discussion of race and policy is not unique to the evaluation literature or its professional association's documents. "[C]onsistently mystify[ing] the process of racial accumulation through occlusion of history and forsaking structural analysis for a focus on the individual" (Leonardo, 2004, p. 141) corresponds with the larger depoliticization of difference and disavowal of systemic oppression that is characteristic of neoliberal notions of "color-blindness" (Doane, 2006). To be sure, among evaluation scholars, it may represent a well-intentioned attempt to avoid prioritizing race among dimensions of difference and social group identification, which also include, but are not limited to, gender, sexuality, class, and ability status. Still, they have yet to connect the salience of these dimensions of difference and social group

---

[19] The small body of evaluation literature focused on context, which "comes from the Latin word *contextus*, meaning 'to join together' or 'to weave together'" (Dahler-Larsen & Schwandt, 2012, p. 75) includes no analysis of race. It is reviewed in Chapter Two.

[20] Because the topic of this study concerns the treatment of racialized difference within the U.S. field of evaluation, specifically, House's more detailed accounts of the process of racialization, neoliberalism, and evaluation's larger societal role, all of which can be found in House's writing in the educational policy and research literature (e.g., House, 1999b) and in evaluation-related publications that were not subject to peer-review (e.g., House, 1998), are not addressed.

identification[21] to the interlocking forms of systemic oppression[22] that include and extend beyond white supremacy, such as cis-hetero-patriarchy, capitalism, ableism, and the intersections[23] therein. Instead, it has retrenched both the field's unit of analysis and its unit of redress to internalized and interpersonal dynamics within the realm of evaluation practice at the expense of the racially stratified structure of the industries in which evaluators play a pivotal and interested role, not just as professionals, but also as members of particular social groups (Greene & Walker, 2001; House, 1990; Young, 2011).

Surrounding structures of racialized neoliberalization. The body of House's writing on social justice within the U.S. academic evaluation literature betrays a critical

---

[21] AEA's nearly 60 Topical Interest Groups (TIGs) include seven that focus on identity-based groups and associated issues: Disabilities and Other Vulnerable Populations; Indigenous Peoples in Evaluation; Latinx Responsive Evaluation Discourse; Lesbian, Gay Bisexual and Transgender Issues; Military and Veteran's Issues; Multi-ethnic Issues in Evaluation; and Youth Focused Evaluation.

[22] Because the focus is on identity rather than underlying structures of oppression, the social groups referred to by these TIGs are treated as requiring special attention even though, together, they form the overwhelming majority of evaluated programs' participants. Only the Feminist Issues in Evaluation TIG refers to a critical analysis (feminism) of power dynamics (patriarchy) rather than a particular social group (e.g., "women in evaluation") or set of interests ("gender issues in evaluation"). In sharp contrast, the name of the International and Cross Cultural Evaluation TIG suggests evaluation across lateral differences in nation and culture, masking the TIG's focus on the evaluation of development aid programs whose target populations are former colonial subjects—disproportionately racially otherized—and which are funded by former European colonial powers and states settler-colonized by Europe. The focus on individual identity as opposed to oppressive structures in AEA's TIGs corresponds with the treatment of these groups as separate special interests. Indeed, only since 2015 have some of what are called "the diversity TIGs" begun to co-program.

[23] The Feminist Evaluation TIG's conference programming and scholarship's analysis of patriarchy has generally failed to incorporate Black Feminist Thought or Postcolonial Feminism—one important exception being Hood and Cassaro (2002). "Intersectionality" was coined by critical race theorist and legal scholar Kimberlé Crenshaw to describe the predicament of an African American woman who was denied a position at an automobile plant because she was considered inappropriate for both the customer service-oriented positions occupied by White women as well as the labor-oriented positions occupied by Black men. Her claim of discrimination on the grounds of both sex and race simultaneously was dismissed because the court felt it constituted "double-dipping." Crenshaw was the first to say that the experience of intersectionality is not unique to individuals who are racially otherized and gendered as women, but that it encompasses multiple dimensions of identity (Crenshaw, 1989). Both the pattern of interaction among the professional association's TIGs and the absence of an intersectional analysis in the literature limit the field's potential to identify and address mechanisms within the industries it serves that negatively affect multiple social groups or that negatively affect social groups with multiple salient identities (Reskin, 2003).

analysis of explicitly racialized social and educational policies in the U.S. (1999a) and a shift from New Deal and Great Society programs to deregulated markets with fewer protections for labor and the environment in addition to privatized social programs that are predicated on an understanding of poverty as the result of individual choices rooted in cultural pathologies and deficits (House, 1990, 2014; Oral History Team, 2015). While he does not refer to it as such in the evaluation literature, the shift that House described at length on multiple occasions throughout his career corresponds with the process of neoliberalization,[24] whose intrinsically racialized nature—not just in terms of impact, but in terms of its active production of racialized bodies—has been described as "reproduc[ing] racial knowledge with every outwardly progressive gesture, which works to normalize racism as just an aspect of life" (Roberts & Mahtani, 2010, p. 254). Even those evaluation scholars who have addressed U.S. neoliberalism directly (Dahler-Larsen et al., 2017; Mathison, 2011; Schwandt, 2005; Sturges, 2014; Tranquist, 2015)—some of whom have drawn from House's work in Deliberative Democratic Evaluation (Mathison, 2000) and others who have invoked the notion of justice specifically (Desivilya Syna, Rottman, & Raz, 2015)—have eschewed the topic of race with which neoliberalism is so closely intertwined (Roberts & Mahtani, 2010).

Multiple conceptualizations of justice. While House draws heavily from Rawls (1971), he acknowledges that "minorities and feminists have challenged the stand-apart conception of justice" (House, 2014, p. 13)[25] represented in Rawls' framework by impartial experts who judge what is just based on fundamental principles:

> The idea that participants can express their own views, values, and interests has supplemented, and in some cases supplanted, the idea of impartial judges. Hence, we have participatory evaluations, including deliberative democratic

[24] The movement of policies and practices to correspond with neoliberal philosophy. Neoliberalization maximizes free trade through increased privatization, deregulation, and competition, as well as deficit reduction, through the reduction of government investment in social and economic protections (Roberts & Mahtani, 2010).

[25] House further apologized "for the dated parts of the book, the politically incorrect usage of the time, and the intemperate remarks," adding, "written works bear the mark of their time, place, and maker" (2014, p. 14).

> evaluations…. [J]ustice applies as a criterion, but what we mean by justice has shifted. (House, 2014, p. 13)

In much the same way that House describes the challenge, supplementation, and supplanting of Rawlsian notions of justice by "minorities and feminists," this dissertation represents a challenge, supplementation, and supplanting of House's interests in justice, race, and the role of evaluation within a larger context by critical race theory, postcolonial theory, and feminist theory.

As much as Rawls prioritizes justice as the primary purpose of social institutions, the paradigm of justice that he advances is distributive. In the distributive conceptualization of justice, notions of equality and fairness form "the ground floor" (Mills, 1994, p. 118) of the social contract. Such a paradigm is associated with liberal individualism, wherein persons are conceptualized as consumers and possessors of static goods or positions (Young, 2011). Race is irrelevant to personhood and would be hidden by the veil of ignorance during deliberations of justice as conceptualized by Rawls (Mills, 1994). Alternative paradigms of justice have emerged from a "cognitively advantaged perspective—the view from the basement" (Mills, 1994, p. 120) as opposed to the ground floor. For example, Young (2011) proposed what she called an enabling conception of justice:

> Justice should refer not only to distribution, but also to the institutional conditions necessary for the development and exercise of individual capacities and collective communication and cooperation. Under this conceptualization of justice, injustice refers primarily to two forms of disabling constraints, oppression and domination. While these constraints include distributive patterns, they also involve matters which cannot easily be assimilated to the logic of distribution: decision-making procedures, division of labor, and culture. (p. 39)

An enabling view of justice is premised upon a foundation of systemic oppression rather than equality and fairness. It centers social groups—as opposed to individuals—as agents of change and producers of knowledge. Arising from movement organizing efforts against systemic oppression, enabling conceptualizations of justice focus on dynamic processes and relations rather than static goods or positions (Young, 2011). Because enabling conceptualizations of justice arise from critical theories of systemic

oppression—which arise from movement organizing traditions and recognize difference as dialectically (re)produced among social groups through asymmetrically structured relations over time rather than static and attached to individuals—enabling conceptualizations of justice emphasize the nature of structural relations. They specifically focus on the extent to which they reinforce or counteract social groups' efforts to exercise collective agency on changing the structural conditions that govern their participation in decision making and socio-economic processes. Under an enabling conceptualization of justice, status, power, and privilege are considered forms of capital—stocks—that flow through structurally mediated mechanisms and media of exchange.

In summary, Kirkhart and Mertens' approach to social justice differs from House's through their attention to individual dynamics within the realm of evaluation practice in contrast to House's attention to dynamics involving evaluation as a field and other institutions in larger industrial and societal contexts; through their emphasis on culture; and through their ahistorical and decontextualized treatment of difference, which differs from House's consistent treatment of race in relation to U.S. history and the increasing neoliberalization of public policy. Still, House's approach to justice is limited by its static focus on the distribution of resources as opposed to a dynamic focus on the structurally mediated socio-economic processes and relations—exchanges of capital—that either enable social groups to organize and dismantle asymmetrical structures or hinder them from doing so. This dissertation is intended to begin addressing that limitation.

Problem Statement

The current climate—not just in the USA but internationally—wherein inequality violence against those who are otherized has not declined but has in fact increased—lends a sense of urgency to questions about evaluation's role with respect to difference, power, and justice. U.S. evaluation literature and professional association documents are replete with acknowledgment of the differences in phenotype, experience, and perspective observed between program participants and program evaluators (Lincoln,

1991; ASDC, 2001). The field has largely framed discussion of this difference in terms of cultural incongruence and responses to it in terms of dynamics within and among individuals under the realm of evaluation practice, however, rather than in terms of structural dynamics within and among actors in the industries within which evaluation is practiced (Kirkhart, 1995, 2010; LaFrance et al., 2012; Lincoln, 1991; Madison, 1992b; Symonette, 2004). Discussing the incongruence without historicizing or contextualizing its ongoing (re)production through asymmetrically structured socio-economic relations suggests that the incongruence is natural, reinforcing the normativity of whiteness.

Such framing further leaves the underlying structures unexamined and potentially intact—reflecting, but also serving, neoliberalism's normalization of racial stratification as simply a natural result of market forces. The field of evaluation will not likely see a purportedly desired change in the incongruence observed between program participants and program evaluators unless it considers its material and discursive role relative to the racially stratified organization of labor and decision-making processes within the industries that utilize evaluation services and results, as well as the neoliberal policies that continue to shape those industries (House, 2017). In contrast with the evaluation field's prevailing focus on interpersonal dynamics rooted in cultural differences between program participants and program evaluators, as well as its explicit rejection of justice in AEA's recently approved competencies (AEA, 2018a), this introductory chapter has illuminated the need for the field of evaluation to approach the incongruence widely observed between program participants and program evaluators with an explicitly enabling conceptualization of justice. Such a conceptualization would require an analysis of systemic oppression and explore remedies directed at the relations and processes through which difference is produced and reproduced.

Study Purpose and Research Questions

The purpose of this study is to advance an enabling conceptualization of justice in the field of evaluation by broadening the current framing of difference in evaluation discourse and directing attention to the historical and ongoing socio-economic relations and decision-making processes involved in producing and reproducing difference. In

particular, it focuses on the material and discursive role that evaluation plays as the exchange of capital within the racially stratified industries that utilize evaluation services and results (House, 2017). It uses critical theories of systemic oppression and systems thinking to examine the construction of racialized difference *in* as well as *through* evaluation. Centering the experiences and perspectives of racially otherized groups as protagonists—as producers of knowledge and agents of change rather than as consumers of evaluated services alone—in its conceptualization of justice as well as its methodology, this dissertation research asks the following:

1. How has the U.S. scholarly evaluation literature constructed racialized difference?
2. How has that construction changed since the field began formalizing in the early 1970s?
3. How does that trajectory relate to the systems surrounding evaluation?

Significance of Study

Crenshaw identified the following "definitional tension" (1988, p. 1336) regarding antidiscrimination law that also applies to the evaluation field's ambivalence with respect to the incongruence observed between program participants and program evaluators:

> Is the goal limited to the mere rejection of white supremacy as a normative vision or *may the goal be expanded to include a societal commitment to the eradication of the substantive conditions of Black subordination*? …When discussed as a normative vision, white supremacy is used to refer to a formal system of racial domination based on the explicit belief that Blacks are inferior and should be subordinated…. [A] society once expressly organized around white supremacist principles does not cease to be a white supremacist society simply by formally rejecting those principles. The society remains white supremacist in its maintenance of the actual distribution of goods and resources, status, and prestige in which whites establish norms which are ideologically self-reflective[26]…. (Crenshaw, 1988, p. 1336, emphasis added)

[26] Norms that are ideologically "self-reflective" of those who established them reflect the ideologies of their establishers—in this case, those racialized as White. This is not to be confused with norms that are ideologically "self-*reflexive*," which would demonstrate an exploration of "*how we are in relation with the contexts we study and with our informants*, understanding that we are all multiple in those relations" (Fine, 1998, p. 135, emphasis added).

A commitment by the field of evaluation to eradicating the substantive conditions of subordination among racially otherized groups should not be confused for activism or advocacy, as it does not favor any political party, policy, or program over another. It favors an end to racialized system dynamics that (re)produce racialized difference.

Using critical theories of systemic oppression and systems thinking to advance an enabling conceptualization of justice, this dissertation research differs from the bulk of existing evaluation literature concerned with the incongruence observed between program participants and program evaluators in the following ways:

- It expands the current boundaries of space to permit examination of the intergroup and intragroup relations that produce and reproduce difference between program participants and program evaluators.
- It shifts the focus from accommodating or compensating for static, individualized differences in culture to inform alteration of the structural mechanisms that mediate the above processes and relations.
- It expands the current boundaries of time to permit examination of the historical and ongoing processes that produce and reproduce accumulated difference between program participants and program evaluators.

By expanding the boundaries of the unit of analysis to include dynamics within and among institutions in the industries that utilize and provide evaluation services and results, this study has the potential to increase investment in interventions that disrupt the unfettered deregulation, privatization, and devolution of government programs, all of which disproportionately harm otherized groups.

Conclusion of Chapter One

This chapter has drawn from critical theories of systemic oppression and systems thinking to problematize the U.S. evaluation literature's emphasis on dynamics within and among individuals involved in evaluation practice at the expense of a structural analysis of the racialized structural dynamics within and among institutions in the industries that utilize evaluation services and results. Offering a critical, systems-oriented

alternative to the liberal construction of difference, it illustrated how difference is (re)produced and identity is (re)constituted cumulatively and interactively through historical and ongoing processes that involve time and asymmetrical structural relations. Additionally, it tied critical constructions of difference to enabling conceptualizations of justice through the structurally mediated ability to exercise, rather than possess, power. Finally, it tied enabling conceptualizations of justice to evaluation's role in an increasingly oppressive national and international climate.

Organization of the Dissertation

Chapter Two provides an overview of racial stratification in the USA and within the nonprofit industry. It then reviews the program evaluation literature that attends to context or multiple levels of analysis or that addresses racialized difference. It closes with an orientation to concepts from systems thinking and critical theories of systemic oppression that can contribute to the field's understanding of and response to the incongruence observed between program participants and program evaluators. Chapter Three details the research methodology, which includes critical discourse analysis (CDA) of the evaluation literature, documents, and interviews with those most closely involved with the evaluation field's construction of racialized difference and development of corresponding remedies. Chapters Four, Five, Six, and Seven focus on the results of the three research questions. Chapter Eight interprets the results and Chapter Nine discusses the implications for future research and practice. Like the prologue that preceded Chapter One, an Epilogue is offered after Chapter Nine in the spirit of counter narrative.

Definition of Key Terms

Ascriptive: Defined by an ascribed characteristic that is based on something outside individual control, such as heredity, e.g., race, sex, age, class, and religion at birth (adapted from Reskin, 2003).

Asymmetrical, hierarchical, or stratified relations: Relationships and interactions wherein one party is capable of disproportionately imposing their will on the other and

setting conditions, making decisions, taking actions, and exercising control that determines the nature or outcome of the relationship (adapted from Guess, 2006).

Capital: Capital is the total sum, store, or stock of accumulated contributions—past flows invested in, minus flows out (depreciation)—of some valuable, positive, or advantageous quality that allows efforts to operate. It is used to create more value as opposed to being used for consumption. Typically in a capitalist economy, where the goal is to maximize shareholder profit, capital is considered in relation to labor and production output. Ecological and cooperative economics proffer multiple types of capital, for example: natural living and non-living capital, relational and human capital, learned and created capital, structural and technological capital, cultural and spiritual capital, economic and financial capital (adapted from Gowdy & Erickson, 2005).

Cultural imperialism: Universalization of the super-ordinated social group's experience and culture and its establishment as the norm. The cultures of other groups are simultaneously invisible, seen only in contrast to the group super-ordinated by systems of oppression (typically but not always as falling short of it) and stereotyped. Those living under cultural imperialism find themselves defined from the outside by those with whom they do not identify and who do not identify with them. To operate within that culture, however, they must to some extent internalize its perspectives (adapted from Young, 2011).

Culture: A social system of meaning, customs, and unwritten rules that a social group has developed to assure its adaptation and survival in a particular physical, geographical, historical, social, political, and economic context. It shapes how its members live together and it distinguishes them from others. It includes values, beliefs, patterns of thinking, styles of communication, aesthetic standards, linguistic expression, and behavioral norms (adapted from *15 Tools for Creating Healthy, Productive Interracial/Multicultural Communities: A Community Builder's Toolkit*, nd).

Currency: The media through which various types of capital are exchanged (Roland, 2011).

Distal and/or cumulative effects: Effects that are not necessarily proximate to their cause in time and space, because the media of exchange needs to flow through the system's structural mechanisms. Cumulative effects occur when a cause within one domain may influence other causes within that domain over time, or in other domains, so that the initial causes produce much larger systemic effects (adapted from Menendian & Watt, 2008).

Dialecticism: The Marxist notion of ongoing tension and contestation among those engaged in class struggle. Here it applies to the process of producing categories of difference by creating groups of "same" and groups of "other" in opposition to each other. The categories are fluid rather than fixed, however, as groups sub-ordinated by systems of oppression continually contest their status while groups super-ordinated by systems of oppression continually find ways to re-establish theirs (adapted from Bonilla-Silva, 1997).

Differential racialization: The way that multiple social groups are racialized differently or the same social group at multiple points in time is racialized differently, depending on the historic, social, political, or economic needs of those racialized as White (adapted from Abrams & Moio, 2009; Delgado & Stefancic, 2001).

Discourse: The social process of making meaning through text, speech, and other vehicles; the language associated with a particular social field or practice; and a way of construing aspects of the world associated with a particular social perspective (adapted from Fairclough, 2012).

Diverse: Assorted, multi-form, varied or with its varieties (OED Online, 2016).

Essentialism: Reductive construction of commonalities, including common social conditions, among otherized social groups as intrinsic and pre-social rather than as (re)produced through social relations with super-ordinated groups, whose identities—in contrast—are typically portrayed as complex and varying from individual to individual. Essentialization is increasingly based on notions of culture (often used as a proxy for

race) rather than heredity (adapted from Bannerji, 2000; Hollinsworth, 2013; Jani et al., 2011).

Exploitation: The process in which the results of the labor and energy expenditure of one social group is steadily carried out or transferred to benefit another, continuously reproducing relations of super-ordination and sub-ordination between them. For example, "menial labor," unskilled, servile, low-paying work lacking in autonomy is often reserved for women—particularly women who are racially otherized globally—but ultimately benefits men—particularly men classified as White in countries that were former colonial powers or are settler colonial states (adapted from Young, 2011).

Feedback loops: A particular type of stock-flow that changes the magnitude of a stock in ways that then influence the original stock-flow, making subsequent change in magnitude either more or less likely. There are two types of feedback: Positive feedback loops reinforce or amplify some phenomenon that a cause helped effect (which may be considered "good" or "bad"—the effect is not necessarily "positive"; it is the direction of growth that is positive). Negative feedback loops balance or counteract some phenomenon that a cause helped create (which may also be considered "good" or "bad"—again, the effect is not necessarily "negative"; it is the direction of growth that is negative.) In positive feedback loops, initial changes become amplified or magnified over time. Achievement of the product encourages the process that created it in what is commonly called a vicious or virtuous cycle. An example is interest-bearing accounts: interest earned on the balance continually increases the balance and subsequent interest earned. In contrast, in negative feedback loops, initial changes are counteracted or balanced out so that conditions remain relatively stable. Achievement of the product inhibits the process that created it. An example is body temperature: heat causes a body to sweat, which cools the body down (adapted from Flood, 2010; Menendian & Watt, 2008).

Flows: Exchanges, relations, processes, dynamics, or activities that take place over time and are measured per unit or interval of time as rates or speeds of verbs. Flows either fill or drain stocks, updating their magnitude. In-flows add to the stock and out-flows

subtract from it. Flows may refer to the total value of transactions, including income, expenditure, saving, debt repayment, fixed investment, inventory investment, and labor utilization during an accounting period (adapted from Sterman, 2001).

Identity: A product of discursive and material social, cultural, and economic interaction; "an internal organization of self-perception concerning one's relationship to social categories, that also incorporates views of the self-perceived to be held by others. Identity is constituted relationally, through involvement with-and incorporation of-significant others and integration into communities" (adapted from Epstein, 1987).

Imperialism: One of two forms of colonialism—"the process by which European powers reached positions of economic, military, political, and cultural hegemony in much of Asia, Africa, and the Americas" (Shohat & Stam, 1994, p. 13). (The other form is settler colonialism, defined separately, below). Imperialism involves distant control of resources. The conquest of territory is linked to the systematic search for markets, social interventions, and exportation of capital by former colonial powers and current settler-colonial states in former colonial subject nations even after independence. Involving both internal processes of cultural appropriation and external struggles over international power, imperialism continues to foster the diffusion of capitalism as a dominant mode of production from Europe and the USA to the rest of the world and institutionalizes unequal economic and power relations internationally. Because imperialism is closely associated with war and extracted elements of nature, it involves the global movement of capital —natural and human "resources." The latter may be in the form of immigrants, or exploitable labor, as well as in the form of refugees, or marginalized/surplus labor (adapted from Midgley, 1998; Volpp, 1996).

Individualism: A normative conception of the self as independent that is central to neoliberal philosophy and policy (Leonardo, 2004). "The authentic self is autonomous, unified, free, and self-made, standing apart from history and affiliations, choosing its life plan entirely for itself" (Young, 2011, p. 45).

Institutional dynamics: Dynamics that occur within institutions, wherein policies and practices—including unwritten customs and traditions—systematically create differential outcomes for social groups, regardless of the policies and practices' intention and of their specification of any group in particular (adapted from Leiderman, Potapchuk, & Major, 2005).

Interest convergence: The need for policies and programs targeted at helping groups sub-ordinated by systems of oppression to coincide with the interests of groups super-ordinated by those systems, even (or perhaps especially) when the latter keep their interests silent (adapted from Stec, 2007). "Acts that directly help blacks [*sic*] must implicate white interests because white economic (and other) interests and black oppression are inextricably interwoven and depend on each other for their survival" (Stec, 2007, p. 31).

Internalized dynamics: Dynamics that occur within individuals living under systemic oppression, wherein members of all social groups—including those experiencing sub-ordination by systems of oppression—willingly or inadvertently maintain or participate in the set of attitudes, behaviors, social structures, and ideologies that undergird the position of the super-ordinated group. Individually or collectively, those who have internalized white supremacy may assume or relinquish decision-making responsibilities, depending on their status as super-ordinated or sub-ordinated; devalue needs and priorities of sub-ordinated groups that are not shared by the super-ordinated group; base standards for what is appropriate or "normal" on the super-ordinated group; or fail to examine and hold accountable the super-ordinated group and its individual members, instead scrutinizing and finding fault with sub-ordinated groups and their members for the conditions they experience (adapted from Bivens, 1995).

Interpersonal dynamics: Dynamics that occur among individuals; intentional or unintentional, overt or covert actions or behaviors perpetrated by individuals against others because of learned or internalized beliefs—whether conscious or subconscious—about groups with which they are identified. Even unintentional actions or behaviors rooted in subconscious beliefs may support or perpetuate systemic oppression and thus be

oppressive in terms of their impact. Without societally or institutionally accorded power and privilege to act on conscious or subconscious negative beliefs about particular groups, however, the beliefs themselves do not constitute systemic oppression. Indeed, the definition of racism as "prejudice plus power and privilege" arose in response to accusations of "reverse racism," which is imagined as beliefs held and actions perpetrated by individuals racially otherized against individuals racially normatized as White (adapted from Bivens, 2005; Leiderman et al., 2005).

Intersectionality: Experiencing varying levels and types of super-ordination and sub-ordination by multiple systems of oppression as a result of identification with multiple social groups and their associated status, each of which interacts with and inflects the others. Every individual and every issue is intersectional—the manifestation of which is greater than the sum of its parts. For example, the experience of cis-hetero-patriarchy—that is, both the sub-ordination of those classified as women and those who threaten the gender binary upon which that sub-ordination depends (perhaps by identifying as LGBTQ) as well as the super-ordination of men and those who maintain that super-ordination (perhaps by identifying as the gender assigned as birth or as heterosexual)—is simultaneously and necessarily racialized, just as the experiences of white supremacy and ableism are simultaneously and necessarily gendered. Any acknowledgment of socio-economic inequality (capitalism), for example, that does not also acknowledge its intersectional production, through the systematic sexual violation (cis-hetero-patriarchy) of women of African descent that was sanctioned under the institution of slavery (white supremacy)—cannot remedy the particular ways in which those who experience poverty are actually sub-ordinated (adapted from Crenshaw, 1989).

Justice: Collective self-determination; the obverse of oppression (which involves being required to perform according to goals and norms that one did not participate in establishing). While the distributive paradigm of justice focuses on altering inequitable distribution within given structures, an enabling conceptualization of justice focuses on altering structures that result in inequitable patterns. In particular, it focuses on creating institutional conditions that would enable social groups to exercise collective agency

against structural inequality (for example, through movement organizing). These institutional conditions include decision-making processes, the unwritten rules of culture, and the organization of labor—all of which shape the opportunities available to develop and exercise individual capacities and collective communication and cooperation (adapted from Mukherjee-Reed, 2014; Young, 2011).

Liberalism: A "natural," pre-social rights-based tradition that prioritizes individual entitlements and personal protections and corresponds with Rawls' theory of justice. Mills (2008) argues that what is referred to as liberalism is really racial liberalism. Conceptions of personhood and resulting agendas of duties and government responsibilities have all been racialized and the social contract underlying Rawlsian justice has been an agreement among contractors classified as White to sub-ordinate and exploit nonwhite noncontractors for the benefit of those classified as White. He further argued that to the extent that its racialized nature is ignored, it is perpetuated (Mills, 2008). Guinier similarly contrasts racial liberalism from racial literacy: "the capacity to decipher the durable racial grammar that structures racialized hierarchies and frames the narrative of our republic" (Guinier, 2004, p. 100).

Liberation movements: Sustained collective challenges—in the form of disruptive direct action—by people claiming to speak on behalf of a constituency with a common purpose and solidarity who lack formal representation in and access to decision making against those who do have such representation and access to make publicly visible demands for changes in the distribution or exercise of power or cultural codes. Common purpose and solidarity are rooted in identification with a shared set of interests arising from a shared experience of sub-ordination. Had they representation, such groups would not need a movement, and had they access, they would not need to engage in disruptive direct action (adapted from Klandermans & Van Stekelenburg, 2013; Stoecker, 1995).

Marginalization: Exclusion from the formal economy. Achieved in wealthy countries through selective policies and enforcement of incarceration, immigration, and segregation—including reservations. Marginalized people are those whom the formal

system of labor cannot or will not use and members of this growing underclass are often dependent on the state (adapted from Young, 2011).

Mechanisms: Specific processes that link individuals' ascriptive characteristics to outcomes (adapted from Reskin, 2003).

Neoliberalization: The movement of policies and practices to correspond with neoliberalism/neoliberal philosophy, which extols the virtues of market forces as natural and fair and emphasizes individual agency. Neoliberalization maximizes free trade through increased privatization and competition, as well as deficit reduction through decreased government investment in social, environmental, and economic regulations and protections (adapted from Roberts & Mahtani, 2010).

Oppression: Involves five dimensions: cultural imperialism, exploitation, marginalization, powerlessness, and violence. These forces systemically counteract social groups' ability to develop and exercise their capacities and express their needs, thoughts, and feelings. Rather than the result of individual intent, oppression is systematically reproduced in the structural features of hierarchies and bureaucratic administration. These include market mechanisms, such as the production and distribution of consumer goods, as well as liberal and "humane" economic, political, health-related, educational, and cultural institutions. They take the form of subconscious and unquestioned assumptions, norms, habits, symbols, and media and cultural stereotypes woven through normal, ordinary interactions, reactions, and processes of everyday life. At the same time, the *conscious* daily actions of many individuals contribute to maintaining and reproducing oppression. Such people are often well-meaning—just doing their jobs or living their lives—and do not understand themselves as agents of oppression. Oppression is thus the vast and deep injustice that members of some groups collectively suffer as a consequence of people—who do not necessarily share the same circumstances or consequences, but might—simply following the rules (adapted from Young, 2011).

Orientalism: A particular form of otherizing; the way that Europe and states settler-colonized by Europe, such as the USA ("the West") draw geographic and cultural

boundaries between themselves and the "exotic" peoples and cultures of Africa, Asia, and Latin America ("the rest"). "The rest" continue to be portrayed as infantile, feminized (or hyper-masculine), and primitive in contrast to the modernity and progress of "the West." Denying the former subjectivity and considering them incapable of governing or developing themselves supports the material processes of military, economic, and cultural oppression involved in imperialism by the latter. "The Orient as an object of knowledge is the product of colonial relations of power" (Grossberg, 1996, p. 95). Orientalism's designation and exclusion of Asian Americans as perpetual aliens—threats to national military and economic security—constitutes one of the three pillars of white supremacy in the USA, the other two being the capitalist exploitation and enslavement of African labor and the settler colonialist replacement of indigenous inhabitants of land (Said, 1976; A. Smith, 2016).

Other: Others are members of the marked category, produced in opposition to the unmarked, default category; the marginal(ized) identity as opposed to the center(ed) identity; and typically, the sub-standard relative to the standard. Thus, otherizing takes place dialectically. Individuals and groups are racially normatized as the "White majority" and/or racially otherized as being "diverse," "of color," "minorities." For purposes of this dissertation, racially otherized groups include those that are indigenous to settler colonial states, including those colonized by countries in the Iberian Peninsula, not because such groups constitute a "race" (to the extent that any group does) but because such groups are racially otherized (adapted from Bonilla-Silva, 1997; Doane, 2003; Omi & Winant, 2014).

Powerlessness: The lack of formal authority. The powerless must take orders and rarely have the right to give them. They have little opportunity to develop and exercise skills; have little to no work autonomy; and exercise little creativity or judgment in their work. Even in their private lives, they stand under the authority of professionals (adapted from Young, 2011).

Processual and/or dynamical: Continuously becoming. Everything changes, although at vastly different timescales. Substance has no effect without process, without dynamically

interacting with something else. It cannot even be perceived, because perception itself is a dynamic interaction. Many processes—thunderstorms, flames, epidemics—are not things even if they involve things that come and go during the lifetime of the process. Both identity and oppression are processual and dynamical (adapted from Jaeger & Monk, 2015).

Racialized: Racially differentiated; racialized structural dynamics are those that allocate differential economic, political, social, and even psychological rewards to social groups along racial lines that have been socially constructed. "White" is as racialized a category as "Asian," "Black," and "Native American," although it is racially normatized whereas the latter three are racially otherized (adapted from Bonilla-Silva, 1997; See "Other" for "Racialized 'Other'").

Relationality: Intersubjective dynamics. The concept of relationality challenges the Cartesian conceptualization of human beings discrete, individuated, self-founding subjects and instead asserts the primacy of relationships in the constitution of subjectivity. Inherently relational beings are co-implicated in each other's lives and drawn into responsibility for those with whom we stand in relation. However, human beings often disavow the very thing (inter-relations) that makes them who they are by opting for the Cartesian binary between subject and object. The notion of relationality allows difference, identity, and status/power/privilege to be defined in fluid rather than fixed or static terms (adapted from Hollway, 2010).

Settler colonialism: One of two forms of colonialism—"the process by which European powers reached positions of economic, military, political, and cultural hegemony in much of Asia, Africa, and the Americas" (Shohat & Stam, 1994, p. 13), the other being franchise colonialism or imperialism (defined earlier). Settler colonialism involves direct European settlement of land inhabited and stewarded by indigenous peoples and thus requires their literal genocide and figurative erasure. Settler colonialism constitutes one of the three pillars of white supremacy in the USA; capitalism (which involves the exploitation and enslavement of those of African descent) and orientalism (which

involves the designation of those of Asian descent as the perpetually foreign threat) are the other two pillars (A. Smith, 2016).

Social group: A collective of persons differentiated from at least one other group by their way of living and, as such, exists only in relation to others. Group members have a specific affinity with one another because of their similar experience or shared interests, prompting them to associate with one another more or in different ways than they do with those not identified with the group. Identification with one group arises from the encounter and interaction with other groups within the same society and the experience of differences in their way of living, perspectives, interests, and experiences (adapted from Young, 2011).

Solidarity: Mutuality, accountability, and the recognition of common interests as the basis for relationships among different groups. Rather than being organized around a shared experience of oppression, the practice of solidarity foregrounds collaboration and praxis—active political struggle—among members of different groups who have chosen to organize themselves around shared interests. Difference is central to solidarity in that it is acknowledged and respected as opposed to erased to serve the common interest (derived from Mohanty, 2003).

Stock: Stocks can be referred to as level variables. If flows are verbs, stocks are nouns that are measured at one specific time and represent a quantity existing at that point. Importantly, stocks accumulated from in-flows (minus what was depleted by out-flows)—over time, of whatever duration—in the near or distant past. Indeed, stocks can only be changed by flows. Stock measurements refer to the value of an asset, resource, or type of capital used for production—financial assets, liabilities, wealth, inventories, and education—as of a particular balance date (adapted from Sterman, 2001).

Structural dynamics: Racialized dynamics within and among institutions; the process through which race-based inequality reproduces itself, regardless of the intentions or behaviors of individual actors involved in any particular event. The placement of social

groups in racial categories involves a hierarchy that determines the socio-economic relations and organization of labor among them (adapted from Bonilla-Silva, 1997).

Sub-ordinated: Targeted or systematically disadvantaged by a system of oppression. Sub-ordinated groups are labeled as defective or substandard. Sub-ordinated groups are usually said to be innately incapable of performing preferred societal roles, a characterization they may internalize. While groups super-ordinated by systems of oppression do not understand the complexity of sub-ordinated groups' experience, sub-ordinated groups are saturated with information about the super-ordinated groups because the complexity of their lives and their worldview are fully explored in multiple media and popular culture. Moreover, sub-ordinated groups must understand super-ordinated groups for survival, which often means not responding to oppressive behavior directly, especially individually, in an attempt to avoid physical harm or even death. Sub-ordinated groups thus develop covert ways of resisting or undermining the power of super-ordinated groups. Another strategy is to avoid the indignities and erasure of identity by avoiding engagement with super-ordinated groups entirely, putting them at a disadvantage when they must engage with them, because they have not learned how.

Super-ordinated: Superior in rank, position, or power relative to someone or something else; belonging to a higher order or category within a particular system of classification (OED Online, 2016). Almost everyone is super-ordinated by some systems and sub-ordinated by others. Groups super-ordinated by systems of oppression tend to

> … receive greater economic remuneration and access to better occupations and/or prospects in the labor market, occupy a primary position in the political system, are granted higher social estimation…, often have the license to draw physical…as well as social…boundaries between [themselves] and other [groups]…. (Bonilla-Silva, 1997, pp. 469-470).

Systemic racialization: A set of social relations and practices based on racial distinctions that develops at all levels (internalized, interpersonal, institutional, structural) during and after a society becomes racialized. The racialized effects of these relations, interacting across levels and compounded over generations, produce and normalize racial categories—including the equation of whiteness with ownership, decision making, and

other forms of power and the equation of racial otherness with deficits in resources, knowledge, and agency (adapted from Bonilla-Silva, 1997; Doane, 2003).

Violence: Less about specific acts of violence themselves than about the threat of potential violence associated with particular systems of oppression that tends to keep particular groups sub-ordinated. Even though members of super-ordinated groups sometimes experience violence simply because of their group classification or membership, they know that if the incident goes to trial, they will likely be supported by a justice system that was created by, is composed of, and tends to be sympathetic toward members of their group (adapted from Young, 2011).

White supremacy: The social, economic, and political system dominating the planet for the past several hundred years, which has resulted in today's racialized distributions of economic, political, and cultural power and which has attained a level of autonomy. White supremacy in the USA consists of three "pillars": settler colonialism/genocide, capitalism/commodified labor, and orientalism/imperial war. The concept focuses attention on the racial dimension of oppression, although it is not being claimed that this is the only dimension or that there are no other systems of super-ordination and sub-ordination (adapted from Mills, 1994; A. Smith, 2016).

Whiteness: Theorized as early as 1910 by W.E.B. DuBois as the unmarked, natural, and expected way of being human; the norm against which other groups' identities are produced and compared. Historically codified by law,[27] classification as White for hundreds of years meant the difference between owning property and being property. Because whiteness continues to be continually associated with unquestioned entitlement, authority, and legitimacy and because—like any form of property—its value rises with its exclusivity, its possession must be actively secured and indeed shored up through everyday decisions and interactions (adapted from Harris, 1995; Nylund, 2006; Rabaka, 2007; Todd, 2011).

---

[27] "Children got by an Englishman upon a Negro woman shall be bond or free according to the condition of the mother" (Harris, 1995, p. 279). This deviation from prevailing patrilineal lines of descent ensured the reproduction of White wealth.

CHAPTER 2: LITERATURE REVIEW

The previous chapter drew from critical theories of systemic oppression to offer an alternative to the prevailing, liberal construction of difference in the field of evaluation, which has long noted an incongruence between program participants and program evaluators (e.g., Kirkhart, 1995; Kirkhart, 2010; Kirkhart & Hopson, 2008; LaFrance, Nichols, & Kirkhart, 2012; Symonette, 2004). In the alternative, critical construction, difference is produced and identity is constituted cumulatively and interactively through historical and ongoing relations and processes mediated by asymmetrical structural mechanisms. When the field of evaluation focuses on the observed incongruence while leaving the processes underlying its production unexamined and unchanged, it fails to disrupt the incongruence by normalizing the association of program participants with racial otherness and program evaluators with racial normativity and instead reinforces the authority of whiteness. An enabling conceptualization of justice, in contrast, focuses attention on understanding and potentially changing the structures underlying the oppressive processes through which that incongruence is continually reproduced (see Figure 1). This interdisciplinary dissertation attempts to advance justice so conceptualized by examining the discursive and material role that evaluation as a field plays in the process of producing "race" or racialized difference—racialization—specifically within the U.S. nonprofit industry.

To begin, this chapter reviews literature that illuminates the incongruence noted in the evaluation literature between program participants and program evaluators from the perspectives of systems thinking and critical theories of systemic oppression established in Chapter One. From such perspectives, the observed incongruence does not occur and cannot be remedied in a historical or contextual vacuum, because program evaluation is woven into a larger social fabric (Davis, 1992). Understanding and addressing the incongruence requires understanding the larger social fabric (its context) and its influence

on and interaction with institutional, interpersonal, and internalized dynamics specific to evaluation in the nonprofit industry.

Critical constructions of difference	Enabling conceptualizations of justice
• Difference is inherently relational, within and between groups.	• Identity is based on shared social group interests, which are based on shared experiences of sub-ordination or super-ordination.
• Hierarchical, stratified structures of exchange mean that the flow of capital is asymmetrical. • The asymmetry/hierarchy/stratification is justified through the dialectical production of difference between the groups benefiting from the structural arrangement and those harmed by it.	• Understanding the structural mechanisms mediating multiple groups' sub-ordination can build intersectional solidarity for collective action.
• The asymmetrical flow of capital creates dynamics of super-ordination and sub-ordination among the groups involved.	• Justice lies in social groups' ability to exercise collective agency and produce knowledge to change structural mechanisms.
• Over time, this flow cumulatively results in differences that have ossified to appear "natural" and "common sense," continually justifying and reproducing themselves.	• Changing structural mechanisms changes structurally mediated dynamics, disrupting the reproduction of difference.

Figure 1. Alignment between critical constructions of difference and enabling conceptualizations of justice

Organization of Chapter Two

The first section provides an overview of the racial stratification of U.S. society and the U.S. nonprofit industry surrounding evaluation. The second section reviews U.S. evaluation literature addressing racialized difference that utilizes (1) a contextual/multi-level analysis and (2) a critical perspective. The third and fourth sections provide an overview of systems thinking and critical theories of systemic oppression, which are this dissertation's analytical and theoretical frameworks, respectively. Together, the sections in this chapter (see Figure 2) are intended to provide understanding of the concepts necessary to shift attention from the observed difference—or incongruence, which is

implicitly racialized—to the structural mechanisms and processes underlying its ongoing reproduction.

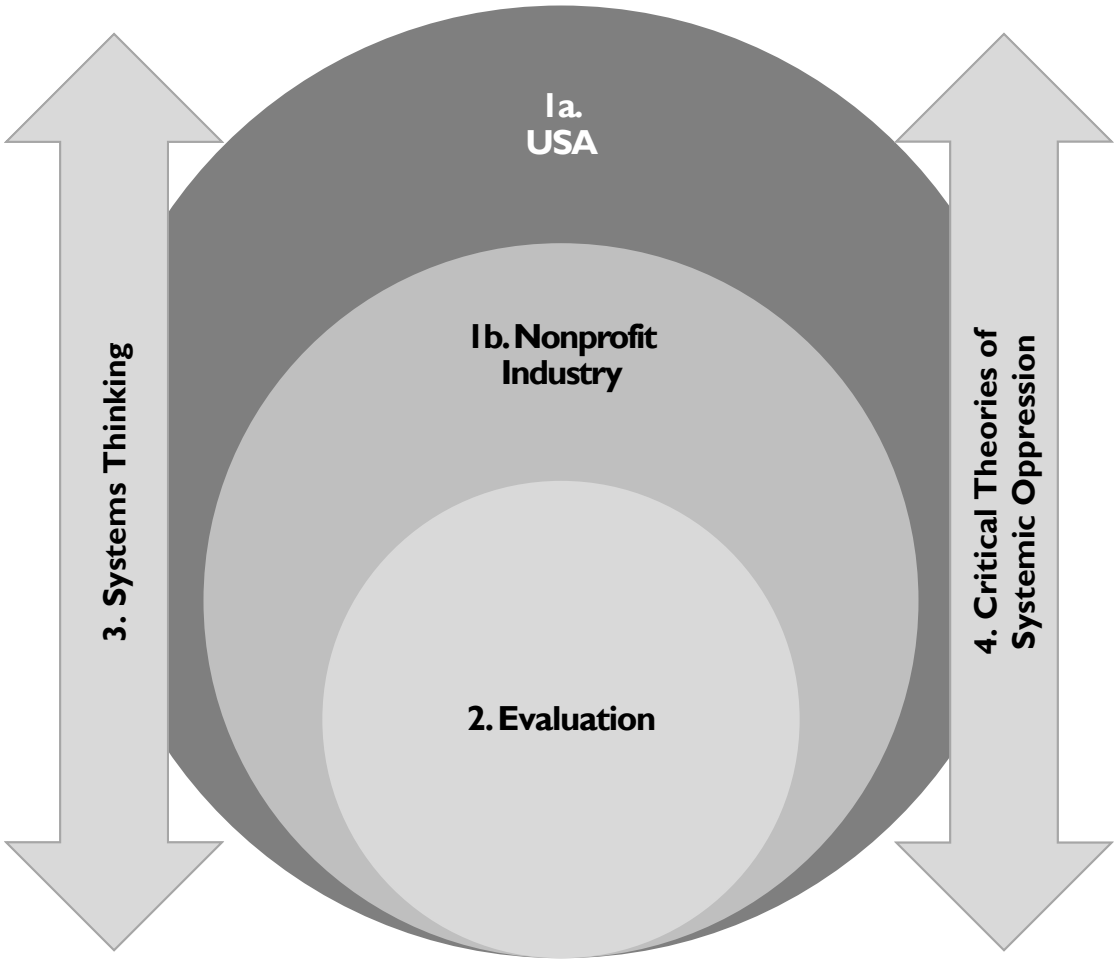

Figure 2. Overall organization of Chapter Two

Racial Stratification

This section provides an overview of racial disparities and disproportionalities in U.S. society and in the U.S. nonprofit industry. The production of racialized difference, observable as racial disparities and disproportionalities, has been structurally encoded into the intergenerational accumulation of wealth—and poverty—since this country's origins. The racialized difference between those whose legal status has enabled them to accumulate wealth and those whose legal status has not only prohibited them from owning wealth, but has also dispossessed them of multiple types of capital—having

capitalized their labor and land—bears implications for the nonprofit industry and evaluation.

Racial Stratification of U.S. Society

Homeownership, household income, employment stability, college education, financial supports by families or friends, and preexisting family wealth account for nearly two-thirds of the racial difference in wealth (Shapiro, Meschede, & Osoro, 2013). All are often portrayed as the result of individual differences and decisions often referred to as personal responsibility (Chiteji & Hamilton, 2005; Darity, 2002, 2005; Hamilton & Darity, 2010; Spilerman, 2000; Personal Responsibility and Work Opportunity Reconciliation Act, 1996). Research suggests, however, that wealth represents a difference between those classified as Black and those classified as White that individuals cannot easily change, certainly within their lifetimes (Lui, Robles, Leondar-Wright, Brewer, & Adamson, 2006; Spilerman, 2000).

For example, the median wealth of households classified as White in the United States of America was 19 times that of those classified as Black[28] in 2009, according to the Pew Research Center (Kochhar, Fry, & Taylor, 2011). While the median adjusted household income has increased slightly for those classified as Black relative to those classified as White (DeSilver, 2013), increases in income do not necessarily result in increases in wealth. Controlling for income leaves as much as three-quarters of the racial wealth gap unexplained (Blau & Graham, 1990). In fact, even in the lowest income quartile—the group that contains the working poor—families classified as Black have 2% of the wealth of families classified as White in the same quartile (Hamilton & Darity, 2010). Similarly, the racialized income gap has narrowed, but the racialized wealth gap has widened. The White-to-Black ratio of median wealth was actually lower (12 to 1) a generation prior. The median net worth for households classified as Black decreased in

[28] Although the racial differences described here apply to groups beyond those that are classified as Black and White, those are the two groups for whom historical data are most consistently available. This is likely due to sample sizes and the Black-White binary underlying the construction of "race" that has influenced how various groups continue to be classified as well as the extent to which and ways in which data for them are collected.

the 25 years between 1984 and 2009, from $7,150 to $5,245, while the net worth of households classified as White increased by almost 20 percent.

Just as decreases in racial income disparities have not led to decreases in racial wealth disparities, decreases in racial educational and employment disparities have not led to decreases in racial wealth disparities. The median wealth of families classified as Black whose head graduated from college is less than the median wealth of families classified as White whose head dropped out of high school (Gittleman & Wolff, 2004). Similarly, even among those with high work stability,[29] those classified as White have $40,000 more net worth than those classified as Black and more than $7,000 in net financial assets (Oliver & Shapiro, 2006, p. 120). Among those with moderate work stability,[30] those classified as White average an income of $20,081, a net worth of $20,000, and net financial assets of $500, whereas those classified as Black average $12,070 in income, $1,740 in net worth, and zero in net financial assets (Oliver & Shapiro, 2006).

In studies of wealth accumulation that control for all factors aside from racial classification among comparably situated families in the current century, a Black-White disparity of 14% grows within just five years because of the interaction among so many structures in which racial classification remains encoded (Conley, 2000). Tracing the same households over the 25-year period referenced earlier (1984 to 2009), Shapiro, Meschede, and Osoro (2013) found that the total wealth gap between families classified as African American and as White nearly tripled, increasing from $85,000 in 1984 to $236,500 in 2009. The increase in wealth disparities over the last generation, particularly between those classified as Black and those classified as White with comparable educational and income levels, is noteworthy when considered in relation to the dramatic growth of philanthropic assets, the nonprofit industry, and human service programs (both nonprofit and for-profit)—many of which provide basic needs and employment

[29] Defined by the study as a maximum of four of the last 39 weeks without a paying job for the household's most experienced worker (Shapiro, 2004)

[30] Defined by the study as five to 34 weeks of not working over the course of a year (Shapiro, 2004)

services—during the same period (Boris & Steuerle, 2006; Grønbjerg, 2001; Salamon, 2002). This industry is discussed in the next segment.

Overview of the Nonprofit Industry

The U.S. nonprofit industry can be traced at least to Benjamin Franklin's organization of a young men's society, volunteer fire company, library, and private academy in the mid-18th century (Hall, 2006). While this segment does not offer a comprehensive history of the industry, it does provide an overview of its origins, which are closely tied to the racialized origins of the USA. It begins, however, by illustrating a pattern in which racialized difference manifests at multiple levels. Those who lead, govern, and fund nonprofit programs and organizations are disproportionately classified as White, while those participating in nonprofit programs are disproportionately classified as not (LeRoux, 2009).

"Nonprofit" has become synonymous with "human services" and "charity" (Grønbjerg, 2001). However, nonprofit organizations and programs include more than those providing human and social services, and human services include more than the administration of government assistance programs and the provision of material assistance or basic needs. Human and social services may be provided by government, 501(c)3 agencies, or for-profit actors—the latter two of which are often contracted to do so by government. Organizations classified under 501(c)3 as nonprofit typically also rely on earned income—some of which comes in the form of client fees, some of which are underwritten by vouchers from for-profit insurance companies or government programs—to administer particular services. Finally, nonprofit organizations are funded by grants from philanthropic foundations and government as well as donations from individuals. Indeed, nonprofit organizations, wealth, and the U.S. government are all intricately tied together conceptually as well as through the exchange of financial, social, human, and knowledge capital. They are thus sometimes referred to as comprising an industrial complex (Gilmore, 2007; Grønbjerg, 2001; Guo, 2010; Hall, 2006; INCITE!, 2007; Rodriguez, 2007; Salamon, 2002; Salamon & Sokolowski, 2006).

Racial Stratification of the Nonprofit Industry

This section establishes the racial stratification of the philanthropic and nonprofit industry. There is considerable acknowledgment in the academic literature that "most nonprofits lack…an explicit group identity base, yet they count large, often disproportionate numbers of racial minorities among their clientele" (LeRoux, 2009, p. 742). However, the academic literature's limited critical consideration of difference, coupled with most nonprofit organizations' limited resources devoted to collecting accurate data, means that information regarding the racial classification of individuals at any level of the industry is also limited. As such, this segment relies partially on administrative data regarding racial classification from government public assistance programs—which are routinely even if imperfectly collected and disaggregated—as an indication of the disproportionalities among their program participants. It relies on limited and often outdated industry literature for data regarding disproportionalities among nonprofit leadership, governance, and funders.

Disproportionalities in human services participation. The relative lack of access to wealth in the form of home equity, savings, and family support established in the previous segment among those in the USA classified as Black, in particular, means that they must often address sudden changes in income or expenses through public assistance programs (Oliver & Shapiro, 2001; Spilerman, 2000).

> In times of economic crisis the wealth principal can be consumed—which is hardly the case with human capital…. [This] point has particular relevance for low income families. Even modest levels of financial assets, which normally provide only a small addition to total income, can cushion a family from the economic shock of illness or job loss, enabling a home mortgage, car loan, and other bills to be paid for a number of months and thereby preventing a temporary loss of employment from snowballing into a wider crisis for the family. Households with few financial assets, especially African-American [*sic*] families (Conley, 1999; Oliver & Shapiro, 1995), are particularly vulnerable to such economic dislocations. (Spilerman, 2000, p. 500)

Perhaps not surprisingly, then, in 2012 members of all racially otherized groups [31] participated in public assistance programs[32] at rates higher than that of those classified as non-Hispanic White. Nearly 42 percent of those classified as Black, 36.4 percent of those classified as Hispanic, and 17.8 percent of those classified as Asian or Pacific Islander participated in one or more government-funded public assistance program for at least one month, compared to 13.2 percent of those classified as non-Hispanic White who did so (Irving & Loveless, 2015). Studies suggest that those participating in nonprofit service programs similarly disproportionately represent groups that are racially otherized (Bandyopadhyay & Pardasani, 2011). The wealth that even average families classified as White accumulate allows many to withstand emergencies without long-term participation in human service programs, while the wealth that families classified as Black manage to accumulate only comes close to what is needed for emergencies (Oliver & Shapiro, 2001; Spilerman, 2000). Indeed, eligibility in many government assistance programs hinges on the absence of wealth in the form of savings or property (Oliver & Shapiro, 2006; Spilerman, 2000).

Disproportionalities in nonprofit leadership. In contrast with nonprofit program participants, the National Urban Fellows reported that in 2012, "the vast majority (88 percent) of nonprofit executives are of White, non-Hispanic heritage" (p. 5). Importantly, research suggests that the racial representation of program participants at leadership levels is more than symbolic.

> Racially representative organizations display increased efforts to provide political education to their clients as well as increased efforts to mobilize them to take action on issues clients have a stake in and help them assimilate as civic participants [which] has important implications for the role nonprofits might play in restructuring the imbalance of influence in larger political and policy-making systems… Persistent disparities in political participation compromise democracy, limit civic discourse, and contribute to inequalities in policy outputs. Some types

---

[31] Rates of participation for those classified as American Indian or Alaska Native were not included in the Census report cited.

[32] These programs provide cash or noncash assistance to individuals and families whose income and/or assets fall below specified thresholds and include General Assistance, Housing Assistance, Medicaid, Supplemental Nutrition Assistance Program/Food Stamps, Supplemental Security Income, and Temporary Assistance for Needy Families.

> of nonprofit organizations may be effective at helping to correct these disparities if their governance is constituted in such a way that adequately represents the people served. (LeRoux, 2009, pp. 757-758)

Writing specifically about the effects of neoliberal privatization in human services in the mid-1990s, including commercialization in the form of increased reliance on fees and service charges, Grønbjerg notes that "the capacity of minority agencies, often relatively new and undercapitalized, to compete with large commercial providers entering the human service field has become one area of concern among policy makers and ethnic and religious community leaders" (Grønbjerg, 2001, p. 290). While what exactly constitutes a "minority agency" is unclear, what constitutes its contrast is clear:

> Most likely, large, multi-service agencies will do better because they have flexibility and discretion in how to allocate their many funding sources among program activities. *They have also powerful board members, strong connections to political actors, and close linkages to lobbying organizations*. (Grønbjerg, 2001, p. 293, emphasis added)

Disproportionalities in nonprofit governance. According to BoardSource, "[b]oard members' racial and ethnic backgrounds closely mirror that of the CEO… Nearly 30% of all nonprofit boards report that 100% of their members are Caucasian [*sic*] with no other racial or ethnic representation" (2012, p. 9). Rather than a snapshot, this figure appears consistent across studies, reflecting "… persistent levels of inequity among African American, Asian American, and Latino populations as documented by the …Nonprofit Governance Index over the past 18 years…." (BoardSource, 2012, p. 9)

The Urban Institute similarly found that on average, 86 percent of nonprofit organization[33] board members are classified as White non-Hispanic (Ostrower, 2007). Even among nonprofit organizations who identified their client population as more than 50 percent African American or Black, 18 percent have no board members whom they identify as African American or Black (Ostrower, 2007).[34] Of more than 500 New York-

---

[33] The cited study was based on Internal Revenue Service 990 forms and thus represents organizations with annual receipts of at least $25,000.

[34] Among organizations whose client population was identified as 25 to 49 percent African American or Black, 36 percent had no board members identified as African American or Black (Ostrower, 2007).

based nonprofit organizations surveyed by Philanthropy New York, nearly 199 self-identified as "minority-*led.*"[35] But of those 199, 37 percent did not have a racially otherized individual serving as executive director. Some of the nonprofit organizations with CEOs classified as White who chose to identify themselves as "minority-led" said they did so because at least half of their board or staff members (the overwhelming majority of whom occupy service and administrative roles) were "minorities." Others said they identified their organizations as such because they are led by those who identified as women, immigrants, LGBTQ individuals, or people with disabilities. Still others said they identified their organizations as such simply because they *served* "minorities" (McGill, Bryan, & Miller, 2009).

Had the above result—in which more than a third of organizations considered themselves "minority-led"—not been probed, it may have masked the extent to which organizations in one of the most multi-racial settings in the USA were led by members of racially otherized groups. More importantly, though, suggesting that an organization that serves racially otherized groups is led by them shows the extent to which nonprofit staff either fail to understand or fail to acknowledge the hierarchical nature of both racialized difference and nonprofit service delivery.

Disproportionalities in philanthropic leadership. According to the *2014 Grantmakers' Salary and Benefits Report* (a proprietary report cited in Bain & Barnett, 2016), of approximately 900 private, public, community, and operating foundations as well as corporate grant-makers and direct giving programs that responded to the Council on Foundations' biannual survey, 92 percent of chief executive officers identified themselves as White.[36] Cohen describes the small number of racially otherized individuals in leadership positions at foundations as follows:

---

[35] The study left this term open for respondents to interpret, asking them to share the reasons underlying their identification as part of the study.

[36] This is consistent with the previous year and up slightly from approximately 91 in 2011 (Cohen, 2012; Prest, 2014). In 2011 and 2013 (data for 2014 are unavailable), 76 percent of full-time staff and 85 percent of trustees were classified as non-Hispanic Whites (Cohen, 2012; Prest, 2014). In 2011, members of racially otherized groups comprised 12.9 percent of executive vice presidents, nine percent of chief financial officers, and 10.5 percent of chief investment officers (Cohen, 2012).

> [R]acial and ethnic minorities are more likely to be CEOs or CGOs in larger foundations. They are also less likely to be employed in professional positions in smaller foundations, and (perhaps not surprisingly, *given the distribution of wealth by race in the U.S.*) less likely to be found in family foundations (which frequently employ their own family members). (2012, paragraph numbers 6 and 7; emphasis added)

Socio-economic Relations Within and Between Groups. The relevance of Cohen's parenthetical comments regarding the distribution of wealth extends beyond the staffing structure of family foundations. While the level of wealth involved in many philanthropic initiatives is beyond the experience of most Americans, whether racially otherized or racially normatized as White, the nearly exclusive representation of those classified as White in positions to influence the direction of nonprofit organizations through their funding streams and established outcomes must be considered in relation to the disproportionate representation of those classified as White in nonprofit leadership and governance. Furthermore, this shared social group status must be considered in relation to the disproportionate representation of racially otherized groups among nonprofit program participants:

> What is often not acknowledged is that the same social system that fosters the accumulation of private wealth for many whites [*sic*] denies it to blacks [*sic*], thus forging an intimate connection between white wealth accumulation and black poverty. Just as blacks [*sic*] have had "cumulative disadvantages," many whites [*sic*] have had "cumulative advantages." (Oliver & Shapiro, 1997, p. 5)

Wealth, philanthropy, and the nonprofit industry. The nonprofit industry has long served as a battleground in terms of its relationship with wealth, the accumulation of which the previous segment established is racialized in the USA. Nonprofit organizations have served as both a vehicle for and a beneficiary of the racialized accumulation and preservation of wealth in the USA. "Many of the great fortunes of Boston, New York, and Philadelphia philanthropists were derived from direct or indirect participation in the slave economy" (Hall, 2006, p. 40). In addition to the use of enslaved labor, commodity agriculture and other sources of philanthropic wealth such as the steel and railroad industries relied on the clearing of indigenous land and exploitative use of immigrant labor (Chan, 1991; Espiritu, 1992). Voluntary associations and organized collectivities—

charitable, educational, and religious institutions that are the precursors of what are today called nonprofit organizations—received much of the bounty derived from enslaved labor, indigenous land, and immigrant labor directly, but also invested their endowments in the steel and railroad industries, allowing those industries to broaden their base of capital and risk in a newly developing economy. Additionally, associations and collectivities provided an option for professional and commercial élites to reassert their influence in a democratic context where wealth was no longer directly tied to political influence as it had been in England. Importantly, however, nonprofit organizations have also served as vehicles for challenging slavery, segregation, and ongoing inequity. The industry's relationship with both wealth and with government has implications for liberation movement organizing among members of otherized groups—rooted in conceptualizations of justice—and for the (re)production of systemic oppression as well as racialized difference more specifically.

Amid the growth of immigrant populations at the turn of the century, associations and collectivities helped preserve class distinctions, with élite private universities limiting the admission of Jews and Catholics and legal and medical professions raising educational standards for admission to the bar and to hospital privileges in an attempt to exclude non-Protestants (Hall, 2006, p. 43). In response to exclusionary tactics by Protestants, Catholics and Jews continued establishing their own hospitals, social agencies, clubs, and national efforts at fundraising well into the next century (Hall, 2006). African American churches had also played a role organizing businesses and a large array of associations and collectivities, including lodges, mutual aid funds, and schools (Hall, 2006; Stanfield, 1993) since before the Civil War. After the war, Booker T. Washington's philosophy of self-sufficiency among ex-slaves involved those classified as Black creating their own economic and social institutions to compensate for their exclusion from institutions and communities designed by and for those classified as White. In the 1920s, those classified as Black traveled to northern cities, where they created communities with churches, associations, and collectivities. While those classified as White had "linked investment capitalism, organized informal and formal political

processes, and organized citizenship participation traditions" (Stanfield, 1993, p. 146), including associations and collectivities since the time of the Civil War, the institutions developed by those classified as Black were prevented from establishing relationships with investment capitalism and government, which are pivotal to mainstream associations and collectivities (Hall, 2006; Stanfield, 1993) in much the same way that individuals classified as Black were excluded from mainstream economic and political life.

While recent studies are lacking and analyzable data are not publicly available, an analysis of total giving by the nation's top 24 independent foundations found that nonprofit organizations that are led by members of groups that are racially otherized[37] received only 3.6% of all dollars and 5.5 percent of all awards granted in 2004. Eleven foundations in the sample awarded less than three percent of grant dollars, and five of the foundations invested less than one percent of grant dollars in organizations led by members of racially otherized groups (Dueñas, Cano, & Mayorga, 2006).[38]

The private-public dichotomy. The U.S. nonprofit industry's roots lie in the early 18th century, when John Locke's ideas about limited government were merged with the spiritual ideas of Calvinism. The result was an emphasis on the "spiritual sovereignty and moral agency of the individual" (Hall, 2006, p. 34), as well as an increasingly politicized clergy that stimulated community-level activity and political engagement through voluntary associations and organized collectivities. After the Revolutionary War, tension arose regarding those same associations and collectivities—in particular between "voice and equality, …majoritarian decision making and…individual rights" (Hall, 2006, pp. 35-36)—that continues today and underlies the distinction between liberal constructions of difference and those rooted in critical theories of systemic oppression.

---

[37] This study defined "minority-led" as those whose staff is 50 percent or more "minority"; whose board of directors is 50 percent or more "minority"; and whose mission statement and charitable programs aim to predominantly serve and empower "minority" communities or populations.

[38] Two years prior, the same institution found that organizations led by members of racially otherized groups received 3 percent of grant dollars (2 percent of grants) awarded by the nation's top 25 community foundations (Aguilar et al., 2005)

In the early 19th century, while there was not yet a legal infrastructure to regulate the amount of economic and political power that charitable trusts could amass, "private initiatives professing to benefit the public" (Hall, 2006, p. 36) were discouraged. An 1817 court case involving Dartmouth College that crystallized this tension between private and public purposes, however, secured the individual rights of donors at the federal level. Thereafter, specific states varied in how broadly they interpreted associational activity and charity. "Where charities and tax laws favored private initiatives… privately supported schools, colleges, and charities were founded in great numbers. [Elsewhere], … state universities and public hospitals…were established instead" (Hall, 2006, pp. 37-38). Importantly, however, evidence suggests that strong government services and strong nonprofit services have not acted as substitutes for one another. In fact, they have depended on each other to degrees that have varied over time (Grønbjerg, 1987, 2001; Ostrander, 1989; Salamon & Sokolowski, 2006).

Interdependence among wealth, government, and the nonprofit industry. Despite prevailing narratives that contrapose what is individual and private with what is structural and public, wealth, government, and the nonprofit industry have mutually benefited from social, economic, and political relations characterized more by interdependence than by independence. The three sectors have increasingly exchanged economic, financial, structural, relational, human, knowledge, and cultural capital since the Civil War, after which Reconstruction required integrating newly emancipated slaves with no education, property, or jobs into the U.S. economic and political system.

Religiously-based efforts, which used volunteers, focused on proselytizing. Secular associations and collectivities focused on reorganizing the South along the model of New England's multi-racial and religiously diverse civil society (Hall, 2006). While both increased, the latter increasingly professionalized, such that by the end of the 19th century, addressing poverty through secular social welfare policies became a full-time occupation. Although the professionalization that accompanied increased interdependence with wealth and government initially shifted the focus of charity from changing individuals experiencing poverty to changing the conditions that create

poverty—an effort that wealth supported by investing in academic disciplines focused on the social sciences, including the type of research that became evaluation—the resulting nonprofit industrial complex has grown increasingly distant from any membership base. It has largely excluded the participation and interests of racially otherized groups, specifically.

By the early 20th century, for-profit corporations combined grant-making with active involvement in the fields they proposed to subsidize. Hall summarized that "[p]hilanthropically supported institutions would play key roles in both moderating the excesses of capitalism and at the same time expanding its reach into every aspect of public and private life" (2006, p. 48). They invested in advertising as they underwrote education in home economics and shop courses that familiarized generations of students with their products. This role served government's public health interests and provided opportunities for their executives to assume leadership roles on the boards of the first modern philanthropic foundations, which had by then been established, and of universities.

The relationship between wealth and public life was not without controversy. Concern regarding philanthropic foundations' ability to influence public policy caused them to retreat from solving social problems through direct political action. In addition to focusing on less controversial topics like healthcare and education, they began influencing public policy through intermediary associations and collectivities (now known as nonprofit organizations). Their efforts at scientific philanthropy—identifying and solving the root causes of social problems rather than their symptoms—were highly influential in government policies and the teaching and research agendas of colleges and universities. They advocated

> … studying conditions, making findings available to influential citizens, and mobilizing public opinion to bring about change. This relationship between academic experts, professional bodies, business, and government would become the paradigm of a new kind of political process… (Hall, 2006, p. 47)

In *American Individualism* (1922), Hoover "envisioned a society self-governed by dense networks of associations working in partnership with government to advance

public welfare by combining the pursuit of profit with the higher values of cooperation and public service" (Hall, 2006, p. 49). Ideas in the book formed the basis for the first phase of the New Deal, whose federal tax policies encouraged private support for charitable institutions. After World War II, in response to the rise of Nazism and communism, efforts to privatize public services and devolve government responsibilities to states and localities began. Hall cites Putnam (2000) in describing post-war associations and collectivities as having "no social dimension: members seldom if ever met face-to-face, individually or collectively. Membership became a political and financial act, not a social commitment" (Hall, 2006, p. 53).

By the late 1950s, attention was drawn to the inequities of the tax code, from which the very wealthy benefited, as well as to the effects on the federal tax base—again crystallizing the tension between private and public gain (Hall, 2013). Liberals advanced a form of third-party government in which federal programs were largely carried out through nongovernmental actors. Conservatives—who typically favored private rather than public efforts—also sought tighter regulation of philanthropic foundations in response to the perception that they were engaging in political activity, specifically funding liberal causes. Eventually, however, conservatives began utilizing the nonprofit vehicle to further their own causes. Ultimately, all donor and recipient institutions alike would form a "third," "nonprofit," or "independent" sector whose well-being was essential to the future of U.S. democracy.

By the time of the federal tax policies of the New Deal, government at all levels depended on the infrastructure of privately funded associations and collectivities not only to implement the provision of services, but also to provide policy expertise. Government now found itself in a position of depending on private organizations for both expertise and implementation (Hall, 2006).

> [I]n terms of its political role, the emergent charitable tax-exempt universe of the postwar era differed dramatically from its associational domain of earlier decades. In the past, when national associations, foundations, think tanks, and other philanthropically supported entities sought to influence government, they generally did so as outsiders. In the postwar decades, associations, now enjoying the benefits of charitable tax-exempt status, increasingly became—if not

> extensions of government itself—an intrinsic part of the organizational field of public governance. The relationship between the Brookings Institution and the government which produced the Social Security Act in the 1930s was exceptional. By the late 1950s, *such relationships were becoming routinized not only on the institutional level (with government contracting with think tanks for all manner of policy and technical services) but on the individual level, as professional careers moved individuals from universities to grant making foundations or from business corporations to government agencies and Congressional staffs—and sometimes to elective office*. (Hall, 2006, pp. 53-54, emphasis added)

The unacknowledged equation of "white" with ownership, decision making, and other forms of power (Doane, 2003) helps explain the intensification of relations across nonprofit organizations, philanthropy, government, academia, and research/evaluation—within social groups disproportionately racially normatized as White—and the disintegration of relations between such groups and those that were racially otherized, for example, through a membership base.

Neoliberalism and the primacy of the market. The second half of the 20th century saw a striking shift in the organization of service provision work. While the number of federal civilian employees remained virtually unchanged, the number of state government employees increased from 4.3 million to 14.7 million, and employment in the nonprofit industry increased from 5.6 million in 1977 to 9.7 million in 1994. At the same time, the flow of direct federal subsidies to nonprofit associations and collectivities increased from approximately $30 billion in 1974 to approximately $160 billion in 1994 (Hall, 2006). Before Reagan's cuts in government spending, public perception had been that nonprofits were entirely privately funded. The policy changes that he made brought the question of the industry's viability without government funding to the research fore, revealing the previous extensiveness of government funding (Grønbjerg, 2001; Hall, 2006).

Since Reagan's devolution of many Great Society programs, federal funding has decreased dramatically as part of the ongoing process of neoliberalization, and the tangle of funding involved in the provision of nonprofit services, in particular, makes traditional forms of accountability difficult (Grønbjerg, 2001; Hall, 2006). Nonprofit associations and collectivities increasingly rely on fees, third party payments from government and for-profit contracts, and corresponding for-profit operations. These involve market-like

expectations, contractual performance outcomes, and proprietary management styles (Grønbjerg, 1987, 2001). "[W]hen present and prominent, [the third-party for-profit] tends to set the major tone of interaction…. It should perhaps not be surprising that capitalist institutions dominate in a capitalist society, when given the opportunity to do so (Grønbjerg, 1987, p. 78). Correspondingly, "in the closing decades of the twentieth century, nonprofits would become increasingly entrepreneurial under the guidance of executives trained as management professionals" (Hall, 2006, p. 55). Indeed, Hall notes that critics of the umbrella term "nonprofit" worry that the sanitized language of law and economics (not distributing dividends and serving a class of people rather than individuals) may obscure these organizations' relationship to wealth and power (2006).

Nonprofit organizations and liberal narratives. Racially otherized groups also received a smaller share of foundation support than their demographic representation in all regions of the USA between 1994 and 2001 (Pittz & Sen, 2004). The lack of responsiveness to racially otherized communities' needs could easily be attributed to lower rates of philanthropic giving among them in much the same way that US racial disparities in wealth and racial stratification of the nonprofit industry are attributed to individual deficits. Assuming that monetary gifts to established organizations constitute philanthropic giving—an assumption not without its critics (Stanfield, 1993)[39]—and considering disparities in wealth, especially that the basis of giving is typically liquid assets gained through inheritance (Conley, 2000), one may expect those classified as Black to give much less than those classified as White at every income level. However, "…controlling for income and net worth, donation rates and amounts are probably higher [for those classified as Black] than the corresponding figures for Whites" (Conley, 2000, p. 538). More importantly,

> … it remains an open question as to whether high rates of philanthropy…have been a positive influence on the overall net worths of African Americans. Namely, if African Americans are giving their money away to charitable causes, then they are not keeping it for themselves. In other words, high rates of giving may themselves contribute to lower net worths if the money given does not have

[39] Stanfield further cites Myrdal (1944) as finding that those classified as Black also contribute more time than those classified as White (Stanfield, 1993).

> an even greater positive effect on the net worths of other African Americans. *This, of course, depends on how much the recipient charities focus on this as a goal and how effective they are*. (Conley, 2000, pp. 538-539, emphasis added)

Conley suggests that philanthropic foundations use their resources to build wealth in under-represented communities to "facilitate racial equity in wealth levels—a goal that will have, of course, an indirect impact on the future levels of White and Black donations" (2000, p. 539).

The various ways that philanthropic foundations describe their grants and that the Foundation Center codes them makes quantifying the size and volume of grant-making for racial (and economic, gender, disability) justice, in particular, difficult. "Racial justice" is not an official grant-making category as a description of a grant recipient's purpose in the Foundation Center's Online Directory of grants. However, the Foundation Center can and did count the number of grants made toward "civil rights and social action" within the broader category of "public and societal benefit" in its analysis of grant making by 1,000 of the largest foundations in the USA. Only 10.7% of grants and 13.7% of total philanthropic contributions in 2007 went to nonprofit associations or collectivities that work for "structural change in order to increase the opportunity of those who are the least well off politically, economically, and socially" (Lawrence, 2009, p. 10). Between 2004 and 2011 (the most recent year for which data are available), the percentage of all such grants made by the Foundation Center's top 1,000 grant-makers hovered between one and two percent (Cohen, 2014, p. 40). Within that structural change work, which the Foundation Center terms "social justice," study participants cited constituent-led organizing, leadership development particularly within groups who are racially otherized, and capacity-building as currently under-funded. Funding the areas that study participants named as under-funded would advance an enabling conceptualization of justice in that they would shift the organization of labor and the center of decision making from those classified as White to those who are racially otherized.

To those for whom structural arrangements have disproportionately albeit to varying degrees worked, even racialized poverty may appear accidental—the natural result of atheorized disadvantage, at best, as opposed to the compounded effect of generations of racialized policy. Having a positive experience with—or at least uncritical view of—policy structures, those in positions with power to conceptualize poverty and related problems, to develop the theory underlying corresponding solutions, to manage those programs, and to measure their outcomes have overwhelmingly focused their efforts on programs designed to change individuals experiencing poverty and other social problems rather than on programs or policies designed to change the structural arrangements themselves (Davis, 1992; Lincoln, 1991; Madison, 1992b).

> Foundation leaders are people who hold considerable power and reap significant benefits from the system as it currently is. They are thus more vulnerable to "overlooking the circumstances of economic injustice that made their philanthropy necessary" (attributed to Rev. Dr. Martin Luther King, Jr.) and, indeed, that made it possible. (Greene, Millett, & Hopson, 2004, p. 106)

This tendency may explain the distaste shared among leaders of nonprofit associations and collectivities, philanthropic foundations, government, for-profit business, and research for direct political action—a tool that those excluded from formal decision-making bodies often employ to influence the structure underlying such exclusive processes (Hall, 2006). In the early part of the 20th century, scientific management spawned welfare capitalism, which provided workers with education, health, housing, and other services to boost their productivity and discourage them from joining unions (Hall, 2006).

Speaking about the need and potential for large, long-established, multi-service agencies to advocate for policy-level changes on behalf of their clients, Ostrander expressed concerns about nonprofit association and collectivity board members that are similar to those expressed about foundation leaders.

> I do not wish to exaggerate the potential for social change as a consequence of advocacy by these agencies. They are limited…by conservative forces on their boards, which represent the traditional historical opposition of the voluntary sector to a large and permanent welfare state (Ostrander, 1987). Established, or

> "mainline" voluntary social service agencies should not be counted on to advocate for fundamental social change in the economic and social institutions of society. They are not likely to put forth a truly radical social agenda. (Ostrander, 1989, p. 42)

Summary of Racial Stratification

To summarize, the nonprofit industry has from its inception occupied a contested third space between private and public as well as between individually- and structurally-focused interventions, which are central to understanding the racial stratification of U.S. society and the nonprofit industry. This segment has connected the nonprofit industry's racial stratification to its development in the context of individualism and the racialized accumulation of wealth in the USA. It further explained the industry's reproduction of structural racialization through its failure to support constituent-led organizing for racial justice, both deploying and solidifying whiteness in the process. The next segment of this chapter reviews U.S. evaluation literature that utilizes contextual/multi-level analysis and that directly addresses racialized difference. The latter often questions prevailing narratives of difference—even while, in some cases, manifesting the authority of whiteness.

Racialized Difference in Program Evaluation

As stated in Chapter One and earlier in Chapter Two, the U.S. field of evaluation has long noted an incongruence between program evaluators and program participants in its academic literature (e.g., Kirkhart, 1995; Kirkhart, 2010; Kirkhart & Hopson, 2008; LaFrance et al., 2012; Lincoln, 1991; Madison, 1992b; Symonette, 2004). Concerns about the representation of evaluators from under-represented groups in the field's professional associations were raised as early as 1976, when Marcia Guttentag, Lois-ellin Datta, and Carol Weiss developed the Evaluation Research Society (ERS)—one precursor of AEA—as an alternative to existing associations that were considered to consist disproportionately of "White men" (Gargani, 2011, p. 430). In 1986, ERS merged with the Evaluation Network (ENet) to create the American Evaluation Association. Soon after, AEA President Bob Covert documented the professional association's demographic characteristics and referred to the "small number of minority members" as

"discomforting" (Covert, 1987a, p. 96). Two years later, AEA President Ross Conner described AEA's "record on involvement of people of color in leadership positions" as not "as strong" as its "history of equal involvement of men and women" (Conner, 1989b, p. 79).

Since the 1990s, scholars have often framed the observed incongruence between program participants and program evaluators within the U.S. evaluation literature as a natural result of recent demographic changes (Dunaway, Morrow, & Porter, 2012; Hood, 2014; Kirkhart, 1995; Mertens & Russon, 2000). For example, Kirkhart stated:

> Multicultural validity may be justified as a topic worthy of development in terms of societal relevance, historical tradition, and social justice. To develop the argument of societal relevance, one would point to the changing demographics of our population… (1995, p. 2)

Mertens and Russon described the demographic changes as resulting from growth in groups considered foreign (A. Smith, 2016), although they did not specify whether this growth is a result of immigration or higher fertility rates among those born in the USA of Latin American or Asian descent:

> [I]n the United States, the Hispanic population grew by more than 35% and the Asian population grew by more than 40% during the 1990s. Lester (1999) quotes Hugh Davis Graham, Vanderbilt University historian, as saying that demographic trends are leading to a time when "everybody's a minority." Under these circumstances, it is imperative that evaluators develop intercultural sensitivity if they want to do their jobs effectively. (2000, pp. 280-281)

Hood referred to countries—including those in Europe, like the colony of Ireland, together with settler colonial states like the USA—that he described as being more homogenous until recently:

> The need to increase the number of culturally responsive evaluators globally continues to be critically important as racial and cultural demographics continue to rapidly change in countries that had been historically more homogeneous… (2014, p. 118)

Dunaway, Morrow, and Porter (2012) explicitly tied U.S. changing demographics with AEA's expectation of cultural competence in evaluators:

> As part of its Guiding Principles for Evaluators, the American Evaluation Association (AEA) requires that evaluators develop cultural competencies. Such competence is especially important, given the dramatically changing composition of the United States. The U.S. Census Bureau (2007) reports that racial and ethnic minorities [*sic*], comprising approximately 100 million people, account for about one third of the nation's population. It is estimated that by 2025 ethnic minorities [*sic*] will comprise 40% of all Americans and that by 2050, non-Whites [*sic*] will become the majority (Barrett & George, 2005; Stanhope, Solomon, Pernell-Arnold, Sands, & Bourjolly, 2005).
>
> This growth in minority [*sic*] populations has led to the expectation that evaluators will work effectively with increasingly diverse groups. Our capacity to do this will depend on the acquisition of cultural competence (Hansen, Pepitone-Arreola-Rockwell, & Greene, 2000; Stanhope et al., 2005). (2012, pp. 496-497)

The above framing of the incongruence misses the racial stratification detailed earlier that underlies the over-representation of racially otherized groups among program participants. Perhaps more importantly, it also ignores the racialized policies that underlie the over-representation of those classified as White among program evaluators. These include the displacement and dispossession associated with settler colonialism, exploited and enslaved labor, and war. Specifically, this involves genocide; cultural imperialism; over-surveillance and mass incarceration; residential, occupational, and educational segregation; and exclusionary immigration policies (Chan, 1991; Espiritu, 1992; Lui et al., 2006; Oliver & Shapiro, 2006; A. Smith, 2016). The failure to name indigenous peoples and those of African descent who survived the Middle Passage suggests that the use of social programs is a temporary and natural result of individual deficits such as a lack of experience in the USA, English language proficiency, or transferable skills. It illustrates the way those perceived as perpetual newcomers—regardless of immigration status or generation of migration—are racialized differently relative to those whose disproportionate social program use defies individualized explanations rooted in language, experience in an industrial economy, or generation in the USA (A. Smith, 2016).

The following section on racialized difference in program evaluation reviews two bodies of evaluation literature. The first body, on context-sensitive evaluation and multi-level analysis, addresses elements of systems thinking in relation to racialized difference

by expanding the boundaries around individuals as the unit of analysis. The second body addresses critical theories of systemic oppression by challenging prevailing, liberal constructions of racialized difference.

Context-sensitive Evaluation and Multi-level Analysis

A significant portion of the U.S. evaluation literature that expands the unit of analysis beyond individuals and examines context beyond evaluation practice draws from emergent realist evaluation's integration of contextual variables (Conner, Fitzpatrick, & Rog, 2012; Vo, 2013; Vo & Christie, 2015). Henry and Mark (2003) and Mark and Henry (2004) employ a multi-level analytical framework—individual, interpersonal, and collective—and examine the pathways and mechanisms that might link these levels to postulate how evaluation can trigger use. Their emergent realist evaluation approach relies explicitly on scientific and "common sense" reasoning to reconcile multiple stakeholder perspectives of reality (Greene & Walker, 2001; Henry & Julnes, 1998; Vo & Christie, 2015).

Search and selection criteria. This dissertation's use of systems thinking as its analytical framework and critical theories of systemic oppression as its theoretical framework led to a search within U.S. scholarly evaluation journals for articles that addressed racialized difference using a contextual or multi-level analysis. The search included all five peer-reviewed, academic journals published in the USA that have "evaluation" in their title, including the two flagship journals published in association with the American Evaluation Association. Included in the search were the *American Journal of Evaluation* (formerly *Evaluation Practice*), *Evaluation and Program Planning*, *Evaluation Review*, the *Journal of Multidisciplinary Evaluation*, and *New Directions for Evaluation* (formerly *New Directions for Program Evaluation*). The search was conducted from the journals' webpages from the various journals' inception through December 2017. Items that expanded boundaries around individuals as the unit of analysis were identified by conducting a search for items containing the following terms: context, level, mechanism, pathway, or system. Results were narrowed to those that explored racialized difference, which was usually expressed in cultural terms, and can be

divided into two broad groups: those that focus on context, and those that focus on multiple levels of analysis.

Expanded boundaries and difference. The social context surrounding the evaluation of social programs—including the risks of homogeneity in terms of social group status, interests, and assumptions—has been acknowledged as influential to evaluation, programs, and society since the evaluation field's inception (Levine & Levine, 1977). Fitzpatrick (2012), LaFrance, Nichols, and Kirkhart (2012), Kirkhart (2011), and Lee and Gilbert (2014) connect discussion of expanded boundaries—whether context-sensitive evaluation or multi-level analysis—to difference. Unlike Henry and Mark (2003) and Mark and Henry (2004), these authors acknowledge reality as mutually constituted.

Context-sensitive evaluation. Fitzpatrick (2012) elaborates on the cultural dimension of context initially identified by Rog (2012) by discussing

> … two areas in which context has been more carefully considered by evaluators: the culture of program participants when their culture is different from the predominant one and the cultural norms of program participants in countries outside the West. We have learned much about how the culture of participants or communities can affect evaluation and should continue our learning there. (p. 7)

While Fitzpatrick extends the questions that Rog raises regarding power dynamics to include the poor representation of particular racial and ethnic groups in decision making more generally, she does so without connecting it to systemic oppression or acknowledging the industrial contexts in which evaluators work as racially stratified (Fitzpatrick, 2012; Rog, 2012).

Perhaps because so much of the literature explicitly emphasizing the importance of context comes from cross-cultural and international evaluation, Fitzpatrick calls one stream of scholarship on context "Culture as Context" (2012, p. 13) and links this stream to U.S. demographic changes:

> As the United States has grown increasingly diverse in race and ethnicity, and as educational and social programs often serve people from minority or disadvantaged groups, evaluators have begun to realize that we need to know

> more about the context and culture of the clients served by the programs we evaluate. (2012, p. 13)

When the culture of program participants differs from that of program evaluators, that of the participants is described as representing "minority" or "disadvantaged" groups, and the culture and status of the evaluators are left unstated. Fitzpatrick notes the racial, ethnic, or cultural background of scholars in her review of literature on context only once,[40] when referring to two African American scholars whose work with African American schools and students in the 1940s and 1950s predates by at least 20 years Stufflebeam's CIPP model (1968). Still, it is the CIPP model that marks the onset of Fitzpatrick's section entitled the "Historical Roots of Context in Evaluation," while Brown and Boykin's scholarship raising awareness of "the special needs and history of African Americans" is reserved for Fitzpatrick's second section entitled "Cross-Cultural Evaluation" (Fitzpatrick, 2012, p. 14). The remainder of scholars in Fitzpatrick's review (including some who have self-identified as African American in their scholarship [e.g., Hood, 2000]) are not identified by racial classification, ethnicity, or culture.

Without contextualizing the disproportionality as artificially produced through the U.S. history of settler colonialism, segregation, and exclusionary immigration policies (A. Smith, 2016), Fitzpatrick later alludes to notions of multicultural validity (Kirkhart, 1995; Mertens, 2007, 2013) and states directly:

> U.S. evaluators, who are primarily from white [*sic*], middle-class, educated backgrounds, have recognized that *their* own personal contexts and values influence how *they* see, or fail to see, other cultures. As such, *our* evaluations are invalid. (2012, p. 14, emphasis added)

Fitzpatrick's demonstrated ambivalence can serve as a microcosm of the program evaluation field: Who is "our," and who is "their"? How do racial classification, ethnicity, nationality, and culture relate to each other? When are they marginal, salient, or central, and for whom?

---

[40] In Fitzpatrick's section entitled "Contextual Factors Arising from Literature on International Evaluations," she refers to David Nevo (1982) as "an educational evaluator in Israel" (Fitzpatrick, 2012, p. 16).

Similarly, the "special needs" (2012, p. 14) that Fitzpatrick refers to are presented as if they are specific to the "…history of African Americans" (p. 14). Their placement in the article's section entitled "Cross-Cultural Evaluation: Culture as Context" as opposed to the one entitled "Historical Roots of Context in Evaluation" suggests that the "needs" are marginal rather than central to both the history of the USA and the field of evaluation and directly related to the very different experience of those racialized as White. They are attached to African Americans—"*of* African Americans"—rather than presented as disparities *between* those racialized as African American *and* those racialized as White. Moreover, they are tied to African American culture when it is African Americans' racial classification that is salient. The "special needs" were produced and continue to be reproduced by laws, policies, and practices contingent on the classification of Africans in an oppressive system of racial stratification—namely, the enslavement and segregated education and employment of those classified as Black. In other words, underlying the incongruence between program participants and program evaluators is a racialized context rather than, or at most in addition to, a cultural context.

Culture overshadows material inequity in LaFrance, Nichols, and Kirkhart's discussion of Indigenous Evaluation Framework's (IEF) prioritization of context (LaFrance et al., 2012). The authors argue that beyond being as important as rigor and stakeholder needs in producing actionable evidence, which is the argument to which they are responding, context actually defines rigor, stakeholder needs, and even actionable evidence (LaFrance et al., 2012; Rog, 2012). While they make no mention of racial classification (considering their focus on IEF and the distinction between indigeneity and "race," despite racialization of the former), they tie the need for IEF directly to a larger context of systemic oppression—namely, settler colonization—through their substantive discussion of sovereignty (A. Smith, 2016):

> … honoring sovereignty recognizes *nationhood*. It reaffirms place, community, *culture*, language, and *political presence*. Indigenous evaluation methodology is explicitly related to *nation building* (Robertson, Jorgenson, & Garrow, 2004). It seeks to contribute to the health and wellbeing of the community first and foremost rather than to generalization to larger audiences or other settings. (LaFrance et al., 2012, p. 72, emphasis added)

IEF, in the sense that it honors sovereignty, is thus a decolonizing praxis—the logical extension of which would be an end to the settler colonization of indigenous lands and peoples (Cavino, 2013; Kerr, 2006), including the nonprofit industry and field of evaluation.

The authors similarly cite Kovach (2010, p. 30) as arguing that in indigenous systems of knowledge, "knowledge is *neither acultural nor apolitical*" (LaFrance al., 2012, p. 62, emphasis added). And finally, they conclude by differentiating the value that IEF places on context from that which Rog (2012) places on it. Building on Rog's metaphor of performance, they state:

> Rog has advanced considerations of context from background to foreground, from a character role to a leading role in our evaluations. An indigenous framing of evaluation illustrates how *context* in fact writes the script and staging and directs the entire performance. (LaFrance et al., 2012, p. 73, emphasis added)

Deviating ever so slightly from the last line of the chapter, however, the title of the chapter is "*Culture* Writes the Script: *The Centrality of Context* in Indigenous Evaluation."

LaFrance, Nichols, and Kirkhart's slippage between culture and context, which are named approximately the same number of times in the body of the chapter—although context encompasses both cultural and political dimensions among others (Fitzpatrick, 2012; LaFrance, et al., 2012; Rog, 2012; Thomas, 2004)—demonstrates a certain ambivalence. While political context is described in the chapter as at least as important to IEF as cultural context, the authors' interchangeable use of "culture" and the higher order category of "context" (Thomas, 2004) serves to privilege cultural context relative to political context, romanticizing culture by suggesting that it is pre-social and essential rather than contextually constituted.

In addition, the authors demonstrate a certain ambivalence between cultural context and political context (as in nationhood and sovereignty), which are distinguished in Rog's framework even if inseparable to those living under conditions of cultural imperialism—particularly to colonized peoples who continue to live under occupation (LaFrance et al., 2012; Madison, 1992b; Rog, 2012; Thomas, 2004; Young, 2011).

Continuing to keep alive ancestral cultures, languages, and spiritualities in a settler colonial state—one that remains occupied by those who would benefit materially (in an economy where land is wealth) from the disappearance of indigenous peoples and practices—itself constitutes an act of political resistance, blurring the lines between political and cultural context (A. Smith, 2016). IEF's prioritization of context—having arisen from a place-based worldview in which life is interdependent and relationally accountable—similarly constitutes a threat to the settler state, which is founded on colonial relations of super-ordination and sub-ordination, as well as on capitalist relations of production and consumption. The practice of IEF thus constitutes the enactment of self-determination and sovereignty, which involves but extends beyond culture. The politically subversive potential of IEF, however, is compromised when it is portrayed principally as cultural rather than at least equally political.

Multi-level analysis. Kirkhart distinguishes between individual and organizational levels of culture in her use of a multi-site evaluation to illustrate dimensions of influence and culture (Kirkhart, 2011). She asserts that

> … [c]ulture leads us to reflect on issues of power and ownership, of self-determination and autonomy, and of proprietary knowledge that is not intended to be exported or shared. Culture also leads us to notice whose agendas are being served and whose interests are ignored or impeded when evaluation exerts influence. (Kirkhart, 2011, p. 81)

She goes on to state:

> Culture teaches us to notice the consequences of evaluation influence and how such consequences relate to broader issues of equity and social justice…. Culture cautions us to temper our understanding of the apparent success of our evaluation implementation (process-based influence) and our dissemination/utilization efforts (results-based influence) with a careful reflection on consequences…. Culture draws attention to historically relevant information, extending the time frame of influence. (Kirkhart, 2011, p. 82)

Despite her multi-level analysis, Kirkhart (2011) fails to specify the pathways and mechanisms—outlined extensively in Henry and Mark and Mark and Henry's framework—through which culture leads, teaches, cautions, or even draws attention, however (Henry & Mark, 2003; Mark & Henry, 2004). Instead, she concedes that culture

is "messy" (Kirkhart, 2011, p. 78). While Kirkhart admonishes against the tendency to consider culture "out there" rather than "in here" (p. 74), her treatment of culture as the subject—and that, too, as the guide—serves to exalt it to mythical proportions, precisely as the "something different, 'special,' foreign, or distant" (p. 74) that she cautions against.

Discussing AEA's pipeline program (the Graduate Education Diversity Initiative or GEDI) as one avenue for doing the latter, Lee and Gilbert (2014) note that while increasing the number of people from traditionally underrepresented groups in any profession is necessary, it is insufficient for improving racial equity in the USA. They allude to the larger context of racial stratification without stating it explicitly:

> This assumption supports the line of reasoning that increasing the number of people in a traditionally underrepresented group *could have positive outcomes on issues associated with the disproportional representation in the first place* (Deaux & Ullman, 1983; Toren & Kraus, 1987). This assumption also demanded a particular response from the evaluation profession—a larger pool of racially, ethnically, and culturally diverse evaluators. This demand compelled Dr. Millett [of the Kellogg Foundation] to fund the BDI. (2014, p. 99, emphasis added)

Lee and Gilbert go on to cite the pipeline literature, echoing Darity, Hamilton, and Stewart's conclusions regarding persistent disparities between those classified as White and those who are racially otherized—even in studies that control for all factors beyond racial classification (Darity et al., 2015). Lee and Gilbert (2014) also cite Collins and Hopson (2007) in their admonition that pipeline programs

> … move beyond the individual and consider the interconnected parts of the ecosystem (i.e., organizational policies and practices, institutional norms) that made it necessary for the programs in the first place and thus play a role in the change process. Focusing on and affecting individuals—the student or intern, in particular—is less threatening than challenging policies and practices that have been in place for generations. (Lee & Gilbert, 2014, pp. 106-107)

Importantly, Lee and Gilbert's 2014 analysis focuses specifically on AEA's pipeline program. Analyzing the effects of the evaluation field and its professional association's similarly individual- and practice-oriented efforts to increase the cultural competence and inclusiveness of evaluators more generally would require broadening the

unit of analysis in ways similar to that of Lee and Gilbert—beyond individual differences, intentions, and behaviors within the realm of evaluation practice to consider the racial stratification of the industries within which evaluation plays a pivotal role. Lee and Gilbert's multi-level analysis of AEA's pipeline program (2014) is rare in its implicit application of the social-ecological model (which they refer to as an ecosystem) in conjunction with critical theories of systemic oppression.

The Limitations of Boundary Expansion. Multi-level and contextual studies of difference can better illuminate the incongruence observed in evaluation if they are informed by systems thinking and critical theories of systemic oppression. What distinguishes Henry and Mark's multi-level framework from both is their means for reconciling multiple stakeholder perspectives of reality without acknowledging that both scientific and "common sense" reasoning are socially located; moreover, reality is inevitably influenced by the perspective—or interests—of the evaluator, which are also socially located (Greene & Walker, 2001; House, 1995; Thomas & Parsons, 2017). What distinguishes the context-sensitive or multi-level analysis of Fitzpatrick (2012), LaFrance, Nichols, and Kirkhart (2012), and Kirkhart (2011) from those advancing systems thinking and critical theories of systemic oppression is the formers' essentialist characterization of difference as cultural rather than as produced and reproduced through ongoing relations and processes that are mediated through structural mechanisms. Lee and Gilbert combine both critical theories of systemic oppression and a systems orientation that considers multiple levels.

Challenges to Liberal Constructions of Racialized Difference

Although some academic literature in the U.S. transdiscipline of evaluation explicitly draws from decolonizing or indigenous frameworks and feminist theory, only a handful explicitly utilizes critical race theory or postcolonial theory (e.g., Bheda, 2011; LaFrance, 2004; LaFrance et al., 2012; Mertens, 1995; Noblit & Jay, 2010; Parker, 2004;

Sielbeck-Bowen, Brisolara, Seigart, Tischler, & Whitmore, 2002a, 2002b).[41] At the same time, the U.S. evaluation literature includes examples that mention critical theories of systemic oppression or their concepts without necessarily elaborating on or utilizing them. Consequently, this review includes literature that challenges prevailing, liberal constructions of difference, sometimes implicitly and even unwittingly. This takes place through the acknowledgment of systemic oppression. Additionally, it takes place through the explication of difference as produced and the complication—however brief or tangential—of essentialist notions of racialized difference. Finally, it takes place through an emphasis on social group interests and an intimation of the discursive power of evaluation with respect to racialized difference.

Search and selection criteria. Whereas House has written extensively and critically about racialized difference in relation to the evaluation of U.S. educational programs (e.g., 1985, 1990, 1999), this dissertation's focus on social programs in the U.S. nonprofit industry—where wealth plays a pivotal role—led to the exclusion of articles and journals specific to education and assessment, whose funding and mandates differ from those of human and social services. The search also excluded journals specific to public health and substance abuse, which—while increasingly integrated with human and social services—still tend to receive different types of funding and adhere to different sets of professional standards, expectations, and practices (Grønbjerg, 1987). Articles focused on education/assessment and public health were similarly excluded in favor of those that either focused on U.S. nonprofit settings specifically or addressed evaluation more generally.

Because the funding structure of social programs and the specific processes of racialization[42] differ internationally, including from one settler colony to another—even

---

[41] As of April 2017, no articles explicitly used queer or crip theory, although the U.S. evaluation literature includes two articles that explicitly utilize a social-ecological model of disability (Gill, 1999; Lee, 1999).

[42] The social, economic, and political construction of racial categories varies internationally, one indication of which is the difficulty in translating racial categories internationally (or even across decades within a lifetime in single country). "Black" was until recently used in the UK to describe those of South Asian descent, for example, and racial categorization in Brazil and South Africa also contrasts sharply with that in the USA.

within North America—the literature included in this review focuses on the treatment of racialized difference in U.S. evaluation contexts and literature. As such, journals published in Canada, the Pacific, and the UK were excluded from this review. Additionally—again, because of the dissertation's focus on U.S. social programs—individual results that were specific to international development were excluded; considering colonial histories, these may otherwise have dominated the review without necessarily contributing a critical perspective (Chouinard & Milley, 2016). Including them could contribute to the conflation of racialized difference with differences in nationality, culture, and language, potentially reinscribing the implicit association among "America," "white," English, and normativity (Darity, 2005). Importantly, however, authors whose work was selected for review represented various national origins and disciplinary backgrounds.

A search was conducted using variations of terms referring to racialized difference (detailed below) in all five peer-reviewed, academic journals published in the USA that have "evaluation" in their title, including the two flagship journals published in association with the American Evaluation Association. Included in the search were the *American Journal of Evaluation* (formerly *Evaluation Practice*), *Evaluation and Program Planning*, *Evaluation Review*, the *Journal of Multidisciplinary Evaluation*, and *New Directions for Evaluation* (formerly *New Directions for Program Evaluation*). The search was conducted from the journals' webpages, from the various journals' inception through December 2017. While 2018 saw a rise in publications addressing racialized difference, including one *New Directions for Evaluation* edition dedicated to Indigenous Evaluation and a new section in the *American Journal of Evaluation* dedicated to Race, the search did not include articles published in 2018.

Terms specific to racially otherized groups were not included in the search. Rather, searching for variations of "race," "white," and "colony" allowed literature that focuses the gaze on particular racially otherized groups—eliding the inherently relational nature of difference—to be excluded from the results (Hall, 1997; Ward Hood & Cassaro, 2002). This should not suggest that no items selected for review focused on a group that

is racially otherized—some did—but simply that all items selected for review included direct discussion of "race," colonization, or the classification of "white." As such, articles that focused exclusively on racial differences in social phenomena, program participation, or program outcomes; psychometric scales regarding racial identity; practice-based "lessons learned" from work with particular groups who are racially otherized; and group-specific methodological instruments or approaches to analysis were omitted from the review.

The search identified items that fell into three broad categories: (1) those that generally acknowledge systemic oppression based on racial classification; (2) those that explicate the production of difference or raise questions related to intersectionality and the essentialization of racial categories; and (3) those that demonstrate an explicit or implicit understanding of social group identification and corresponding interests, particularly with respect to racialized categories. Within the literature that raised questions regarding social group identification and interests were (a) those that advance notions of self-determination and (b) those that reflect an often-unacknowledged identification with whiteness.

Systemic oppression. Five items generally acknowledge systemic oppression based on racial classification, often in conjunction with oppression based on other dimensions of difference and social group identification (Chouinard & Cousins, 2009; Lincoln, 1994; Mertens et al., 1994; SenGupta et al., 2004). Some of these also analyze essentialism or social group identification. Mertens, Farley, Madison, and Singleton (1994) and Lincoln (1994) both share contributions from scholars drawing from critical theories of systemic oppression based on racial, gender, and ability status, among others, contemplating the implications for evaluation. As such, both also demonstrate an acknowledgment of racially otherized groups as protagonists as opposed to consumers of evaluated services alone. SenGupta, Hopson, and Thompson-Robinson (2004) and Chouinard and Cousins (2009) take a much broader look at identity through the lens of culture. While SenGupta, Hopson, and Thompson-Robinson elaborate on structural racism, Chouinard and Cousins elaborate on colonialism.

Racialized difference as produced. Evaluation scholars engaging critically with racialized difference challenge prevailing notions of identity as pre-social, essential, or fixed (Birman, 2007; Chouinard & Cousins, 2009; Davis, 1992; Evergreen & Cullen, 2010; Ward Hood & Cassaro, 2002). They also attempt to tease apart the relationships among classifications of race and culture as well as other dimensions of difference and social group identification. For example, Davis (1992) points out how, in program evaluators' efforts to explain differential program effects and populations, they often categorize by racial classification unquestioningly, at the point of data coding, without necessarily realizing that "the use of race as a variable in evaluation design and analysis is not always benign" (p. 58). Those engaging critically with racialized difference describe how "program evaluation becomes a social practice that influences how evaluators construct the social realities of program participants and how they analyze results" (Davis, 1992, p. 60), with material effects on participants and others involved in the enterprise (Clayson, Castañeda, Sanchez, & Brindis, 2002; Ward Hood & Cassaro, 2002).

They further describe how racial classification sometimes serves as a proxy variable for socio-economic class status, language, nationality, generation in the USA, documentation status, trust in program staff, and aspects of service delivery that would affect differently-positioned groups differently (Birman, 2007; Davis, 1992). Over-reliance on racial classification can mask within-group variation—including the inevitable identification with other intersecting and sometimes conflicting social groups (some of which engage in oppression and others of which suffer under oppression)—representing a failure to attend to the complexity of the lives of racially otherized groups (Birman, 2007; Davis, 1992; Ward Hood & Cassaro, 2002). The above challenges to essentialized notions of identity and difference rest on an understanding, perhaps implicit, of social groups and associated interests, which are discussed below.

Social group rather than individual interests. As discussed in Chapter One, pivotal to the production of difference is the formation of social groups around shared interests. Shared group interests arise from shared experiences of socio-economic

relations and decision-making processes—including those of program evaluation—that are structurally mediated (Bonilla-Silva, 1997; Chouinard & Cousins, 2009; Young, 2011). Several evaluation scholars have considered shared group interests at multiple levels in relation to racialized difference (Clayson et al., 2002; Evergreen & Cullen, 2010; Freeman, Preissle, & Havick, 2010; Hood, 2000, 2001; Lee & Gilbert, 2014; Madison, 2007; Sturges, 2015). For example, Hood revealed to the readership of *New Directions for Evaluation* that having collectively experienced democracy predominantly under conditions of enslavement and segregation, those classified as Black do not necessarily share the positive associations of democracy held by those whose freedoms it has protected—such as the majority of those classified as White (Hood, 2000). He tied this revelation to House and Howe's prescriptions for deliberative democratic evaluation, but also to the evaluation field's failure to use evaluators from groups who are racially otherized (House & Howe, 2000). While this difference in experience and subsequent interest is rooted in racial classification, Hood characterizes his perspective as "'tinted' by [his] racial and cultural heritage" (Hood, 2000, p. 79).

Unlike Hood, Evergreen and Cullen suggest that shared racial classification—or more specifically, "color" (2010, p. 133)—does not necessarily correspond with shared interests, particularly when considering other dimensions of difference and the wider range of socio-economic variation among those classified as Black relative to those classified as White (Conley, 2000). As they challenge essentialized notions of racialized difference and allude to intersectionality by mentioning dimensions of difference beyond racial classification, Evergreen and Cullen place primacy on socio-economic class status, financial access, and resources. In the process, they negate the ongoing importance of phenotypic variation and subsequent racial classification from the perspective of disparities and disproportionalities in nearly every dimension of life, as detailed in this chapter's segment on stratification of U.S. society (Conley, 2000; Darity, 2002, 2005; Darity et al., 2015; Darity & Nicholson, 2005; Hamilton & Darity, 2010; Nam et al., 2015).

Evergreen and Cullen also fail to acknowledge that ongoing segregation and the wider level of socio-economic variation among those who are racially otherized manifests within families and communities, such that even members of racially otherized groups who are considered middle- or upper-class tend to have familial and community-level ties with those who enjoy considerably fewer financial resources and less access, whether in the USA or in their countries of origin (Chiteji & Hamilton, 2005). This failure represents an ongoing tension, not limited to the field of evaluation, in how to conceptualize and communicate the socially-produced and simultaneously materially-experienced nature of racialized difference that is directly related to persistent racial stratification, structural racialization, and implicit bias (powell, Heller, & Bundalli, 2011).

While Hood (2000) and Evergreen and Cullen (2010) consider shared social group interests in relation to evaluation generally, other scholars have accounted for such interests as they manifest in the structured context of evaluation practice (Clayson et al., 2002; Freeman et al., 2010; Sturges, 2015). Describing evaluation practice as a "politically contextualized act" (p. 34), Clayson, Castañeda, Sanchez, and Brindis (2002) identified four distinct stakeholder groups—funders, community-based organizations, community members, and evaluators. Members from these stakeholder groups may share various dimensions of difference and social group identification. However, in the authors' experience evaluating programs in Latinx communities, each group had different interests, and the authors describe their interactions as shaped by unequal power. Their complication of the strong and yet not absolute correlation between racial classification and socio-economic class status urges more focused attention on the mediation of socio-economic relations and decision-making processes through structural mechanisms, which may enable or constrain the expression of social group interests by racially otherized groups.

Clayson, Castañeda, Sanchez, and Brindis characterize their own roles as "interpreters, translators, mediators, and storytellers" (2002, p. 35), identifying only two members of their team in terms of national origin and language proficiency and doing so

directly in relation to discussion of their storytelling role. In contrast, both Freeman, Preissle, and Havick (2010) and Sturges (2015) use the word "complicit" to characterize their role in the reproduction of racialized discourse and dynamics as it played out in evaluations they personally conducted. Although the latter both frame their experiences in terms of ethics and recommend establishing at the outset greater clarity about evaluator positions and transparency about how those might affect practice, Freeman, Preissle, and Havick self-identify as White (2010, p. 52) and share their feelings of shame for their moral failing (p. 51), whereas Sturges does not identify himself racially and shares the concrete tools he now uses to hold himself ethically accountable. The differences in these three accounts of the evaluator's role in relation to structural asymmetries that correspond with larger patterns of racial and ethnic stratification point again to questions of agency. Under which conditions are evaluators' alliance with particular social groups and not others chosen by the evaluators, and under which conditions is it attributed to them by others? Whose credibility is damaged by such alliances and whose is strengthened?

Still other scholars have considered social group interests as they manifest in the structured context of knowledge production, the academic canon, and professional advancement. Hood and Madison each question which scholars and which scholarship, respectively, the evaluation transdiscipline has claimed and advanced, shining light on the canon's omission of knowledge produced and agency exercised by members of racially otherized groups (Hood, 2001; Madison, 2007). Hood contemplates the field's disproportionate whiteness, typically considered a natural vestige from a distant past, in ways that correspond with the persistent disparities described in the section on stratification of U.S. society earlier in this chapter:

> I find it curious that African American evaluators went unnoticed when a number of them had received their degrees from prestigious public and private universities with reputations not only for their educational research but also for the emerging field of educational evaluation…. (Hood, 2001, p. 35)

> Clearly, early African American educational evaluators did the studies and published in the so-called right journals. But the bulk of their work appeared in the distinguished *Journal of Negro Education*, which was and remains largely

> unread by the white [*sic*] scholarly community (or at least is not cited in their publications)…. (Hood, 2001, pp. 40-41)

> Some of these African American evaluators worked in close proximity to major figures such as Tyler, Cronbach, Bloom, and others who shaped evaluative thinking on education—but they never made it into the academic inner sanctum… (Hood, 2001, p. 41)

Like Hood's search for evaluators classified as Black in the evaluation canon, Madison searches for the discipline's coverage of underrepresented groups:

> … low-income, nonwhite ethnic and racial groups who have little or no input in policy, program, and evaluation decision making…share a history of political powerlessness, social oppression, and economic exploitation…[and] depend on governmental interventions and interventions sponsored by nonprofit philanthropic organizations to correct deficiencies in the allocation of societal resources. (Madison, 2007, p. 107)

From her review of 20 years of *New Directions for Evaluation* editions, she similarly concludes that this social group, also—distinct from, but likely of interest to the responsive evaluation scholars discussed by Hood—continues "to be presented as *subjects of evaluation* rather than as *invested stakeholders*" (Madison, 2007, p. 113, emphasis in original).

Self-determination. Published more than a decade before the above review, Madison's "Primary Inclusion of Culturally Diverse Minority Program Participants in the Evaluation Process" explains that primary inclusion of program participants in evaluation is necessary regardless of the recruitment and advancement of evaluators from groups that are racially otherized because many "middle-class minorities…are too removed from the worldviews and experiences of the program participants…." (Madison, 1992b, pp. 35-36). While they may share the same racial classification as participants, they do not necessarily share the same culture.

Madison defines primary inclusion as "direct participation of program participants in all phases of program development, from the construction of problems to the evaluation and the interpretation of findings" (Madison, 1992b, p. 35). She spends considerable time explaining not just how the conceptualization of social problems is

culturally influenced, but also how the influence of culture plays out within larger social, economic, and political systems. Her explanation of cultural dominance, or what Young (2011) would call cultural imperialism, is rare in the U.S. evaluation literature for connecting the experiences of racially otherized groups in the USA to that of those outside it under what Smith (2016) would call the larger umbrella of white supremacy. For example, she states,

> The dynamics of dominant colonial culture-third-world culture interactions can be used to describe patterns of cultural dominance and subordination in this country…. [M]inority low-income ethnic communities are very much like third-world countries in that they are politically and economically dependent on the dominant groups for survival, yet they maintain their own core social values that guide their understanding of their living environment and their responses to the environment. (Madison, 1992b, p. 36)

Madison—like Davis—explains how policy makers, program planners, and evaluators, including from otherized groups, who come from an otherizing gaze (Hall, 1997) often interpret social problems as individual deficits. She asks why evaluation could not serve as a tool for program participants to advocate for their own best interests. While infrequent, Madison's conceptualization of the historical and potential role of program evaluation in the context of liberation movements is not unique in the evaluation literature. It is represented more frequently, but not exclusively, by evaluators utilizing decolonizing and indigenous frameworks (Cavino, 2013; Johnston-Goodstar, 2012; Kerr, 2006; LaFrance, 2004).

Evaluators working with decolonizing and indigenous frameworks recognize evaluation as inherently political—still part of the ongoing project of colonization rather than simply having historical roots in colonization (Cavino, 2013; Johnston-Goodstar, 2012; Kerr, 2006; LaFrance, 2004). As such, colonization—and decolonization—are central rather than tangential to evaluation. Kerr (2006) describes her experience of AEA's annual conference in ways similar to Hood's and Madison's experience of the scholarly canon and publication history of evaluation, in that racially otherized groups were largely invisible except possibly as subjects featured in presentations by those from

Europe and settler colonial states (Hood, 2001; Madison, 2007). Also like Madison (1992b, 2007), she draws parallels with international work rooted in colonial relations.

While sovereignty is specific to colonized indigenous peoples in ways that are not trivial, this centering of the collective experiences of marginalization and exploitation is characteristic of critical theories of systemic oppression more generally, as are questions about the gaze. Self-representation constitutes a decolonizing act (Cavino, 2013; Johnston-Goodstar, 2012; Kerr, 2006; LaFrance, 2004). Again, in ways similar to Madison's (1992b) notion of primary inclusion, Johnston-Goodstar (2012) proffers evaluation advisory groups as sites of both knowledge production and agenda setting in ways that differ fundamentally from notions of cultural competence, "where the focus is on difference, competency, and issues of access" (Cavino, 2013, p. 339). While many scholars drawing from decolonizing and indigenous methodologies refer to cultural competence, none advocate for it, and some critique it (Cavino, 2013). Even LaFrance (2004)—in a piece entitled "Culturally Competent Evaluation in Indian Country"—does not necessarily advance the notion of cultural competence. Mentioning it briefly only three times, she states that

> … the goal of a competent evaluator, especially in Indian Country, should be to actively seek cultural grounding through the ongoing processes of appreciating the role of tribal sovereignty, seeking knowledge of the particular community, building relationships, and reflecting on methodological practices. (p. 39)

Perhaps more importantly, LaFrance sees the real value of culturally competent evaluators as lying in their increased ability to change structures as opposed to their ability to attend to issues of culture within evaluation practice:

> Consequently, evaluators who learn how to practice in a culturally competent framework have the potential for changing not only the field of evaluation but also conversations on knowledge creation, its components, and its ramifications. (2004, p. 42)

In Maack and Upton's (2006) study of how and why independent evaluation consultants form collaborative relationships with each other, one presumably indigenous

respondent lists their reason as one of preserving and protecting tribal sovereignty and self-determination:

> Evaluation, like land or gaming rights, is a federally protected aspect that Indians have a legal right to…. "[O]utsiders" often come in during research and evaluation projects and do not conduct themselves in an ethically, culturally, or scientifically appropriate way. My role is a gatekeeper, translator (across/between tribes and funders), and capacity builder for Natives and non-Natives where evaluation is concerned. (p. 65)

Similarly, Kerr (2006) explicitly questions rhetoric about "partnership" with non-indigenous individuals and organizations. Her critique of the notion of partnership contrasts sharply with accounts like that of Letiecq and Bailey (2004), who describe their experience using Tribal Participatory Research on an American Indian reservation. Having self-identified as White, they consistently refer to themselves as from the "majority culture" without ever historicizing that status. They similarly use the phrases "cross-cultural" and "partnership" to refer to their work with a tribal nation, despite its fundamental asymmetry. While conscious of their own racial classification as White, explicit about their "outsider" status, and aware of the effects of colonization (although colonization itself is never mentioned in the article and sovereignty is only mentioned once), Letiecq and Bailey do little to disrupt, and, in fact, they reinforce prevailing understandings of difference through their detailed examination of whether a "racial/ethnic match" is necessary considering the lack of availability of trained indigenous evaluators in rural areas as well as their laundry list of the challenges of working in Indian country, including resistance to evaluation and differing conceptualizations of time. In the end, they conclude that

> … research partnerships between native and nonnative researchers can be advantageous, as long as the partnership seeks equal voice and input by all of its members, and the diverse expertise and knowledge shared by partners is given equal weight in the design and implementation of the research. (p. 349)

Such accounts merit a deeper understanding of whiteness, which follows.

Whiteness. The notion of whiteness appears once in the U.S. peer-reviewed evaluation literature, in the context of cultural competence. Citing Pon (2009), Evergreen

and Robertson argue that "the whole idea of cultural competency…is a problematic concept because it inherently keeps Whiteness [*sic*] at the center, with some 'otherized' culture as the topic matter on which to become competent" (2010). As part of the same series of articles examining cultural competence, Evergreen and Cullen advance the notion of cultural humility (Tervalon & Murray-Garcia, 1998) instead, stating, "We have encountered situations where our findings or presence on the evaluation team have been called in to [*sic*] question on the basis of our sociocultural backgrounds, despite more than adequate cultural understanding" (2010, p. 131). While neither set of authors (Evergreen and Cullen or Evergreen and Robertson) defines whiteness, their critiques—which cite an analysis of cultural competence rooted in critical race theory (Pon, 2009) while simultaneously claiming authority—provide insight into its meaning. Its meaning is expanded upon below.

As important as the association of whiteness with authority is that "whiteness itself has become a form of economic capital supporting the reproduction of dominance and intergroup inequality" (Vijaya et al., 2015, p. 7). Whiteness is a material possession—property—one that can be deployed as a resource and enjoyed whenever a person classified as White deliberately or unthinkingly takes advantage of the privileges and protections afforded to them simply because of their racial classification (Harris, 1995; Vijaya et al., 2015). Because the hegemony of racialized neoliberalism makes it difficult for those classified as White to see themselves as part of a social group rather than as individuals, to see structures of stratification, and to see their current position as anything other than a result of their personal accomplishments and merit, it may be difficult for those classified as White not to take advantage of or exercise such privileges and protections (Oliver & Shapiro, 2001; D. J. Roberts & Mahtani, 2010; Spilerman, 2000; Stec, 2007).

Like any form of property, the value of whiteness rises with its exclusivity, and thus its possession must be actively secured (Harris, 1995; Nylund, 2006; Rabaka, 2007; Todd, 2011). And like other identities, it is not essential, but constituted and reconstituted through social group membership and non-membership (Bonilla-Silva, 1997; Doane,

2003; Doane & Bonilla-Silva, 2003; Young, 2011). This means that whiteness is performed through everyday decisions and interactions, including the assertion of professional credentials and competitive positioning with respect to discourses valued by the field and surrounding industries (Nylund, 2006; Rabaka, 2007; Vijaya et al., 2015).[43] Discourse becomes a form of currency in which members of the professional class in evaluation and the surrounding nonprofit industry—whose disproportionate classification as White was established in the previous segment—trade. Diversity's value as currency is illustrated in the following finding from Maack and Upton's (2006) study regarding collaborative relationships in evaluation consulting that was referenced earlier:

> [H]alf of the respondents had brought in others to projects as part of expanding their diversity (defined as skills diversity), and almost as many had been asked to participate in a project partly because of their race or ethnicity. Both [the study authors] have similarly had client requests for "representative" consultants or deliberately sought out ethnically diverse evaluators to strengthen teams, especially in projects that include a focus on minority [*sic*] populations. However, as respondents emphasized, one's research and writing skills are essential. Although representative evaluators may be sought initially on the basis of race/ethnicity, they must first and foremost have the experience and abilities needed to fulfill project requirements. One very experienced consultant wrote: "I am black [*sic*], so it is not unusual for people to seek me out to 'represent' or to enhance project diversity. This has happened many times, but at this stage in my career, people invite me for my reputation first." (Maack & Upton, 2006, p. 65)

Noteworthy in the above finding is that the asker—the one with agency—is unnamed. The racialized evaluator is simply "asked"—and that, too, only if they can demonstrate the necessary skills. Those who actively "brought in others as part of expanding their diversity" are unidentified with respect to racial classification, reinforcing the normativity of whiteness. Those who "had been asked to participate in a project" are racialized in that they were brought in "partly because of their race or ethnicity" as "ethnically diverse evaluators…especially in projects that include a focus on minority populations." Thus, their value lies in their representativeness rather than their role with respect to knowledge production; even the notion of validity is not raised here

[43] See "How to Hoard Opportunities" (Tilly, 1998).

(Kirkhart, 1995; Mertens, 2007). They are not necessarily sought out in "regular old" evaluations in which racial classification is presumably not salient—in which "regular old" evaluators, presumably classified as White, would suffice—but in those that focus on "minority" populations and often upon the request of clients (Weisinger et al., 2015). As such, while both whiteness and blackness—or more broadly, "diversity"—are property, those classified as White are the only ones who can own or determine the value of either (Crusto, 2005; Harris, 1995; A. Smith, 2016; Vijaya et al., 2015). While members of racially otherized groups necessarily participate to varying degrees in the commodification of their "otherness"—intentionally and unintentionally—their doing so does not, by itself, change the ownership structure.

Use of the word "partly" to qualify why members of racially otherized groups are asked to participate is important—because "one's research and writing skills are essential." While the research and writing skills of those who brought in others go unquestioned, the skills of those marked as having been "sought initially on the basis of race/ethnicity" must be legitimized. Just as it is difficult for those classified as White not to take advantage of whiteness, it is similarly difficult for members of racially otherized groups not to participate in the contestation for incremental changes in their position relative to whiteness—such as increased entitlement, authority, and legitimacy—that result from their association with whiteness. The "very experienced" respondent identified as Black reveals this tendency by stating that "at this stage in my career, people invite me for my reputation first," as if their blackness could somehow be separated from their reputation in a field that is so disproportionately composed of those classified as White.

It is significant that, again, "people" is unqualified with respect to both racial classification and worthiness. "People" is the unmarked and "unexamined center" (Doane, 2003, p. 7). "People" are the producers of knowledge and agents of change in the nonprofit industry, brokering deals with "people" in government, for-profit businesses, and research/evaluation. Members of racially otherized groups "are asked" by "people" to provide diversity when "people" deem it advantageous (Crusto, 2005).

Summary of Racialized Difference in Evaluation

This segment reviewed literature addressing racialized difference in the U.S. field of evaluation from either a contextual/multi-level or critical perspective. It included literature that expands boundaries beyond individuals as the unit of analysis and that questions prevailing, liberal constructions of racialized difference. The latter consists of literature that generally acknowledges systemic oppression based on racial classification; addresses the production of racial categories, including the notions of intersectionality and essentialism; and demonstrates some understanding of social group interests. Both bodies expose an ambivalence within the field regarding racialized difference as well as a nuanced tension with regard to evaluation's role in advancing social group interests—between an approach consciously rooted in self-determination and one that less consciously normatizes whiteness.

Review of the evaluation literature demonstrates the insufficiency of frameworks that expand the unit of analysis and consider mechanisms that shape relations among levels within the industries surrounding evaluation—exemplifying concepts from systems thinking—without challenging prevailing notions of difference. It also reveals a small, but long-standing body of evaluation literature that challenges prevailing notions of difference—exemplifying concepts from critical theories of systemic oppression—without directly considering the mechanisms that shape relations among levels within the industries surrounding evaluation in an effort to advance justice conceptualized as self-determination. Attempting to bridge this gap, the next segment reviews social-systems theories within systems thinking and critical theories of systemic oppression as the analytical and theoretical frameworks that, when combined, may address the field's ambivalence and tension regarding racialized difference.

Systems Thinking

As discussed in the segment above on context-sensitive and multi-level evaluation, Lee and Gilbert (2014) encourage a move beyond the individual to "consider the interconnected parts of the ecosystem" (p. 106) that contribute to the disproportionalities observed in the field. Their repeated references to "the ecosystem"

and reference to "policies and practices that have been in place for generations" (p. 107) suggest both a systems orientation and a framework drawn from critical theories of systemic oppression. The final two sections of this chapter address each of these bodies: systems thinking as the analytical framework, and critical theories of systemic oppression as the theoretical framework underlying this research. Chapter Two closes with examples applying parts of the analytical framework to inequality, specifically.

As the systems field is vast, with more than 800 theories, this section is limited to systems concepts that cut across the field. Furthermore, it focuses on areas of overlap with critical theories of systemic oppression: (1) boundaries of space and time; (2) relations, exchanges, or flows of stock/capital through structurally mediated mechanisms, including feedback loops and their potentially distal or cumulative effects; and (3) different perspectives or interests (Flood, 2010; powell, 2010; powell, Heller, & Bundalli, 2011; Stave & Hopper, 2007; Thomas & Parsons, 2017; B. Williams & Hummelbrunner, 2011). The literature that connects structural racialization and systems thinking, and that connects both to evaluation, is largely not peer-reviewed, but from john powell's work with racial justice and disparities-reduction practitioners through the University of California at Berkeley's Haas Institute for a Fair and Inclusive Society and previously through the Kirwan Institute at Ohio State University.

Analytical Concepts

According to open systems theory—a response to the prevailing reductionism of components—the survival of any system depends on how well it interacts with its environment (Kreitner, 2008). An organism as a whole co-exists

> … in relation to an environment. Its functions and structure diversify or are maintained by management of a continuous flow of energy and information between organism and environment. Flows occur in an organism through its many interrelated parts. Parts are interrelated through feedback loops. (Flood, 2010, p. 271)

The next section describes the nonprofit industry surrounding evaluation as a system in social-ecological terms.

Social-ecological Systems Theory. Social-ecological systems theory derives from open systems theory, psychoanalytic thinking, and an action orientation (Flood, 2010). The nonprofit industry can be considered a system that co-exists in relation to its environment. Figure 3 illustrates a slice of the social ecology within and surrounding the nonprofit industry (depicted in Figure 2), which includes evaluation and which is also included, among other systems, within larger societal structures. While evaluation's role in the industry—particularly with respect to the generation and use of valid results—has been acknowledged in the evaluation literature (e.g., Kirkhart, 1995, 2010; Mertens, 2007, 2013), the evaluation literature has yet to examine its role in relation to the process of systemic racialization. Nor has the nonprofit industry—a system—as a whole been examined in relation to the process of structural racialization and reproduction of racialized difference. The nonprofit industry and evaluation's role within it are discussed below through the concepts of boundaries, relations, and perspectives that are common to all systems thinking (B. Williams & Hummelbrunner, 2011) and salient in system dynamics and complex adaptive systems.

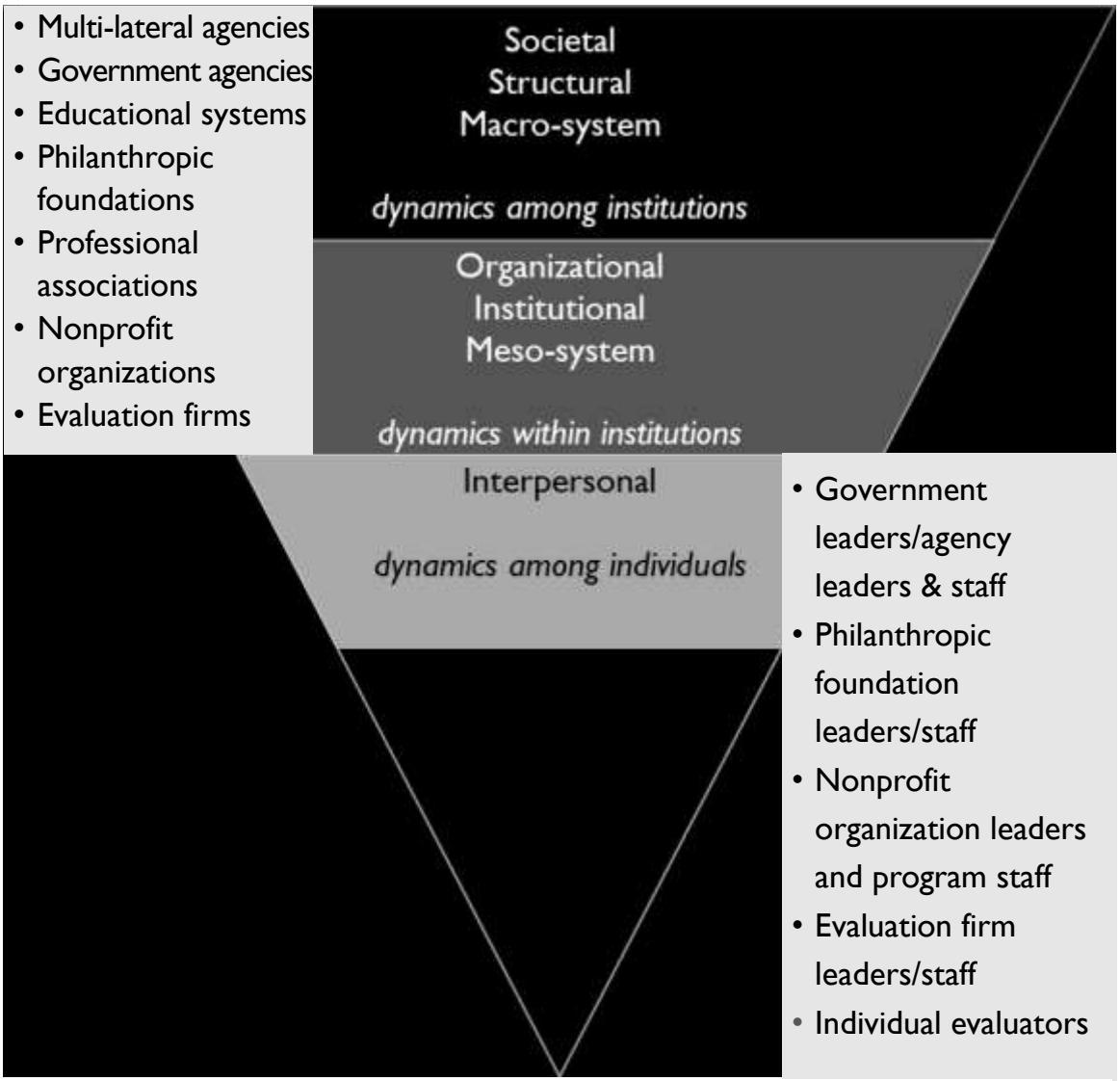

Figure 3. Social- ecological model of nonprofit industry surrounding evaluation

Boundaries. It is physically and intellectually impossible for decision makers to consider the full contexts of their situations. Making any decision requires defining a reference system of concern by establishing a boundary somewhere (Stephens, 2012; Valentinov, 2012; B. Williams & Hummelbrunner, 2011). The determination of what lies inside a system necessarily involves marginalizing what remains outside, whether implicitly or explicitly (B. Williams & Hummelbrunner, 2011). In other words, the choice of boundaries determines what is relevant and what is not and is thus an ethical

decision. Questions arise regarding who is involved in determining what gets in and what stays out (Valentinov, 2012; B. Williams & Hummelbrunner, 2011).

Boundaries apply to both space and time. Systems thinking involves a spatial expansion of the boundaries around the current unit of analysis—beyond the knowledge, behaviors, and conditions of individual program participants and individual program evaluators as well as beyond the interpersonal relationships between them—to consider the structural arrangements among and within the institutions that shape those internalized and interpersonal dynamics (Reskin, 2003). It also involves expanding the view of present-day incongruences as distal effects—the intergenerational legacy of racialized structural arrangements encoded into this country's historical foundation that continue to influence the organization of labor, participation in decision making, and access to various forms of capital.

Narrowly framing boundaries around individuals involved in evaluation practice presents risks for developing remedies in that while attitudes and behaviors may manifest at individual levels of interaction, they reflect collective processes of definition and redefinition arising from hierarchical differences in power between social groups rather than individual tastes (Davis, 2014). Reframing the current boundaries of space and time from around the observed incongruence between program participants and program evaluators in the realm of evaluation practice to instead provide a view of the whole system can allow for greater understanding of how the parts of the system, including the historical foundations of racialized inequality, relate to the whole (Stave & Hopper, 2007). It may make visible that "racism [is] essentially a relationship, and not…an attitude or idea" (C. Williams, 2008).

Critical Systems Thinking. Critical systems thinking (CST) emphasizes the boundary-setting process, which is rooted in decision makers' perspectives. From a practical and ethical perspective, it encourages reflecting upon one's interests and being intentional about boundary setting to avoid exacerbating the problem to be addressed (B. Williams & Hummelbrunner, 2011). It also encourages boundary critique, in which decision makers question and negotiate each other's boundary judgments (Valentinov,

2012). CST benefited from a sophisticated understanding within systems thinking of knowledge, ethics, and power as well as participatory practices (Stephens, Jacobson, & King, 2010) and ties existing structures of wealth, status, power, and authority to boundary setting. Inasmuch as it relies on Soft Systems Thinking—its antecedent, which advances the notion that social reality is the construction of people's interpretation of their experiences and which thus actively engages people's points of view and intentions to yield changes in social systems—CST treats economic and political structures as real in the sense that they perpetuate social arrangements. Transforming these structures is thus considered as necessary for transforming people's worldviews as is the reverse (Flood, 2010; Shalhoub & Al Qasimi, 2005).

Beyond a commitment to systems thinking in general, CST adheres to five core commitments: (1) critical awareness, (2) social awareness, (3) human emancipation, (4) theoretical complementarity, and (5) methodological complementarity (Flood, 2010; Laszlo & Krippner, 1998). Critical awareness involves questioning the assumptions, values, and theoretical underpinnings of systems and systems approaches. Social awareness involves considering societal or organizational expectations and acceptance of any particular systems approach at any particular time. Human emancipation, like an enabling conceptualization of justice, involves attending to people's collective development and exercise of their potential. Theoretical and methodological complementarity are considered necessary to support such work (Flood, 2010; Laszlo & Krippner, 1998).

Relations. Reframing the boundary around individuals involved in evaluation practice can make visible the nature of interpersonal and institutional relations among evaluators, funders, nonprofit leaders, policymakers, program participants, and others—some of whom share social groups and associated interests or perspectives and others of whom do not—who were outside the initial boundary (House, 1990, 1995; powell, 2010). All relations are mediated through structural mechanisms—which may be characterized by symmetry or asymmetry, yielding dynamics within and among individuals and institutions that are enabling or oppressive—underlying the nonprofit industry's

organization of labor and decision-making processes (Reskin, 2003; Young, 2011). As such, what may work at the micro level to improve internalized and interpersonal dynamics within and among individual program participants and program evaluators may not work at larger levels (powell, 2010).

As mentioned in the discussion of boundaries, time is as integral to the understanding of relations as space is. Systems-oriented thinking requires a shift from linear notions of cause and effect—wherein analysts identify and isolate proximate causes for any given effect—toward an approach that recognizes that each effect has multiple causes and each cause has multiple effects, many of which may be distal (powell, 2010). The incongruence between program participants and program evaluators observed today is a product of multiple, reinscribed interactions within the system that may have started long ago (powell, 2010). The notion of emergence describes how multiple, mutual, and cumulative causality often lead a whole to be greater than the sum of its parts. What seems to ameliorate a problem in the short term may thus not solve it in the long term (Flood, 2010; Laszlo & Krippner, 1998; Menendian & Watt, 2008; powell, 2010; Thomas & Parsons, 2017; B. Williams & Hummelbrunner, 2011).

System Dynamics. Systems thinking posits that at any specific moment, systems consist of elements that have a stock value, which represents the accumulation of past events. These values change over time through inflows and outflows. System Dynamics is an approach to understand this dynamic behavior in social systems particularly (B. Williams & Hummelbrunner, 2011). The approach involves mapping the relations and modeling the feedback processes among parts of a system to focus attention and information gathering on how these change over time (Behrens & Kelly, 2008). Systems thinking can thus be considered a reduction to dynamics, in contrast with the more common reduction to components (Laszlo & Krippner, 1998).

Mapping and modeling system dynamics require capturing the structure and processes of a system from those participating in the system. Delayed effects of feedback influence both the functioning of the system and participants' experience of the system and are therefore included in the maps and models (Behrens & Kelly, 2008; Flood, 2010;

B. Williams & Hummelbrunner, 2011). Once a conceptual map of the relations within a system has been developed, the map can be modeled quantitatively or qualitatively to allow examination of interrelations among variables and simulate the effects of changes in one or more of them (Fredericks, Deegan, & Carman, 2008; Nguyen et al., 2011).

Attention to system dynamics can reveal change over time—especially important in relation to race, racism, and racial justice, all of which are "slippery" rather than static concepts (Byng, 2013; Davis, 2014; powell, 2010; Thomas & Parsons, 2017). Delayed feedback and the system's adaptation to dramatically positive changes in the short term may ultimately result in no change at all, or they may result in a backlash. For example, the end to formal segregation set in motion dynamics that have since undermined the long-term success and stability of the policy level changes associated with Civil Rights wins (powell, 2010).

Evaluation as feedback. Reframing boundaries of space and time around the current incongruence observed between program participants and program evaluators reveals how what may seem to help in the short term may actually harm in the long term (powell, 2010). For example, it is possible that individually and interpersonally focused responses at the level of evaluation practice provide a buffer between nonprofit leadership and nonprofit clients, whose social groups and associated perspectives and interests typically diverge (National Urban Fellows, 2012). This buffer may allow nonprofit organizations that are led by those classified as White and that provide services to client populations disproportionately consisting of racially otherized groups to continue accumulating legitimacy and receiving funding at the expense of constituent-led organizations working toward redistributive policy changes at the structural level (Weiss, 1993). The latter tend not to receive philanthropic funding and have less access to evaluation services and associated evidence or other sources of legitimacy (Cohen, 2014; Lawrence, 2009). The above possibility of displacing leaders who are racially otherized is significant from the perspective of an enabling conceptualization of justice and self-determination, particularly as critical race theorists believe that only liberation

movements—as opposed to economic assimilation—can generate the structural changes that could ultimately lead to racial justice (Byng, 2013).

Moreover, critical race theorists believe that the production of critical knowledge about structural racialization from the perspective of racially otherized groups is pivotal for such movements (Byng, 2013; Parker, 2004). Evaluation represents a form of knowledge production. Bonilla-Silva (2010) and Doane (2006) are credited with incorporating discursive practices into theorizations of race and racism, although several scholars have noted the hidden, unrecognizable nature of present-day racism, including through invocations of culture (Byng, 2013; Pon, 2009; C. Williams, 2008; Wolfe, 2010).

> "[D]ifference" can be fetishized, investigated, and tolerated. At the micro level familiar-difference can make it seem as though our society is racially flexible and tolerant in spite of macro level policies and meso level discourses of intolerance. (Byng, 2013, p. 710)

Fitzpatrick's (2012) discussion of context and culture, discussed in the previous section, illustrates the corollary importance of what is not said—"where race is foregrounded and where it is subtext" (Byng, 2013, p. 710)—from the perspective of understanding evaluation's discursive role in systemic racialization.

Perspectives. Individuals' perspectives are related to their interests, motivations, and intentions, which previous sections have established are informed by their identification with the structural location and experience of one or more social groups (Thomas & Parsons, 2017; Young, 2011). Along with other system dynamics, perspectives impinge on the points of leverage available to encourage movement in any direction (Thomas & Parsons, 2017). Mapping a system's dynamics allows for the identification of points that may provide leverage to shift the system's dynamics (Stave & Hopper, 2007).

With regard to enabling conceptualizations of justice, what is important is expanding the boundary beyond individual perspectives, which cannot be empirically observed and measured, to consider the structural mechanisms that produce particular interpersonal dynamics and individual-level effects—both of which can easily be empirically observed and measured. It is structural mechanisms that can either allow or

blunt the effects of individual motivations and intentions to discriminate against members of particular groups. It is also structural mechanisms that can either allow or blunt the effects of individual motivations and intentions to advance justice (Reskin, 2003).

Summary of Systems Thinking

While the effects of feedback loops are not contingent upon the intention or motivation of their source, but rather on the mechanism involved in their transmission, there is a place for volition in relation to the reproduction of structural racialization (Reskin, 2003; C. Williams, 2008): "There is a contradiction in a professed belief in equality and justice, but a societal willingness to tolerate and accept racial inequality and inequity" (Parker, 2004, p. 86). Rather than continuing to answer questions about the origins and effects of racism, critical knowledge produced from the perspectives of racially otherized groups might answer questions like: What processes produce and reproduce disadvantage? What processes continuously inscribe race into American society? How is white supremacy (re)produced, and what are the (pre)conditions of its (re)production? (Byng, 2013; C. Williams, 2008).

As detailed in Chapter One, individuals who share particular experiences of super-ordination and sub-ordination develop shared perspectives and interests rooted in those experiences. They identify with other individuals who share the same perspectives and interests, forming social groups that—in the process—come to develop associated cultures (Young, 2011). Importantly, this process of identifying with social group interests applies both to those who accumulate certain kinds of capital under particular structural arrangements as well as those who do not—who in fact expend capital under them (Bonilla Silva, 1997). Using critical theories of systemic oppression, the next section explicates the oppressive dynamics of super-ordination and sub-ordination. It also illuminates how this varies among indigenous peoples experiencing settler colonization, survivors of the Middle Passage from Africa, and peoples representing areas violated by European and European-settler imperialism and militarism. Each has different perspectives and interests that are rooted in their different experiences suffering under, and inevitably participating in, systemic oppression—specifically white supremacy.

Critical Theories of Systemic Oppression

Critical theories of systemic oppression include variations of critical race theory, feminist theory, postcolonial theory, indigenous studies, queer theory, and crip theory, the first three of which inform this dissertation directly. After a brief orientation to critical race theory, this section of Chapter Two reviews two pieces of inherently intersectional literature that together form the theoretical framework for this dissertation: (1) heteropatriarchy and the three pillars of white supremacy, and (2) the five faces of oppression.

Critical race theory (CRT) grew out of the critical legal tradition in the 1980s, in response to that tradition's relegation of legal issues relevant to racially otherized groups. It rests on the foundation that racism is pervasive and hegemonic—a normal fact of daily life in society—rather than individual acts of discrimination (Parker, 2004). It similarly holds that the law and liberalism more generally will not create an equitable, just society (Parker, 2004). CRT challenges the essentialism of race, arguing that race is socially produced and that there is greater variation within racial (and other) groups than between them; it further challenges the experience of those classified as White as the normative standard. Instead, CRT grounds itself in the experiences of racially otherized groups, often through the use of literary, narrative knowledge and storytelling. In that sense, critical race theorists acknowledge cultural imperialism, hegemony, and dominance, but believe that a focus on cultural differences obscures racism (Wolfe, 2010).

Two remaining tenets of CRT are interest convergence and differential racialization. Interest convergence refers to the pattern in which decision makers address the demands made by racially otherized groups only when they happen to converge with the interests of those classified as White, who are otherwise loath to relinquish the material and psychic advantage, including a state of relative comfort, that they receive from white supremacy. Differential racialization refers to the way that multiple social groups are racialized in different ways or that a single social group is racialized in different ways at different times, depending on the historic, social, political, or economic needs of those classified as White (Abrams & Moio, 2009; Delgado & Stefancic, 2001).

Theoretical Concepts

This section elaborates on the CRT concepts of interest convergence and differential racialization briefly defined above. It also addresses the corollary concept of whiteness.

Interest convergence. CRT's notion of interest convergence, also known as materialist determinism, hinges on the dialectical relationship between those classified as Black and those classified as White, wherein White advantage is inextricably tied to—indeed, interdependent with—Black disadvantage. While decision makers may make concessions in response to the demands of racially otherized groups, to the extent that they do not hinder interests of those classified as White, they will avoid altering the structural arrangements that underlie relations of super-ordination and sub-ordination between members of groups who are racially normatized as White and members of groups who are racially otherized. Bell famously describes the *Brown v. Board of Education* decision as an example of interest convergence between those classified as Black and those classified as middle- or upper-class White. Desegregation was viewed as important for the USA to appear credible to, and win the support of, non-aligned countries in its competition with the U.S.S.R.—more largely between capitalism and communism. Additionally, segregation was viewed as a barrier to industrializing the South.

> [T]he fourteenth amendment, standing alone, will not authorize a judicial remedy providing effective racial equality for blacks where the remedy sought threatens the superior societal status of middle and upper-class whites…. Racial remedies may instead be the outward manifestations of unspoken and perhaps subconscious judicial conclusions that the remedies, if granted, will secure, advance, or at least not harm societal interests deemed important by middle and upper-class whites (Bell, 1980 as cited in Guinier, 2004, p. 94).

Guinier explicitly "modifies and elaborates on" (2004, p. 92) the interest-convergence dilemma by proffering the notion of interest *divergence*. Unlike liberalism, which treats racism as a psychological (internalized) and interpersonal problem, critical theories of systemic oppression define racism as structural. Premised on the idea that "[r]acism is a structural phenomenon that fabricates interdependent yet paradoxical

relationships between race, class, and geography" (p. 100), the interest-divergence dilemma requires "racial literacy, meaning the capacity to decipher the durable racial grammar that structures racialized hierarchies and frames the narrative of our republic" (p. 100).

Differential racialization. Differential racialization refers to changes in the racialization of any group, in relation to others and over time, to serve the interests of those classified as White. While not unique to groups classified as Asian, the most common example concerns the racialization of that social group. When super-ordinated groups need exploitable labor, groups classified as Asian are racialized as harmless—and sometimes even favorably—to fill that need. The need among groups super-ordinated in U.S. for workers to build transcontinental railroads in the mid-1800s, for math and science professionals to compete with the U.S.S.R. in the mid-1900s, and for software technicians in the late 1900s all resulted in changes in U.S immigration policy that permitted, re-instated, or expanded migration from Asia.

When groups classified as Asian American are perceived as succeeding at rates higher than that of those classified as White, however—threatening the latter's super-ordinated status—the former are racialized as unscrupulous and conniving and subsequently excluded. For example, the perceived success of those of Japanese descent in the USA is partly what led them to be considered a threat and thus incarcerated during World War II in ways that those of Italian and German descent were not. When groups super-ordinated by current structural asymmetries fear that their status is threatened by demands for structural change from racially otherized groups, though, they attribute the success perceived among groups classified as Asian to natural intelligence and cultural values that prioritize education, hard work, and compliance rather than corruption or dishonesty. Asian success is then presented as evidence against racial discrimination and structural barriers in what is called the Model Minority Myth (Wu, 2002). The Model Minority Myth illuminates the implications of framing inequities in terms of culture as opposed to race.

Whiteness. "The fact that [racially otherized groups'] behavior is subject to such intense scrutiny means that the 'true' Americans—the established 'white' Americans—sit in perpetual judgment of groups who they do not embrace as true Americans" (Darity, 2005, p. 148).[44] The above examples of differential racialization—the specific ways that groups over time continue to be racially otherized—or differentiated from each other and from Whites, who continue to be racially normatized as the "true Americans"—illustrate the ongoing social, political, and economic contestation of racial otherness in relation to whiteness (Bonilla-Silva, 1997). While the power to scrutinize and judge—what feminist and postcolonial scholars refer to as "the gaze" and critical race scholars refer to as "surveillance"—is discussed at greater length in relation to the power that lies in discourse (naming, categorizing, differentiating) in Chapter Three (e.g., Delgado & Stefancic, 2001; Hall, 1997; Said, 1978), the concept of whiteness is discussed below.

While closely connected with class and deracialized notions of culture, as suggested by the connotations of both high class and high culture being White (Vijaya, Eshleman, & Halley, 2015), whiteness is typically described (including in Chapter One) as the unmarked, default category—the absence of color.

> The "hidden nature" of whiteness is grounded in the dynamics of dominant group status…. [W]hites in the United States have used their political and cultural hegemony to shape the racial order and racial understandings of American society…. Historically, white-dominated racial understandings have generally focused upon the characteristics (i.e., "differences") of subordinate groups rather than the nature of whiteness. This emphasis by whites upon the racial 'other' has gone hand in hand with the politically constructed role of whiteness as the 'unexamined center' of American society…. Consequently, in a discourse that focuses upon differences and the racialized "other," white becomes a default category…." (Doane, 2003, p. 7).

Ahmed's body of work complicates this perspective regarding the invisibility of whiteness, as well as the understanding of whiteness as the default—which risks essentializing whiteness as a given. Challenging that idea, she shows how whiteness, like

---

[44] While APA style calls for the first letter of racial identities to be capitalized, Darity challenges the association of color with group identity and does not capitalize the first letter of black or white.

all racialized difference, is actively produced through the repetition of routinized actions—a bad habit. Its power lies in the invisibility of its (re)production.

> The institutionalization of whiteness involves work: the institution comes to have a body as an effect of this work…. Institutions involve the accumulation of past decisions about how to allocate resources, as well as "who" to recruit. Recruitment functions as a technology for the reproduction of whiteness" (Ahmed, 2007a, p. 157).

Ahmed also explains how whiteness is not reducible to white skin. The presence of any racially otherized bodies at all in white institutions is perceived as success. These bodies are simultaneously invisible, when they blend in, and hypervisible when they do not.[45] Making whiteness visible, thus, only makes sense from the point of view of those for whom it is invisible. Moreover, identifying or raising whiteness for discussion in such settings "is read as a sign of ingratitude, of failing to be grateful for hospitality we have received by virtue of our arrival" (Ahmed, 2007, p. 164).

The three pillars of white supremacy. One of the limitations of CRT, which continues to be addressed through numerous traditions that the movement spawned, is that while its beginnings are in the experiences of those classified as Black, which are informed by the country's foundation in the violence of enslaved African labor, it did not until recently incorporate experiences originating in the country's foundational violences of genocide and imperialism. Nor did it systematically incorporate an analysis of cis-hetero-patriarchy and capitalism. Indeed, Crenshaw's work on intersectionality is seminal in that regard. Described below, Smith's heteropatriarchy and the three pillars of white supremacy (2016) aim to remedy both these omissions by intersectionally conceptualizing the difficulty that multiple racially otherized groups experience in liberation movement organizing against gender-based violence, in part as a result of interest divergence and differential racialization.

---

[45] Here, Ahmed is referring quite literally to color as opposed to identity, and as such, the first letters of "black" and "white" are not capitalized. In her other writing, Ahmed, who is based in the UK and Australia, discusses her Muslim name, explicitly identifies as a Black feminist, and describes herself as being of Pakistani and English ancestry, illustrating how racial categories are constructed—and constructed differently—in different contexts.

Under Smith's conceptualization, effective organizing across otherized groups cannot take place until each sees how it participates in and benefits from the subjugation of the others, albeit to a lesser degree than those classified as White. The logic underlying each group's historical and ongoing subjugation differs, however—pointing to the difficulty in developing a mass movement. Those classified as Black and as Asian, for example, benefit from the genocide—including the ongoing literal and figurative disappearance and erasure—of indigenous peoples that arises from settler colonization of the Americas. They internalize narratives of indigenous peoples as a savage, nearly-extinct species and their ancestral land as there for taking. Indigenous peoples and those classified as Asian benefit from the commodified labor and servitude of those classified as Black under capitalism. They internalize anti-Black narratives regarding violence and laziness that are necessary to support the over-surveillance and mass incarceration that allow the exploitation and enslavement of those classified as Black to continue. And indigenous peoples and those classified as Black benefit from U.S. military and economic aggression discharged wherever the USA has imperial interests, which ensure the global movement of different kinds of capital—natural or human "resources," the latter being exploitable or surplus labor. They internalize narratives of those classified in the USA as Asian, Middle Eastern/North African or Muslim, and Latinx—regardless of nationality, religious identity, or number of generations in the USA—as perpetually foreign and potentially threatening aliens, whose names, foods, clothes, languages, and belief systems will never be considered "American."

These three narratives, illustrated in Figure 4, constitute the process of racialization. The Model Minority Myth described earlier is one example of how they may change over time, but are continually (re)produced dialectically—in relation to each other and in relation to whiteness—to uphold white supremacy. Smith describes the heteropatriarchal family as the hierarchical, authoritarian model of white supremacy. It enforces the narrative (re)production of these three groups—like a father characterizing his three children—effectively impeding them from organizing across racially otherized groups to advance a shared interest in ending gender-based violence. Tuck and Yang

(2012) further complicate the possibilities for cross group solidarity and collective action by introducing the ethic of incommensurability, wherein the gaps rather than the commonalities in experiences of oppression and interests among groups are sought. Under the ethic of incommensurability, it is the gaps—as opposed to the lowest common denominator—that are the areas of possibility.

The five faces of oppression. Young (2011), whose work regarding the constitution of identity through social group interests was discussed in Chapter One and whose enabling conceptualization of justice informed Chapter Two, articulates five faces

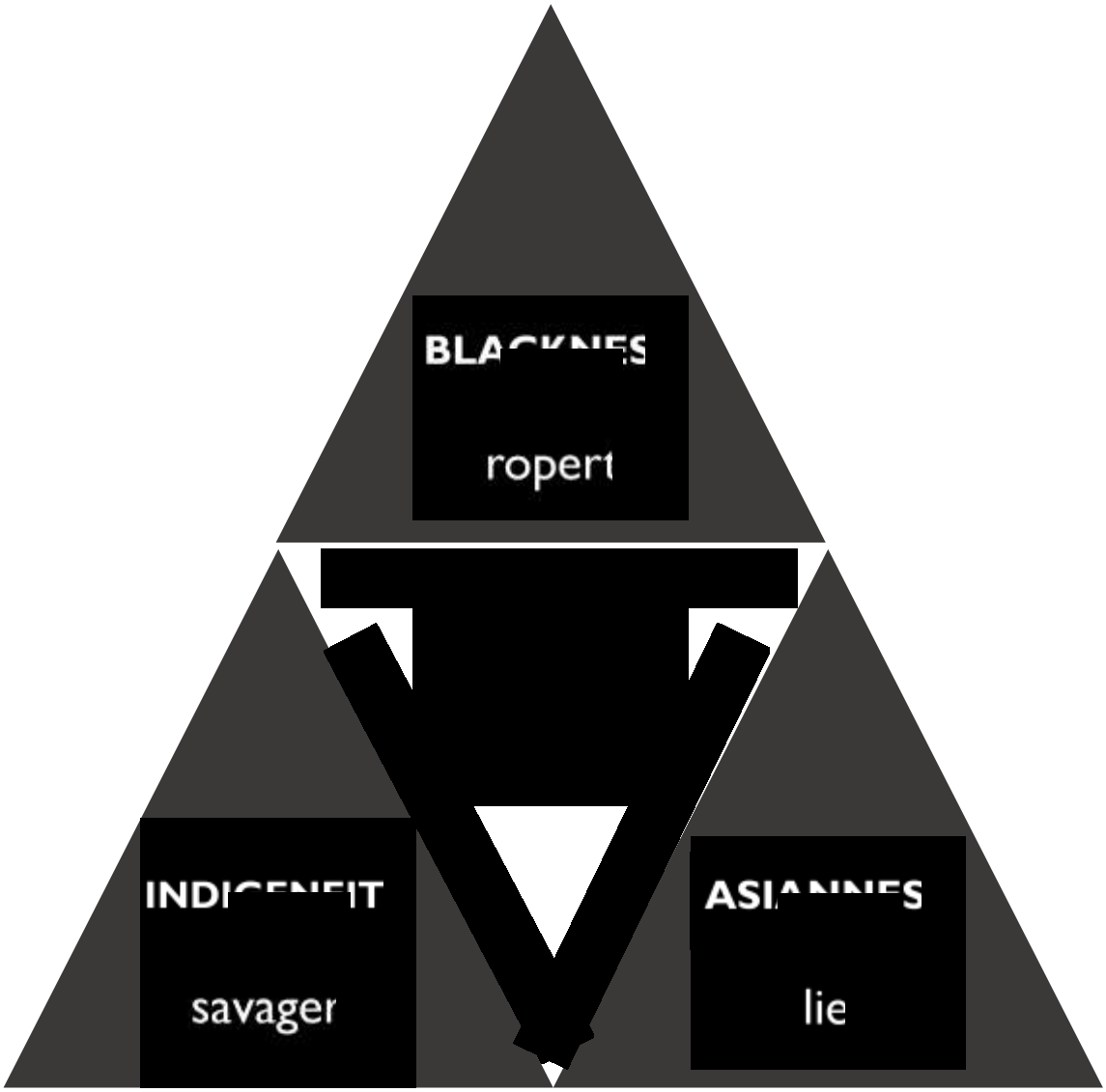

Figure 4. The three pillars of white supremacy

of oppression. According to Young, oppression is the obverse of justice. It systemically counteracts social groups' ability to develop and exercise their capacities and express their needs, thoughts, and feelings.

Rather than being the result of individual intent, oppression is systematically reproduced in the structural features of hierarchies and bureaucratic administration. These encompass market mechanisms, including the production and distribution of consumer goods, as well as liberal and "humane" economic, political, health-related, educational, and cultural institutions. It takes the form of unconscious and unquestioned assumptions, norms, habits, symbols, and media and cultural stereotypes woven through normal, ordinary interactions, reactions, and processes of everyday life. At the same time, the *conscious* daily actions of many individuals contribute to maintaining and reproducing oppression, as well. Such people are often well-meaning—just doing their jobs or living their lives—and do not understand themselves as agents of oppression. Oppression is thus the vast and deep injustice that members of some groups collectively suffer as a consequence of people—who do not necessarily share the same circumstances or consequences, but may—simply following the rules.

Oppression's five faces—cultural imperialism, exploitation, marginalization, powerlessness, and violence (each of which is described below, under Young's three broader categories of culture, organization of labor, and decision-making structures and processes)—apply not only to racialized difference, but cut across all dimensions along which social groups are differentiated and suffer under oppression, all of which intersect. Thus, they are inherently intersectional. For example, those who identify or are classified as middle-class White gay men, on average, experience lower levels of economic exploitation relative to some other groups; they do, however, experience cultural imperialism and the ongoing threat of violence, including sexual violence and exploitation. Identifying or being classified differently with respect to any one of those dimensions of identity would change how they experience the others.

Culture. (1) Cultural imperialism is the universalization of the super-ordinated social group's experience and culture and its establishment as the norm. The cultures of

other groups are simultaneously invisible, seen only in contrast to the super-ordinated group (as falling short of it) and stereotyped. Those living under cultural imperialism find themselves defined from the outside by those with whom they do not identify and who do not identify with them. To operate within that culture, however, they must to some extent internalize its perspectives.

Organization of labor. (2) Exploitation is the process through which the results of the labor and energy expenditure of one social group are steadily transferred to benefit another, continuously reproducing the dynamic of super-ordination and sub-ordination between them. (3) Marginalization is exclusion from the formal system of labor altogether. It is achieved in wealthy countries through policies and enforcement of incarceration, exclusionary immigration policies, and reservations. Members of this growing underclass are often dependent on the state.

Decision-making structures and processes. (4) Powerlessness is the lack of formal authority. The powerless must take orders and rarely have the right to give them. They have little opportunity to develop and exercise skills, have little to no work autonomy, and exercise little creativity or judgment in their work. Even in their private lives, they stand under the authority of professionals. (5) Violence is less about specific acts of violence themselves than about the threat of potential violence that often serves to keep members of particular groups sub-ordinated. Even though members of super-ordinated groups sometimes experience violence simply because of their group status, they know that if the incident goes to trial, they will likely be supported by a justice system that was created by, is composed of, and tends to be sympathetic toward members of their group.

The five faces of oppression describe five types of structurally mediated processes, dynamics, relations, exchanges, or flows of different types of stock or capital that characterize oppression. Because they describe dynamics of super-ordination and sub-ordination—the structural mechanisms mediating them are inherently asymmetrical, stratified, or hierarchical. As such, these types of relations (re)produce difference exponentially over time.

Summary of Critical Theories of Systemic Oppression

This segment provided an overview of CRT, including elaboration of its central concepts of interest convergence and differential racialization and its related concept of whiteness. These three concepts call attention to the material exchange of capital, even if through currency that is psychic or discursive, involved in systemic oppression. In reviewing heteropatriarchy and the three pillars of white supremacy and the five faces of oppression, this segment also provided frameworks for understanding and theorizing not just the material of exchange, but also the nature of the exchange—first in the process of racialization and second in the (re)production of difference more generally.

Applying Systems Thinking and Critical Theories

The following section aligns four approaches to inequality that use concepts from systems thinking and critical theories of systemic oppression, even if implicitly. Coming from economics, sociology, and anti-racist praxis, their alignment provides an alternative to multi-level analyses that address incongruence without the benefit of a critical understanding of systemic oppression.

Exemplifying the discipline of stratification economics, Oliver and Shapiro's (2006) analysis of the Black-White wealth gap corresponds with both systems thinking and critical theories of systemic oppression, although they refer to neither directly. Their emphasis on structural relations and the passage of time in the production of racialized differences in wealth parallels both systems thinking and critical theories of systemic oppression. Moreover, their summary that "wealth…captures the historical legacy of low wages, personal and organizational discrimination, and institutionalized racism…" (p. 5)—which they organize further into racialized state policy, economic detours, and the sedimentation of inequality—represents an expansion of the boundaries of analysis beyond individuals and individual lifetimes, as is characteristic not only of stratification economics, but also of both systems thinking and critical theories of systemic oppression.

In situating individual- and institutional-level dynamics within larger structures, Oliver and Shapiro approximate social-ecological systems theories, in particular, which expand boundaries of analysis. Their 2006 analysis of the racial wealth gap also aligns

with critical theories of systemic oppression through its establishment of a dialectical relationship between those classified as Black and those classified as White. Finally, their notion of the sedimentation of racial inequality refers not to another level of analysis or type of factor contributing to the racialized differences in wealth, but rather to the cumulative interaction—feedback loops—among factors categorized within racialized state policy and economic detours. As such, it resembles the systems concept of emergence, in which multiple causes and effects and mutual causality cause the whole to be greater than the sum of its parts (Flood, 2010; Laszlo & Krippner, 1998; Menendian & Watt, 2008; Thomas & Parsons, 2017; B. Williams & Hummelbrunner, 2011).

Reskin's (2003) mechanisms-based model of ascriptive inequality focuses on mechanisms that link multiple levels of analysis, like Henry and Mark (2003) and Mark and Henry (2004). Unlike Henry and Mark and Mark and Henry, however, it calls attention to the structural mechanisms that mediate interaction among levels—exacerbating or mitigating intergroup inequality. powell, Heller, and Bundalli's (2011) spheres of systemic racialization focus less on the mechanisms linking levels, but specify the dynamics within each level of analysis as being within or among institutions on one hand, and within or among individuals on the other. They also narrow the application of their multi-level analysis to the process of racialization. Byng's (2013) social process theory integrates the mechanisms that link levels of analysis with the "flexibility and persistence of racism and race" by positing that "racism is a social process where the meanings of race identities are traded across macro, meso, and micro levels of society" (p. 705). Important in her social process theory is the passage of time and role of discourse; she proposes that the power of racism lies in the transmission of racial meanings across society (Byng, 2013). Table 1 summarizes four pieces of literature that examine inequality using a multi-level and context-sensitive analysis, but that differ from Kirkhart (2011), LaFrance, Nichols, and Kirkhart (2012), and Fitzpatrick (2012) through their use of concepts from critical theories of systemic oppression. The studies are organized in ascending order in terms of their explicit incorporation of critical theories of systemic oppression and systems thinking, particularly with respect to discourse and

implicit bias. In other words, Oliver and Shapiro (2006) name phenomena that correspond with systems concepts whereas powell, Heller, & Bundalli (2011) name systems concepts explicitly.

Table 1. *Implicit and explicit application of systems thinking and critical theories to inequality*

Social-Ecological Levels	Oliver & Shapiro (2006)	Reskin (2003)	Byng (2013)	Powell et al. (2011)
Most external to individual	Racialized State Policy: • Difference between being property and potentially owning property • Differential access to homesteading, specific occupations, and residential and educational segregation • Current policies that discourage wealth accumulation	Societal Mechanisms: • Public systems of education, health, justice, labor, transportation, etc. • Normative considerations within industries • Collective bargaining agreements • Enforcement agencies	Macro: • Where categorical racial identities are written into society in the form of laws and policies • Where rationales and methods that can be used to advance social groups' interests and resource access are codified	Structural: • Dynamic • Cumulative among institutions • Durable
	Economic Detours: • Differential access to housing, credit, and participation in the open market	Organizational Mechanisms: • Shaped by societal mechanisms • Policies and practices through which organizations somehow link workers' ascriptive characteristics to work outcomes • Can check or permit the effects of intrapsychic and interpersonal mechanisms	Meso: • Where public discourse legitimates how racial meanings are enacted at the macro and micro levels of society • Media, university curricula, employee training programs, the questions on standardized exams, etc. create public knowledge and "common sense"	Institutional: • Bias in policies and practices in schools, agencies, etc.

Levels	Oliver & Shapiro (2006)	Reskin (2003)	Byng (2013)	powell et al. (2011)
	• Interpersonal violence against successful businesses owned by those classified as Black	Interpersonal mechanisms: • Can convert intrapsychic mechanisms into differential behavior, depending on whether organizational mechanisms intervene to blunt or eliminate their effects		Interpersonal: • Bigotry and implicit bias between individuals
Most internal to individual	Individual: • Schooling • Wages • Savings	Intrapsychic Mechanisms: • Largely beyond empirical observation • Unnecessary to observe for remedying intergroup inequality	Micro: • Where macro- and meso-level definitions of race identities come to rest on what actually happens to someone and how they interpret those events • Site of lived experiences and identity negotiations for individuals and groups	Internalized: • Beliefs within individuals, including stereotype threat

This collection of multi-level frameworks provides several points of entry for understanding the racial stratification in the nonprofit industry within which evaluation is embedded and remedying the incongruence observed between program participants and program evaluators. Of the four pieces, only Byng (2013) and powell, Heller, and Bundalli (2011) explicitly draw from critical theories of systemic oppression, and only powell, Heller, and Bundalli—writing for an audience of practitioners rather than in a peer-reviewed journal—explicitly draw from systems thinking. Still, the four items illustrate the application of concepts from systems thinking and critical theories of systemic oppression in ways suggested by Lee and Gilbert (2014). Figure 5 links the theoretical and analytical frameworks to the problem statement through concepts from critical theories and systems thinking.

Constructions of Difference		Research Frameworks	
Liberal ***Difference is…***	**Critical** ***Difference is…***	**Theoretical** ***Systemic Oppression***	**Analytical** ***System Dynamics***
• Natural • Inherent	• Dialectically produced	• 3 Pillars of White Supremacy	• Differences/Perspectives
• Individual	• Relational • Structurally mediated • Within and between groups		• Containers/Boundaries
• Random ("diverse") • Lateral ("cross-cultural")	• Asymmetrical • Hierarchical • Stratified	• 5 Faces of Oppression	• Exchanges/Relations/Processes • Flows of stock (capital/currency)
• Pre-social • Static	• Processual, dynamical • Cumulative • Fluid		• Feedback loops

Figure 5. Alignment of theoretical and analytical frameworks

Conclusion of Chapter Two

Expanding the boundary around the incongruence observed between program participants and program evaluators, Chapter Two provided an overview of the history and context surrounding racialized difference as manifested in disparities and disproportionalities societally and within the nonprofit industry. Through its focus on racialized difference in the evaluation literature, it distinguished between approaches to the incongruence that reinforce the normativity of whiteness and those that disrupt the ongoing process of structural racialization. The latter reverse the gaze and claim the means and ends of knowledge production for purposes of self-determination. Finally, Chapter Two reviewed analytical and theoretical alternatives for examining evaluation's understanding of and response to the incongruence it observes. Chapter Three details the research methodology, which includes textual analysis of the field's academic literature; analysis of changes in language over time; analysis of changes in relation to the professional association's documents; and interviews with evaluation scholars and practitioners most closely involved with the field's construction of racialized difference.

CHAPTER 3: METHODOLOGY

Chapter One contrasted two constructions of difference—liberal constructions and critical constructions. In the former, particularly in a climate of racialized neoliberalism, difference is constructed as natural and individual while structures are obscured. In the latter, difference is constructed as produced and reproduced over time through socio-economic relations and decision-making processes. These relations and processes, characterized by super-ordination and sub-ordination, are mediated by asymmetrical structural mechanisms. Chapter One also introduced an enabling conceptualization of justice, which—in contrast to oppression—is characterized by conditions that enable social groups sub-ordinated by systems of oppression to exercise collective agency in creating more equitable structures. As such, an enabling conceptualization of justice encompasses both structure and agency.

Chapter Two presented literature detailing some specific processes through which racialized difference has been and continues to be produced in the USA, illustrating the limitations of individual agency relative to structural forces. It also described the tension between individuals and structures as central to the nonprofit industry from its inception through the present day and raised questions about the extent to which the industry's racially stratified structure has amplified or counteracted collective agency exercised by racially otherized groups. While it revealed a similar ambivalence in the evaluation literature with specific respect to social group identification and whiteness, it did identify one body of evaluation literature—largely but not exclusively representing indigenous and decolonizing frameworks—that focuses on self-determination or collective agency among racially otherized groups. Finally, Chapter Two showed how—in the absence of critical theories of systemic oppression—even evaluation literature utilizing contextual or multi-level frameworks to examine racialized difference fails to consider the processual,

dynamical nature of difference and the importance of structural arrangements in mediating relations among groups.

Organization of Chapter Three

Chapter Three details the methodology that this dissertation employed to link critical theories of systemic oppression with systems thinking to "… move beyond the individual and consider the interconnected parts of the ecosystem (i.e., organizational policies and practices, institutional norms) that made it necessary for the programs in the first place…" (Lee & Gilbert, 2014, pp. 106-107). Relying heavily on Pearce (2015) and to a lesser extent on Hall and Howard (2008) for its approach to research design, Chapter Three starts with an overview of the dissertation's philosophical and practical positions. It proceeds by detailing the approach to data collection, analysis, and interpretation according to Teddlie and Tashakkori (2006). It concludes by attending to the study's delimitations and ethical considerations, which are informed by Guba (1981), Guba and Lincoln (1994), and Lather (1986), and by reflecting on the researcher's role and positionality.

Philosophical and Practical Positions

Pearce (2015), among others (e.g., Mertens, 2016), describes the philosophical positions of research as consisting of axiological, ontological, and epistemological dimensions. This dissertation's attention to the social construction of difference as well as the interplay between agency and structure, both of which were reviewed at the beginning of this chapter, reveals ontological and epistemological positions rooted in social constructionism and intersubjectivity, respectively. It is openly value-based, or ideological, in that it centers relatively subjugated knowledge—"the view from the basement" (Mills, 1994, p. 120)—and aims to advance justice as conceptualized by sub-ordinated social groups, as stated in Chapter One (Hesse-Biber, Rodriguez, & Frost, 2015; Lather, 1986). The practical positions Pearce (2015) describes are the research logic, locations on axes of inquiry, and methods, each of which is described below.

Logic. Having identified the research problem as a focus within evaluation on dynamics within and among individuals at the expense of dynamics within and among

institutions, this dissertation contributes to a body of research that increases understanding of the latter, building empirical evidence that can inform remedies that focus on structural dynamics. To do so, it asked:

1. How has the U.S. scholarly evaluation literature constructed racialized difference?
2. How has that construction changed since the field began formalizing in the early 1970s?
3. How does that trajectory relate to the systems surrounding evaluation?

The design underlying this dissertation research sought convergence and synergy across multiple strands that mix approaches to inquiry (Hall & Howard, 2008). In much the same way that racialized difference—and difference along other dimensions that are salient in the organization of contemporary U.S. society—continues to be produced and reproduced dialectically, the difference between "quantitative methods" and "qualitative methods" has been dialectically produced (Pearce, 2015). Not only are quantitative and qualitative constructed as binaries, but they are also personified such that quantitative is associated with rational reasoning and qualitative is associated with creativity and feeling (Pearce, 2015). In fact, they are gendered and racialized. Resistance by social groups sub-ordinated by decontextualized and dehistoricized representations of their lived experience has been reduced to stereotypes that they "naturally" gravitate toward qualitative as opposed to quantitative data, each of which is assumed to correspond with particular philosophical and practical positions.

To be clear, the sequential multi-strand, mixed approach of this research was born not out of a pragmatist or critical realist paradigm, but out of an intersectional paradigm rooted in critical theories of systemic oppression (critical race theory, postcolonial theory, and feminist theory) and systems thinking. Intersectionality as a paradigm encompasses five dimensions of complexity: (1) among categories; (2) between categories; (3) within categories; (4) in a given historical moment as well as over time (time dynamics); and (5) in terms of how individual-institutional interactions shape social group identification and classification (Hancock, 2013). Mixed research corresponds with intersectional theory's critical construction of identity and systems of oppression in that it offers opportunities to

collect and analyze data in ways that account for "power, history, and context" (Hankivsky & Grace, 2015).

Intersectionality and systems thinking both correspond with mixed research in that all three reject dichotomous, binary thinking; foreground contradictions and tensions; and allow for interplay between agency and structure (Hankivsky & Grace, 2015; Lane, 2001a, 2001b). Just as intersectional theory recognizes that the experience of multiple identities is greater than their sum, multi-strand research that mixes approaches by designing interaction among them achieves the systems concept of emergence or synergy—the idea that two or more options interact so that their combined effect is greater than the sum of their individual effects (Creswell & Plano Clark, 2017, referring to Hall & Howard, 2008). In addition to being methodologically mixed and paradigmatically intersectional, however, this dissertation is transdisciplinary in that it drew from academic literature beyond evaluation and literature beyond the academy. Specifically, it drew from literature by john powell (powell, 2010, 2013; powell, Heller, & Bundalli, 2011) that brings together critical theories of systemic oppression, systems thinking, and evaluation to support movement organizing efforts toward racial justice (Szostak, 2015).

Locations on axes of inquiry. The interaction that this dissertation research sought among strands that inhabit varied locations on the subjective-objective and inductive-deductive axes characterizes it overall as intersubjective. It was similarly planned as an effort to explore content and explanations that have not previously been considered and is thus inductive overall. No attempt is made to generalize beyond the evaluation context, although findings may be transferable to other practice-oriented disciplines. Finally, this dissertation research is highly reflexive in that research memos were used to document all phases of data collection and analysis. This included the researcher's awareness, reflection, and transparency regarding her relationship to the topic and participants as well as how her own social group classification and identification, position in society, and interests may have influenced her approach to all phases of research, from the choice of topic and formulation of questions through the

inquiry process to the interpretation and use of findings. This process began with the dissertation's prologue, continues with a section on positionality after the segment of this chapter devoted to ethical considerations, and ends with an epilogue.

Strands of inquiry. Two primary strands of inquiry interacted to answer the three research questions listed earlier—critical discourse analysis (CDA) and critical systems thinking (CST)—both of which engage with the relationship between agency and structure (Fairclough, 2003; Lane, 2001a, 2001b). These were arranged sequentially in that results of the CDA informed conceptualization and implementation of the CST strand of inquiry, and results of the CST strand of inquiry informed interpretation of the CDA results. Importantly, neither CDA nor CST dictates an analytical procedure. Details regarding the conceptualization and implementation of the CDA and CST strands of inquiry follow and are summarized in Figure 6.

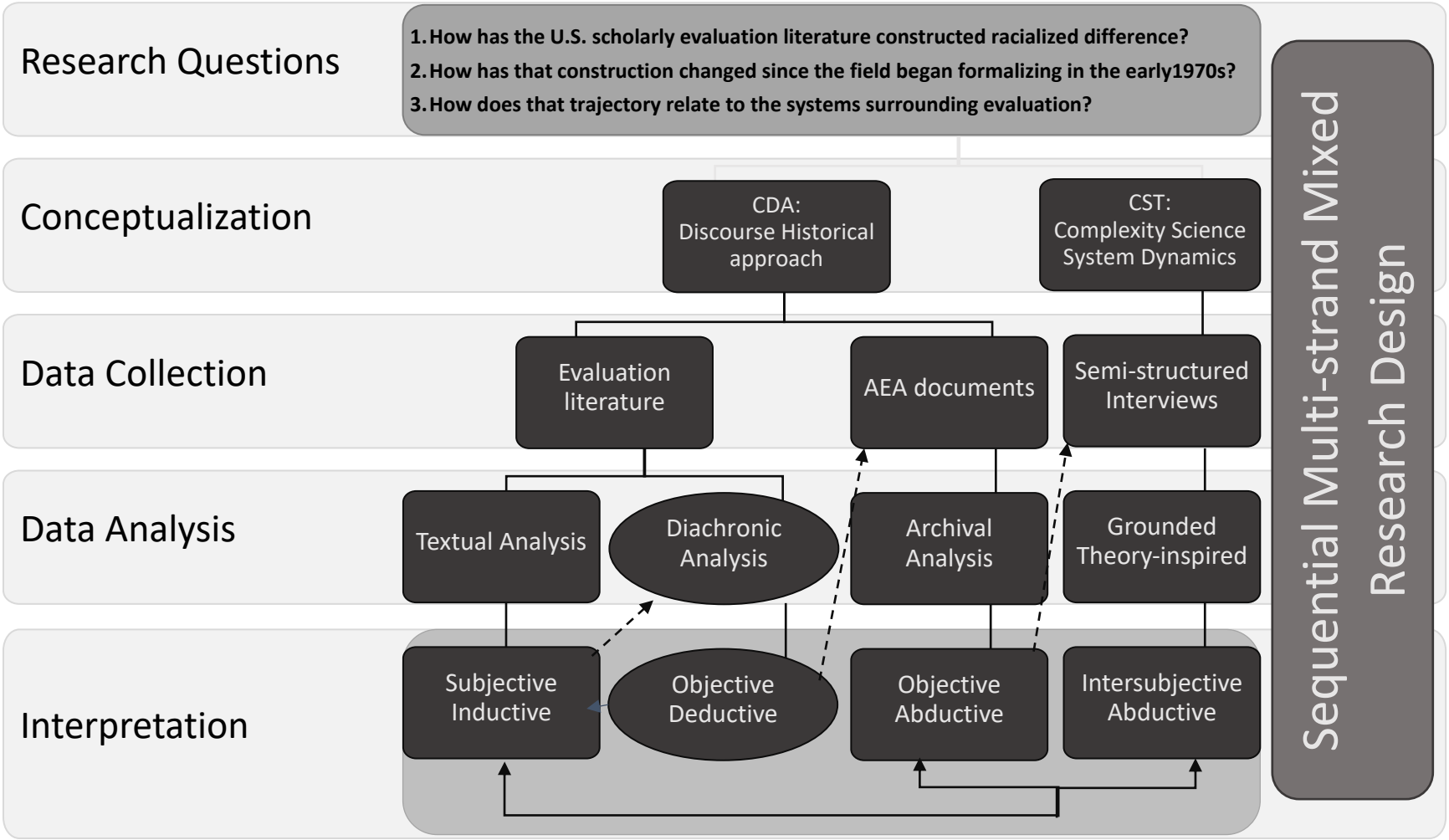

Figure 6. Summary of research design

Critical Discourse Analysis

Discourse encompasses meaning-making as an element of the social process, the language associated with a particular social field or practice, and a way of construing aspects of the world associated with a particular social perspective (Fairclough, 2012). Critical discourse analysis, like evaluation, is considered a transdiscipline (Fairclough, 2012). "It brings the critical tradition of social analysis into language studies and contributes to critical social analysis a particular focus on discourse and on relations between discourse and other social elements (power relations, ideologies, institutions, social identities, and so forth)" (Fairclough, 2012, p. 9).

Purpose and approach. This dissertation used the discourse-historical approach (Wodak, 2001a) and drew heavily from the scholarship of van Dijk (2015), who, like Wodak, "studies the way social power abuse, dominance and inequality are enacted, reproduced and resisted by text and talk in the social and political context" (O'Keeffe, 2012, p. 448). Wodak and van Dijk consider discourse an interface between the micro- and macro-levels and discourse analysis a framework that relates discourse, cognition, and society (van Dijk, 1993, p. 98). Linguistic practices constitute social and political processes and actions—both discursive and non-discursive—in the fields surrounding them; they are simultaneously constituted by social and political processes and actions. Thus, the discourse-historical approach incorporates into the analysis and interpretation of (1) the immediate language of discursive events (talk, text, and other vehicles); (2) the intertextual and interdiscursive relationship between them and the larger genre; and (3) investigation of the historical context and fields surrounding them (meso-level theory), including through fieldwork and analysis of the ways in which language changes over time (Wodak, 2001a). It also includes (4) "grand theory," which Wodak defines as "the broader socio-political and historical contexts" (2001b, p. 29) in which the discursive practices are embedded and to which they are related.

Data collection and processing. The CDA strand of inquiry included three sources of text-based data (as opposed to interview data transcribed into text, which form the only source of data for the CST strand of inquiry). The first source of data included

all items aside from book reviews in peer-reviewed U.S. evaluation journals since the date of their inception, the earliest of which is 1977. The second source of data, which was used to measure diachronic change in language, consisted of the entire population of material (aside from book reviews) published in peer-reviewed journals in the USA, Canada, and the UK. Documents published by AEA comprised the third source of data. Sampling criteria for each of these data sets follow.

Sampling of data for textual analysis. Evaluation literature included in the textual analysis was purposively selected in largely the same way that literature was selected for the review of evaluation literature in Chapter Two, but with one important difference. Only peer-reviewed articles from U.S. evaluation journals that somehow problematized racialized difference were selected for inclusion in Chapter Two's literature review; in contrast, the pool of texts considered for textual analysis consisted of *all* items (aside from book reviews) in U.S. evaluation's peer-reviewed journals, including but not limited to academic articles, editorials, and announcements. Only those that represented a change in language or that introduced language that had not previously appeared in U.S. peer-reviewed evaluation journals were selected for inclusion in the textual analysis portion of this dissertation's CDA strand of inquiry.

To identify academic journal items that used language or discursive strategies that had not previously been used in the evaluation literature to refer to or discuss racialized difference in the U.S. evaluation context, a search for variations of terms referring to racialized difference (detailed below) was conducted in all five peer-reviewed, academic journals published in the USA that have "evaluation" in their title, including the two flagship journals affiliated with AEA. The *American Journal of Evaluation* (formerly *Evaluation Practice*), *Evaluation and Program Planning*, *Evaluation Review*, the *Journal of Multi-disciplinary Evaluation*, and *New Directions for Evaluation* (formerly *New Directions for Program Evaluation*) were included in the search. Because the purpose of this dissertation was to examine the construction of racialized difference in and through evaluation with a particular focus on social programs in the U.S. nonprofit industry, journals published in Canada, Australasia, and the UK were excluded from the search.

Journals specific to education, assessment, public health, and substance abuse were also excluded.

The search was conducted from the journals' webpages[46] starting from the various journals' inception—the earliest of which was 1977—through December 31, 2017. This date was selected to allow the findings from the textual analysis to inform the diachronic analysis, archival analysis, and interviews. As with the literature review in Chapter Two, the search terms included variations of the words "race" (i.e., races, raced, racial, racialize, racializing, racialized, racism, racist), "white" (i.e., whites, whiteness, white privilege, white supremacy), and "colony" (i.e., colonialism, colonial, postcolonial, anticolonial, de/colonizing, de/colonized, de/colonization). Terms and phrases that are often used to refer to racially otherized groups—for example, "African American" or "Black"—were not searched for directly.

Any item that contained variations of one or more search term in its full text was considered for textual analysis. As with the literature review, results that were specific to international development, education/assessment, and public health were excluded in favor of those that either focused on U.S. nonprofit settings specifically or addressed evaluation more generally. The pool of resulting journal items was coded with their dates of publication and sorted chronologically. Each was scanned and coded initially for references to and treatment of racialized difference, including the terms it used as well as its linguistic and inter-textual characteristics (Fairclough, 2003). Those that represented a shift from earlier references to racialized difference were included in the textual analysis sample.

Sampling of data for diachronic analysis. All items (aside from book reviews) from all five peer-reviewed evaluation journals published in the USA listed above were included in the diachronic analysis of language referring to racialized difference (operationalized below, under Analysis). In addition, items from peer-reviewed

[46] The earliest journal material was scanned from paper newsletters as images without recognizable characters and thus searches from the journal websites may not find all search terms. Still, manual comparison of the results of this search strategy and searches conducted from databases such as the Web of Knowledge suggest that searches from the journal websites produced more accurate results.

evaluation journals published in English-dominant countries beyond the USA (*Canadian Journal of Program Evaluation* and *Evaluation*) were included in hopes of examining the extent to which the language trajectory may be specific to the USA and potentially related to systems surrounding evaluation in the USA or if they were reflective of larger trends. *The Evaluation Journal of Australasia* was not included owing to the search features of the journal's website at the start of data collection.

Titles of journal items identified as containing the search terms were entered as rows in an Excel spreadsheet with columns for their years of publication, author names, journal titles, volume numbers, issue numbers, page numbers, AEA affiliation, and countries of publication. Columns were also created for each search term identified through the textual analysis as a signifier of racialized difference in the evaluation literature. One column for each search term was for the binary (0/1) code of whether or not the item contained it. Initial counts by search term and by year were conducted using a pivot table. Another column, created only for search terms with wide variation in meaning, was for an ordinal variable (1 through 4) regarding relevance and substantiveness. To ensure that the diachronic analysis reflected counts of items that used the terms identified through the textual analysis to refer to differences in social group—including racialized difference—and were not artificially inflated by other usages or by usage in the notes or references, search results for terms whose counts were especially high and whose meaning might be unrelated to difference in social group (such as proper nouns and those in phrases like "space race," "learning culture," or "colony of bacteria") were manually scanned[47] and coded with respect to the search terms' relevance and placement. Only items that were identified as using the terms to refer to racialized difference (perhaps among other dimensions of difference) were included in the diachronic analysis. Finally, columns were created for abstracts and direct links to the content.

---

[47] Manual scans of identified material for purposes of determining whether to include them in the sample might have been affected by the fact that the earliest material consisted of printed newsletters that were scanned as images. Thus, to the extent possible, material scanned as images was converted to text using optical character recognition.

In addition to counts of items containing search terms identified through the textual analysis, counts of total items (aside from book reviews) published each year were generated from the tables of contents of each issue of each journal included in the study since the journal's inception. These annual counts were entered, by journal, into a separate tab of the same Excel spreadsheet used for search terms.

Sampling of data for archival analysis. Results of both the textual analysis and the diachronic analysis revealed decisive authors, documents, and periods during which the relative frequency of specific terms used to signify racialized difference in peer-reviewed, English-language evaluation journals increased or decreased. These results guided the identification of additional items in AEA-affiliated journals (for example, conference proceedings and calls for papers) as well as documents published by AEA that might explain the changes in language revealed through both the textual analysis and the diachronic analysis. The search for related journal items was conducted from the *American Journal of Evaluation* and *New Directions for Evaluation* websites. The search for professional association documents was conducted on AEA's website. Documents internal to AEA were not included in the archival analysis.

Analytical procedure and methods. The analysis of each set of data collected for the CDA influenced the collection of the next and interpretation of the others as follows.

Textual analysis. The sample selected for textual analysis, already broadly coded for topic and publication date, was recursively coded at the sentence level for content (terms used to signify racialized difference) and discursive strategies. Content codes were subjectively and inductively generated in that, while rooted in the specific texts selected for analysis rather than existing scholarship, they necessarily had been influenced by the researcher's positionality (which includes an academic and professional background in justice-oriented work, detailed separately in this chapter). The sample was also subjected to a close reading in terms of discursive strategies (self-representation, perspectivation, mitigation, argumentation) and linguistic features (agents, time, tense, modality, syntax) (Wodak, 2001a, 2001b). Additionally, codes for term(s) used to refer to racialized

difference were objectively and inductively generated *in vivo* from this close reading. They provided the terms searched for and analyzed diachronically to describe the trajectory of evaluation literature's construction of racialized difference. Results of the textual analysis also guided the search for additional journal content and professional association documents from within evaluation as well as literature from outside evaluation to interpret the results (Kelle, 2007).

Diachronic analysis. Annual counts of items containing relevant, substantive uses of variations of each search term identified through the textual analysis were divided by annual counts of the total number of items (aside from book reviews) published to arrive at annual percentages of items that contained variations of each search term. A pivot table and chart allowed these percentages to be analyzed (and subsequently reported) longitudinally—overall, by publication, by country of publication, and using various combinations of search terms. Because the diachronic analysis started with keywords identified through the textual analysis, the diachronic analysis can be considered deductive; because the keywords' relative frequency each year was calculated as a proportion based on an automated count of items that contained one or more keywords relative to the total amount of items published each year, it can also be considered objective, although the criteria for what to count cannot. Results of the textual analysis and diachronic analysis together were used to identify pivotal years, authors, and journal items, and professional association documents in evaluation's construction of racialized difference. They formed the basis of the search for archival material and documents from AEA's journals and website.

Archival analysis. AEA documents and additional journal items identified through results of the textual analysis and diachronic analysis were coded with their year of publication and content-related codes generated during the textual analysis to link them topically and chronologically with results of the textual and diachronic analyses. By iteratively relating discursive events (texts) and meso-level theory, the archival analysis can be considered abductive rather than purely inductive or deductive (Wodak, 2001a). Results of the textual analysis, diachronic analysis, and archival analysis were used to

generate the list of scholars and practitioners closest to the construction of racialized difference within evaluation.

Interpretation and use of results. Results of the textual analysis are described in narrative form, chronologically within broad thematic category by selected text. They informed the diachronic analysis and archival analysis as well as the CST strand's purposive sample of interview participants.

A pivot table and chart allowed relative frequencies of terms used to signify racialized difference to be analyzed (and subsequently reported on) longitudinally—overall, by AEA affiliation, by publication, by country of publication, and using various combinations of keywords. Results of the diachronic analysis were used to check preliminary inferences about the trajectory of constructions of racialized difference in evaluation drawn from results of the textual analysis and to inform the archival analysis as well as the CST strand's purposive sample of interview participants. Archival analysis of AEA journal content and documents informed initial results of the textual and diachronic analyses, forming the basis of the interviews that were conducted in the CST strand of inquiry, whose design follows.

Critical Systems Thinking

Critical systems thinking (CST), like CDA, encompasses a wide range of approaches and practices. As described in Chapter Two, it considers the establishment of boundaries an ethical issue and advances boundary critique such as that represented by this dissertation's expansion beyond dynamics within and among individuals. Beyond a commitment to systems thinking in general, it is committed to questioning the assumptions, values, and theoretical underpinnings of systems and systems approaches; considering societal or organizational expectations and acceptance of any particular systems approach at any particular time; attending to people's collective development of their potential; and employing theoretical and methodological complementarity, not unlike the intersectional paradigm and transdisciplinarity discussed earlier in this chapter (Flood, 2010; Laszlo & Krippner, 1998).

Purpose and approach. The CST strand of inquiry was intended to disconfirm, support, and explain or extend the relationship of the trajectory of the construction of racialized difference within evaluation to the systems surrounding it that was preliminarily established through the CDA strand. Its emphasis on the meso-level of theory could be considered part of the CDA strand and is distinguished from it only because it represents a critique of boundaries that is characteristic of CST (Flood, 2010; Lane, 2001b). Rather than individual self-reflection regarding power and difference, which does appear in U.S. evaluation literature and professional association documents as described in Chapters One and Two, this dissertation represents an interest in collective reflection by the field of evaluation on its role in reproducing or disrupting the racial stratification of the systems surrounding it.

The CST strand was thus concerned with increasing understanding of the systemic patterns surrounding evaluation in an effort to identify patterns that reinforce and those that counteract the efforts of social groups that are racially otherized and sub-ordinated by multiple systems of oppression. To conceptualize the systems surrounding evaluation and generate understanding of its behavior, this strand drew from system dynamics and complexity science—using expanded boundaries, structurally mediated relations, and social group perspectives or interests as an organizing framework. Systems theorists note the value of qualitative system dynamics, which shares much with the soft systems methodologies from which CST arose, for structuring and analyzing "ill-defined" systems (Fredericks, Deegan, & Carman, 2008). Qualitative modeling may propose interrelations without ever resulting in the simulation of a mathematical system dynamics model (Luna-Reyes & Andersen, 2003).

Data collection and instrumentation. Interviews of evaluation scholars and practitioners most closely connected to evaluation's construction of racialized difference—as indicated through the textual analysis, diachronic analysis, and archival analysis—elicited their mental models of the systems surrounding evaluation (Martinez-Moyano & Richardson, 2013). Mental models figure prominently in both CDA, where

they are associated with cognition (van Dijk, 1993), and systems thinking, where they are associated with perspectives (Lane, 2001a; Thomas & Parsons, 2017).

Sampling and participant recruitment. Interviewees were members of a purposive, theoretical sample of scholars and practitioners most closely connected to evaluation's construction of racialized difference since the field's inception. An original list of 80 such scholars and practitioners (listed in Appendix A) identified through the results of the textual analysis, diachronic analysis, and archival analysis was narrowed to a list of 20 (listed in Appendix B), with the intended maximum number of interviews being 15 and the intended minimum being data saturation. Members of the list were invited to participate (see Appendix C) using information available through AEA's website and journals. Interviews took place until data saturation was reached. A total of 11 interviews were conducted.

Implementation, setting, and instrumentation. All interviewees received preliminary results of the diachronic and archival analyses electronically in advance. All interviews took place by phone or Zoom and were audio and/or video-recorded and transcribed. They were largely unstructured interviews that began with participants identifying the social groups with which they identified most closely, after which the researcher requested their reaction, interpretation, and storytelling about the observed trajectory (see Appendix D). The researcher shared an image of the social-ecological model and used probes to clarify elements of the story that might relate to participants' perceptions of the system's boundaries, structural mechanisms, relations, and perspectives or interests. For example: When did certain events in their responses happen? Where did they happen structurally? While participants were asked to title their interviews with something other than their name, most declined to do so and in fact preferred that their names be associated with their interviews. Interview transcripts were initially coded to allow the researcher to determine the point of data saturation.

Analytical procedure and methods. The researcher plotted the main events according to each participant on the same timeline to begin a process of analysis inspired by grounded theory and detailed by systems theorists (Kim & Andersen, 2012).

Grounded theory's use of axial coding, in which initial (open) codes are examined and coded in terms of their relations with each other, lends itself to an understanding of relations in terms of time, space, and causality that is characteristic of complexity science and system dynamics. Because collection and analysis of the CST interview data informed and were informed by results of the CDA strand, the CST strand of inquiry can be considered intersubjective. Because it relied on constant comparison of the data with emerging ideas, which guided the search for additional literature, it can be considered abductive (Reichertz, 2007).

Meta-interpretation and action stimulus. The CST-strand interviews resulted in a qualitative model of the complex adaptive system surrounding evaluation that disconfirmed, supported, explained, and extended preliminary results from the CDA strand. The qualitative model identified patterns that enabled sub-ordinated social groups to produce knowledge and exercise collective agency on unjust structures as well as those that counteracted such efforts.

Delimitations and Disadvantages of Research Design

This study operationalized the trajectory of constructions of racialized difference in evaluation as changes in the relative frequency of journal items containing variations of particular terms rather than to changes in the frequency of their use *within* journal items. It similarly limited itself to scholarly print and online publications as opposed to evaluation textbooks, conference presentations, professional development opportunities and materials, blogs and list-serves—whether affiliated with AEA or not. Additionally, this study was intentionally of items in U.S. peer-reviewed journals that contained the word "evaluation" in their titles. It is possible that peer-reviewed journals of related fields, such as human services or social work, contain literature that critically considers racialized difference in relation to program evaluation. The decision to delimit this source to U.S. peer-reviewed evaluation journals was more theoretically than practically motivated as it had the potential to illuminate what is structurally and institutionally sanctioned by the field of evaluation within the U.S. context of racialization.

Similarly, this study intentionally included contextual items, documents, and perspectives of scholars and practitioners who were affiliated with AEA, as the USA's professional association of evaluators. Although there are no other national professional associations for evaluators in the USA, there may be evaluators unaffliated with AEA—perhaps affiliated with local chapters, the Center for Culturally Responsive Evaluation & Assessment (CREA) at the University of Illinois-Urbana/Champaign, or voluntary organizations for professional evaluation internationally—who are advancing more critical constructions of race through their practice and scholarship. While the research focus on the construction of racialized difference in the USA likewise led to its exploration of U.S. journals, scholars, and practitioners, the research risked reinforcing the hegemony of U.S. evaluation internationally. At the same time, delimiting the U.S. context was complicated in that the identities and professional affiliations of scholars and practitioners are fluid. Many work, publish, and presumably influence and are inflenced by writing and thinking in more than one country or discipline.

The study did not include analysis of AEA's internal documents integral to the field's construction of racialized difference, including those produced by and for the association's Multi-ethnic Issues in Evaluation (MIE) Topical Interest Group (TIG). Rather, it relied on those published in AEA-affiliated journals and on AEA's website and on the perspectives of evaluation scholars and practitioners most closely involved. Importantly, it did not include the perspectives of other stake-holding participants in the system, including U.S. evaluation scholars and practitioners—whether racially otherized or racially normatized as White—who were not closely involved in the field's construction of racialized difference between 1977 and 2018. Nor did it include foundation and nonprofit leaders, staff, or nonprofit program participants, all of whom directly affect or were directly affected by evaluation's construction of racialized difference.

Ethical Considerations

Guba (1981) discusses rigor in terms of credibility, transferability, dependability, and confirmability, many of which are achieved through the same or similar means. The

first set of means includes reflexive awareness and documentation of the process, including the researcher's role, through memos and logs that can track and explain variance, whether between research design and implementation; results of multiple methods; or recordings and transcripts. The researcher's reflections and process were documented through memos, as were ongoing research decisions and variances identified through multiple methods and various checks. The second set of means for achieving rigor involves intentionality and coherence with respect to research design, sampling, and data. The rationale underlying the research design and purposive sample was detailed earlier. Coherence with respect to data was achieved through constant comparison and recursive coding in the analysis of textual and interview data. The last set of means for achieving rigor includes transparency with results and the pursuit of disconfirmation, achieved through thick description, convergence of methods, member checks, and peer debriefs. The CDA strand employed thick description, and disconformation was sought through the multiple methods. The processes for member checks and peer debriefs are described below along with the protection of human subjects.

With participants' informed consent, interviews were recorded and transcribed to provide an audit trail and allow for member-checks. Recordings were uploaded onto the researcher's personal, password-protected computer and deleted from the recording instrument. Transcripts were entered into an Nvivo qualitative software database on the same computer. The study's theoretical framework was presented at AEA's annual conference in 2016, and preliminary results were presented at the same conference in 2017. Near-final results were presented in 2018.

The interview sample did not include members of vulnerable populations, but rather members of a scholarly and professional community with considerable influence. Still, the pool of scholars and practitioners most closely connected to the field's construction of racialized difference is small, and its demographic markers systematically differ from those of the field overall. To avoid putting interview participants in a position of relative vulnerability within the field by being easily identifiable, interview participants were invited to describe the social groups with whom they identify, without

being asked direct questions about standard demographic characteristics. This was an attempt to ensure that the research centered relatively subjugated knowledge ("the view from the basement" [Mills, 1994, p. 120]) in generating understanding of the system surrounding them. Many but not all of the interviewees, however, chose to be identified by name rather than to name their interviews.

Researcher Role and Positionality

This section is written in the first person. Both CDA and CST value transparency regarding the researcher's role and positionality. My selection of CDA and CST, within a larger paradigm of intersectionality, theoretical framework of critical theories of systemic oppression, and analytical framework of systems thinking, as well as the topic of racialized difference in the context of nonprofit program evaluation, was heavily influenced by my relationship to justice-oriented movement organizing. It was also influenced by my experience as a racially otherized woman who has engaged in the implementation and evaluation of nonprofit and government programs and who continues to participate in the local and national U.S. evaluation communities, including AEA, sometimes alongside potential interview participants. The experience of families like mine, who emigrated from South Asia to the USA, attuned me to the produced nature of racialized difference, the influence of structural dynamics, and the intricacies of intersectionality as discussed below.

Before 1965, U.S. immigration policy limited migration and naturalization from countries outside Europe in an explicit attempt to maintain the White majority that was artificially produced through the genocide of peoples indigenous to the continent (Chan, 1991; Espiritu, 1992). South Asian men—primarily from Punjabi farming communities in India—whom the U.S. immigration quotas did accommodate were not permitted to bring wives or to marry women classified as White, and South Asian women were not permitted entry. The immigrants were excluded from mainstream society and denied citizenship for decades explicitly because they were not considered White. DNA and linguistic evidence suggests that some South Asians are of Caucasian ancestry, and some South Asians contested their racial classification by the U.S. government based on the

Aryan invasion theory underlying what is now called India. This contestation continued until the Immigration and Nationality Act of 1965, after which time East Asians and South Asians, in particular, experienced an increase in status (Prashad, 2000).

In the wake of Civil Rights legislation, the USA found itself behind the USSR in the space race (Prashad, 2000). Realizing it could not grow its own labor force trained in math and science in time to compete with the USSR, the USA once again opened its borders to immigration. Europeans were generally no longer emigrating to the USA, however, and so the 1965 immigration act allowed for immigration from Asia, explicitly prioritizing two-parent families whose heads had advanced degrees in math or science. Meanwhile, India had just graduated its first generation since gaining independence from Britain. Post-independent India had deliberately emphasized training in math and sciences in pursuit of economic development. It had also begun to grapple with centuries of caste oppression, in which a minute percentage of the population enjoyed nearly exclusive access to literacy, higher education, and positions of influence. While traditionally prohibited from accumulating wealth, Brahmins as the most literate group tended to be selected to occupy administrative positions during British colonization.

With India's post-independence implementation of a reservation system—comparable to Affirmative Action in the USA except that a quota of educational opportunities and government jobs is reserved for each caste according to its proportion of the total population—Brahmins and other oppressive castes that were afforded high status according to Hindu law and according to British colonial practices for the first time experienced the decoupling of perceived effort and reward that the majority in an densely populated country like India had experienced for centuries. A disproportionate number of post-1965 immigrants to the USA from India—including my parents—were members of these oppressive castes, trained in math and sciences and eager for opportunities in a country in which they believed their efforts would yield corresponding rewards.

Immigrants to the USA from oppressive castes came not only with high levels of education and jobs secured, but also with the experience of having exercised relationship and cultural capital in their homelands and with fall-back positions in their countries of

origin that generally included family and property. As a result, the apparent financial success of families like mine was not surprising. The structural factors that contributed to it, however, are obscured. Rather, the apparent success of Americans of South Asian descent—and of Asian descent more generally—is typically attributed to notions about "Asian culture," which is portrayed as singular and unchanging despite vast differences informed by class, caste, region, religion, and migration among several other factors.

Having published *The Negro Family: The Case for National Action* during the same year as the change in U.S. immigration laws, Daniel Patrick Moynihan and others struggled to reconcile the poverty he saw among families classified as Black with the financial success of families from South Asia who were comparably complected. The Model Minority Myth, brewing since the incarceration of Americans of Japanese descent during World War II, would now flourish. The myth suggests that the success of families classified as Asian American is attributable to the perceived value that "Asian culture" places on education and conformity. As such, Asian Americans have been differentially racialized—our racialization since the 1965 Immigration and Nationality Act represented a change from the earlier part of the century. "Asian American" as a classification developed in direct opposition to "Black" and, to a lesser extent, "American Indian" in a deliberate attempt to discredit these groups' longstanding efforts to change oppressive structural dynamics. This was part of an ongoing strategy to shift the focus from civil rights and government programs to cultural pathology and individual responsibility.[48]

---

[48] "'Let me be blunt,' Daniel [Patrick] Moynihan wrote in 1968, 'if ethnic quotas are to be imposed on American universities and similarly quasi-public institutions, it is the Jew who will be almost driven out.' The role played by the figure of the Jew in the 1960s was to be farcically adopted by the Asian American from the 1980s onward. And we heard it spectacularly from Ronald Reagan, who called Asians 'our exemplars of hope and inspiration.' …. The 'Jew' and the 'Asian American' provide a singularly useful way to attack the problem of equity. Phrased in terms of 'overrepresentation' and 'merit,' these minorities, it is argued by some, would be hurt by social engineering since they are (1) already overrepresented in the professions and (2) they would face quotas that would impinge on their métier. During a Heritage Foundation event on affirmative action [*sic*] in the 1980s, Representative Dana Rohrabacker (Rep-CA) had the bad taste to say that he used Asians as 'a vehicle to show that America has made a mistake on affirmative action [*sic*].' Asians are used in this instance, then, as a weapon against the most modest form of redistribution devised by the state." (Prashad, 2001, p. 42)

At the same time, those who benefit from the current asymmetrical and racially stratified structural dynamics have used the Model Minority Myth to discredit Asian American experiences of structural racialization. The apparent success of those classified as Asian American continues to be misrepresented to suggest that the real threat to the advancement of those classified as Black and American Indian—and in fact to poor people classified as Whites—is not White wealth, but rather, immigrants and foreigners (Wu, 2002). However, even the wealthiest families classified as Asian American enjoy less financial and professional success relative to comparably situated families classified as White when compared within education level, number of earners, and number of family members to support. Additionally, increasing proportions of families—particularly since the 1980s—have arrived in the USA from Asia under much less favorable structural conditions. Indeed, Southeast Asian Americans' experience of forced migration and assimilation in the USA as refugees from the U.S. war in Southeast Asia—an imperial war against communism—was in many ways more similar to the experience of those classified as American Indian and Black than it was to the experience of Asian Americans who voluntarily migrated to the USA seeking financial opportunities. The Model Minority Myth masks differences in terms of success within U.S. educational systems and economic systems among those classified as Asian American as well as differences in these underlying conditions. Finally, the value that "Asian culture" places on conformity was also distorted—those classified as Asian American, like our Asian ancestors, have dissented, organized, and built alliances with other racially otherized groups since our arrival in the USA, but even members of those other groups tended not to recognize these misrepresentations and distortions as intentionally produced by white supremacy.

My personal experience as the daughter of Brahmin immigrants from India, who arrived in the USA in 1967 as a direct result of the 1965 immigration act with $8 in their pockets and who ultimately enjoyed considerable financial success, mirrors the collective experience of those classified as Asian American and South Asian American in particular. My siblings and I began school during the era of bussing in a Midwestern

suburb with no other racially otherized families, let alone families of South Asian descent. We were perceived as Black, treated accordingly, and—once the public schools began to diversify—felt a subsequent sense of kinship with students whose ancestors, unlike our own, had survived centuries of enslavement and segregation. When we explained that we were Indian, we were perceived as American Indian and, again, treated accordingly. By adolescence, we began attending an elite, single-sex, college preparatory school, where we were no longer subjected to overtly discriminatory interpersonal dynamics, but gained insight into racialized institutional dynamics as well as the intersections of race, gender, and class. We had already been primed for the latter through our understanding that the wealthiest Indians in India—who do wield economic power—are of "middle" (merchant) caste status and wield less social and cultural power than the "highest" (priestly) caste; in contrast, Brahmins historically lived on alms.

In addition to having been perceived as African American and understood as indigenous, members of my family of origin (whose eye shape, hair texture, skin color, and other phenotypic features vary considerably) have been read—and treated—as Muslim, Latinx, Greek, Italian, and generically as White. Our racially ambiguous status and periodic invisibility, combined with markers of class privilege and racialized, gendered stereotypes of conformity, have allowed me to witness how the construction of racialized difference by highly educated professionals classified as White influenced their decision making in government and institutional situations. I continue cultivating awareness of the propensity under conditions of white supremacy in the USA for my positionality—whether as a Brahmin during India's colonization by the British or as a professional who is read in contemporary Minnesota as an Asian American woman—to be used to maintain current structural dynamics of super-ordination and sub-ordination.

At the same time, my Brahmin status of unearned privilege and sense of entitlement provides me with insight into whiteness. Each day brings me opportunities to reflect on how the intersections of my positionality shape my credibility with various social groups. This includes the evaluation scholars and practitioners—both racially

otherized and racially normatized as White—whom I interviewed as part of this dissertation research.

Conclusion of Chapter Three

This chapter provided an overview of the dissertation's philosophical and practical positions, which are rooted in a framework that links theoretical concepts from critical theories of systemic oppression with analytical concepts from systems thinking. It described its multi-strand design, using CDA and CST, as stemming from an intersectional paradigm and transdisciplinarity. It also detailed the approach to textual and interview data collection, analysis, and interpretation. Finally, it concluded with the study's delimitations and limitations, ethical considerations, and the researcher's role and positionality.

CHAPTER 4: RESULTS OF TEXTUAL ANALYSIS

Chapters Four, Five, Six, and Seven describe the findings from each method of each strand of the study by research question. Within each chapter, findings are organized by strand and the sequence of the design, as summarized in Figure 7.

Chapters by Research Question	Strand	Method
1. Chapter Four How has the U.S. scholarly evaluation literature constructed racialized difference?	Critical Discourse Analysis	Textual analysis of U.S. evaluation literature
2. Chapter Five How has that construction changed since the field began formalizing in the early 1970s?		Diachronic analysis of U.S. evaluation literature
3. a. Chapter Six How does the trajectory of language relate to the systems surrounding evaluation?		Archival analysis of AEA documents
3. b. Chapter Seven How does the trajectory of language relate to the systems surrounding evaluation?	Critical Systems Thinking	Grounded theory-inspired analysis of interviews

Figure 7. Organization of findings by method, strand, and research question

How Has the U.S. Scholarly Evaluation Literature Constructed Racialized Difference?

Using the discourse-historical approach (Wodak, 2001a), the CDA drew from the scholarship of van Dijk (2015) to examine the way social power, relations of super-ordination and sub-ordination, and inequality are enacted, reproduced, and resisted by text and talk in social and political contexts (O'Keeffe, 2012, p. 448). This examination started with the question of how the U.S. scholarly evaluation literature has constructed racialized difference.

Critical Discourse Analysis (CDA): Textual analysis

The first source of data included all items aside from book reviews in peer-reviewed U.S. evaluation journals since the date of their inception. This data source corresponds with the internalized level of the social-ecological model

Data collection. Because findings from the textual analysis were intended to inform the subsequent methods, the search only included items published through 2017. The search for U.S. peer-reviewed evaluation literature that signified racialized difference between the field's inception in 1977 and 2017 yielded 365 articles, regular columns, editorials, keynote addresses, announcements, calls for submissions, and other items aside from book reviews. These items were identified by a search for variations of "race," "white," and "colony" in the five U.S. journals with "evaluation" in their name that are not specific to education, assessment, public health, or substance abuse. Those that contained some discussion of a relevant variation of one or more of the search terms in their full text (as opposed to merely in their references or in a list of variables) were considered for textual analysis. Table 2 summarizes the search results by search term.

Table 2. *Count of items published between 1977 and 2017 in U.S. peer-reviewed evaluation journals containing some variation of "colonization," "race," or "white"*

Search term	Count
Coloni*	69
Rac*	252
White*	159
Total	365

Note. Some items contained more than one search term.

Data processing. The 365 items identified were manually scanned and narrowed to 46 that directly addressed, as opposed to just mentioning, racialized difference. These 46 fell into four broad, sometimes overlapping categories in their overall treatment of racialized difference: (1) diversity, (2) culture, (3) inclusion, and (4) critical. "Critical" was defined broadly. Literature was categorized as critical if it simply accounted for one or more of the following: (1) the produced nature of racialized difference; (2) the relational nature of racialized difference—among groups or through macro-level social structures; (3) the asymmetry of those relations or stratification of those structures; and (4) history or the effects of time with respect to racialized difference. The critical category formed a less coherent body than the other three in that the items within it did not necessarily cite each other, nor had they necessarily subsequently been cited within

evaluation. In fact, some were print versions of guests' keynote addresses at AEA annual conferences. The 46 items were categorized by their industry specificity: education/assessment, health, international development, U.S. nonprofit, or evaluation more generally. Only those from the critical category that were specific to the U.S. nonprofit industry or that pertained to evaluation generally were considered for the textual analysis. Table 3 lists these by broad category and industry.

Table 3. *Count of U.S. peer-reviewed evaluation items discussing racialized difference by industry*

Broad category	Education	Health	International	U.S. General/ Nonprofit
Diversity	1	0	1	6
Culture	5	2	13	14
Inclusion	5	6	4	13
Critical	8	0	1	20
Total				46

Note. Some items are listed in more than one category.

Sample. 20 of these 46 items were selected for inclusion in the textual analysis portion of this dissertation's CDA strand of inquiry because their signification of racialized difference represented a change in language or introduced a way of thinking about racialized difference that had not previously appeared in U.S. peer-reviewed evaluation journals. These shifts in the construction of racialized difference are summarized in Table 4 and detailed in the subsequent paragraphs, with phrases identified through the textual analysis as representing shifts in the signification of racialized difference within the peer-reviewed evaluation literature bolded.

Table 4. *Summary of discursive shifts in U.S. peer-reviewed evaluation literature between 1977 and 2017, contributing to evaluation's construction of racialized difference*

Author	Year	Title	Discursive strategies and shifts
Campbell, D.	1972/ 1991	Methods for the experimenting society	1. Racialized evaluators as **White** 2. Contextualized the exploitative dynamics of research within an asymmetrical incentive structure
Besag, F. P.	1981	Social Darwinism, race, and research	3. Exposed **race** as scientifically produced and named the asymmetrical incentive structures mediating its reproduction
Merryfield, M. M.	1985	The challenge of cross-cultural evaluation: Some views from the field	4. Described **cultural** differences as asymmetrically structured through settler **colonization** and international development
Covert, R. W.	1987	President's corner	5. Described increased **minority** representation in AEA as in issue of fairness in access to professional advancement
Covert, R. W.	1988	Ethics in Evaluation: Beyond the Standards	6. Acknowledged structural asymmetry at multiple levels (state, institutional, interpersonal, and internalized) in discussion of **racism** and bias 7. Racialized conference-goers as **White**
Conner, R. F.	1989	President's corner	8. Used **people of color** and **diversity** for the first time 9. Described representation in AEA as in issue of fairness
Hilliard, A. G.	1989	Kemetic (Egyptian) historical revision: Implications for cross-cultural evaluation and research in education	10. Historicized the deficit-based view of Africans/African American **culture** within the asymmetrical structure of **colonization** and **white** supremacy in the USA 11. Contextualized **race**, **colonization** of **indigenous** Africans, enslavement of African labor, **people of color**, and **culture** within **White**-led structures of power
Richardson, V.	1990	At-risk programs: Evaluation and critical inquiry	12. Historicized the relational and structural production of "at-risk" as a categorical difference 13. Considered the implications of using biological metaphors

Author	Year	Title	Discursive strategies and shifts
Madison, A. M.	1992	Editor's notes	14. Used **inclusion** for the first time in the context of under-represented groups (as opposed to stakeholder inclusion) 15. Described the structural asymmetry of **cultural** dominance 16. Historicized and contextualized the deficit-based view of otherized groups in U.S. government policies and practices 17. Illustrated intersectional analysis by distinguishing racialized difference from differences in socio-economic class status
Davis, J. E.	1992	Reconsidering the use of race as an explanatory variable in program evaluation	18. Detailed the process and implications of the production and reproduction of **race** over time 19. Historicized and contextualized the deficit-based view of otherized groups in U.S. government policies and practices 20. Illustrated intersectional analysis by distinguishing racialized difference from differences in socio-economic class status
Stanfield, J. H.	1999	Slipping through the front door: Relevant social scientific evaluation in the people of color century	21. Historicized and contextualized **race**, **colonization**, **indigenous**, **people of color**, **culture**, **diversity**, **inclusion**, and **minorities** within the structural asymmetry of **white supremacy** at multiple levels
Hood, S.	2000	Commentary on deliberative democratic evaluation	22. Historicized and contextualized U.S. democracy within the enslaved and exploited labor of African Americans
Hood, S.	2001	Nobody knows my name: In praise of African American evaluators who were responsive	23. Historicized and contextualized the invisibility, absence, and need to recruit evaluators representing racially otherized groups within the structural asymmetry of evaluation as a discipline/field
Ward Hood, D. & Cassaro, D. A.	2002	Feminist evaluation and the inclusion of difference	24. Named **white** supremacy for the first time within the evaluation canon (as Hilliard and Stanfield were guests) 25. Challenged essentialism by exploring intersectionality, women **of color**, and Black feminist thought

Author	Year	Title	Discursive strategies and shifts
Hood, S.	2004	A Journey to Understand the Role of Culture in Program Evaluation: Snapshots and Personal Reflections of One African American Evaluator	26. Connected African American and African evaluators, acts of emancipation within evaluation and by abducted Africans 27. Connected **race** and class, contextualizing the exclusion of groups within evaluation and policy decisions within the structural asymmetry of the U.S. educational system 28. Asserted a need, beyond the elimination of bias, to recruit and train evaluators from racially otherized groups
LaFrance, J.	2004	Culturally competent evaluation in Indian Country	29. Described **Indigenous** Evaluation Frameworks as arising from indigenous knowledge systems 30. Detailed culturally competent evaluation in Indian country as honoring and serving sovereignty 31. Historicized and contextualized current programs and evaluation issues in Indian country within settler colonization
Johnston-Goodstar, K.	2012	Decolonizing evaluation: The necessity of evaluation advisory groups in indigenous evaluation	32. Historicized evaluation within the structural asymmetry of colonization 33. Expanded on de**colonizing** evaluation as distinct from but related to evaluation that honors or serves sovereignty
Lee, K. & Gilbert, B.	2014	Embedding the Graduate Education Diversity Internship (GEDI) Program Within a Larger System	34. Contextualized the absence and invisibility of evaluators representing racially otherized groups in the asymmetrical policies and practices of the surrounding industries
Hood, S.	2017	African Americans in the Early History of Evaluation in the United States: Contributions of Ambrose Caliver in the U.S. Office of Education	35. Contributed to the evaluation cannon the name of one more evaluator representing a racially otherized group—specifically African American and operating within legal segregation in the interest of emancipation
House, E. R.	2017	Evaluation and the Framing of Race	36. Historicized and contextualized U.S. wealth in the structural asymmetry of capitalism and enslaved and exploited labor of African Americans 37. Illustrated the role of racial narratives, specifically the **White** racial frame, in fueling the asymmetry

Findings. Analysis of the intra- and inter-textual linguistic characteristics and discursive strategies in the final 20 items suggest that AEA's response to the incongruence between program evaluators and program participants and to racialized difference, specifically, has a history that predates its current formal initiatives focusing on diversity and culture and, in fact, predates formation of the professional association itself. Overall, findings from the textual analysis revealed a sporadic sequence of articles that constructed racialized difference critically, going back to the field's inception, despite a sudden shift toward liberal constructions of racialized difference that followed publications (discussed in the segment below) by Hood (2001) and Ward Hood and Cassaro (2002) and continued through 2016. Only in 2017 did literature critically examining racialized difference resurface.

Naming whiteness and structural incentives for scientific research. Between the early 1970s and early 1980s, four authors directly acknowledged the structural incentives underlying White researchers' approach to racialized difference. The way that each of them—drawing examples from large-scale government programs, educational research, and international development—challenged prevailing notions of racialized difference is described below.

Campbell on social science research. In "Methods for the Experimenting Society," originally written in 1971, but reprinted in its entirety in a 1991 issue of *Evaluation Practice*, Donald Campbell acknowledged in passing that research on the success of anti-poverty programs could be "seen as exploitative, as providing jobs for white [*sic*] middle-class researchers using money that would be better spent on helping poor folks. Such research is also seen as helping the researcher through the articles and theses that get written" (Campbell, 1991, p. 242). This early attention to the incongruence between researchers and program participants remained rare in the evaluation literature in that the author directed readers' gaze at the researchers—whose presumably homogenous whiteness he named and correlated with their socio-economic status and interest in upward mobility. He explicitly racialized the researchers, but not the program

participants, whose racial classification as *not* White he established simply by implying difference between researchers and program participants.

Racializing one group explicitly without racializing another raises questions about what the author assumed or felt confident that readers would infer without being told. The racialization of researchers/evaluators as opposed to program participants contrasts with the bulk of the evaluation literature in the USA since Campbell's piece was originally written in 1971 and even since it was reprinted in 1991. Additionally, Campbell's reference to the publication of articles and theses as helping researchers implied that some structure, presumably academic, surrounds researchers and influences their interests.

Still, Campbell's *appositional reference to researchers' racial classification*—in relation to his *omission of program participants' racial classification*, which apparently went without saying at least in the wake of legally enforced segregation and the Civil Rights movement—suggests that the incongruence between researchers and program participants was unambiguous. Furthermore, Campbell erased African American researchers—who were completing doctorates in educational research, publishing, and evaluating programs at the time that Campbell was (S. Hood, 2001; S. Hood & Hopson, 2008); African American researchers were invisible in Campbell's account. While throughout the article Campbell looked carefully at the structure of government in relation to program experimentation and evaluation, he did not examine the racial incongruence that he briefly mentioned between individual researchers and program participants in relation to the racial stratification of the macro-level socio-economic structures—academic or government—to which he referred.

Besag on the scientific production of racialized difference. In 1981, Frank Besag looked critically at the structural arrangements surrounding educational research and evaluation and the motivating effect such structures might have on the construction of racialized difference. Besag (1981) drew attention to the origins of the now-routine test of statistical significance, which Sir Francis Galton developed to establish the significance of difference in skull size and other presumed indicators of intelligence among groups

differentiated based on selected phenotypic characteristics. This was part of his work founding the International Society of Eugenics, which was dedicated to "the study of the best conditions for the maintenance and development of the [European] race" through selective breeding (Lenz, 1924, p. 223). In the section of the article describing the eugenics movement, Besag explained how early educational researchers sought to explain differences in power and advantage through racialized notions of genetic superiority and inferiority. He drew attention also, though, to the structurally-produced incentive for educational researchers to continue to find statistically significant differences between differently racialized groups—long after the demise of the eugenics movement—to advance themselves professionally. Like Campbell, Besag directed readers' gaze to researchers, acknowledging their professional interests.

Going a step farther than Campbell, Besag connected their professional interests to the structure surrounding educational research and considered the negative effects on those racialized as something other than White. Unlike Campbell, however, Besag did not consider the racial classification of researchers or evaluators themselves. It is unclear whether this omission was deliberate. It is possible that it reflected the author's point that racial classification was produced rather than essential and thus not worth considering. It is also possible that it reflected his focus on structural arrangements that professionally reward all researchers who produce evidence of racial difference, regardless of their individual racial classification or motivation. Finally, it is possible that the omission reflected the author's failure to recognize how the extent to which researchers benefit from such structural incentives varied depending on their racial classification as individuals.

Merryfield on evaluation, colonial, and capitalist interests. Merry Merryfield attended critically to the structure of development aid, naming it as an enterprise and tying it to macro-level systems of oppression—specifically settler colonization and imperialism—in her empirical study of international development evaluators. Her findings included a section devoted to ethical concerns (Merryfield, 1985). Among the examples that she listed were a study in which those affiliated with industries profiting

from an intervention being tested were employed to evaluate the intervention; a funding agency that mandated the evaluation to justify its decision to de-fund the program; disparities in income between program evaluators and participants; and sentiment among evaluators that they must prioritize satisfying the people paying. Merryfield quoted one interview respondent as saying, "These cultures have had thousands of years to get their act together and haven't been so successful of late. Why not instill a profit motive, an idea of society, rather than a family or tribal orientation, and a pride in getting the job done?" (p. 12).

In these examples, the structural relations between the funder and the evaluator—and, indeed, the larger economic relations between corporations and funders based in countries that continue to benefit from colonization, the trans-Atlantic Slave Trade, and imperialism on one hand and program participants in countries whose land and people continue to be extracted and exploited by those dimensions of white supremacy on the other—were paramount. Merryfield acknowledged this. The solutions proposed, however, were methodological and practice-based as opposed to structural. She closed by emphasizing the need for "understanding cultural norms and values" between the "ethnocentric West" (p. 16) and poor countries, each of which she presented as culturally homogenous. This conflation between cultural differences and differences in power resulting from structural asymmetries has, since then, become part of the evaluation discourse. Applied in the above example in international contexts, it is routinely applied to differences between groups within the USA.

Hilliard on the production of African inferiority. Asa Hilliard's 1989 article entitled "Kemetic (Egyptian) Historical Revision: Implications for Cross-Cultural Evaluation and Research in Education" was a transcription of his keynote at the 1988 annual conference under AEA president Michael Quinn Patton. Its focus on culture corresponded with the attention that Patton had drawn to culture in evaluation through the *New Directions for Program Evaluation* edition he edited in 1985. Somewhat like Besag's article nearly a decade prior, though, Hilliard's talk documented "scientific" attempts to justify racial classification and implicated what he called the "use of old boys'

networks for legitimation" (p. 18) among researchers. As such, like Merryfield's study, it was less concerned with culture than with the economic, political, and academic structures surrounding evaluation.

Hilliard framed his talk in the context of politics, which he explained as the manipulation of power. He further explained power as "the ability to define reality and to get others to respond to that definition as if it were their own" (p. 8) and named academics and other members of the research and evaluation community as "deeply involved in defining what others will come to regard as real. As a result, we are involved in the manipulation of power and, therefore, in politics in its truest sense" (p. 8). Perhaps more importantly, however, he named contemporary research and evaluation as having originated in the service of colonization and other forms of systemic oppression:

> [C]ontemporary academic disciplines were honed, and some were actually born, in an era of colonization, slavery, and segregation/apartheid…. It is not just that they were born and honed in the era of colonization, slavery, and segregation/apartheid, it is the case that disciplines cannot escape the political environment and are indeed shaped by the political environment. (p. 8)

Hilliard continued:

> We should also not be surprised to discover that anthropology has been used to assist in the furtherance of the national aims of colonizing nations…. This development has occurred over several centuries and requires an affirmative act or actions, to ensure that the same disciplines are made to serve more equitable ends. (p. 9)

In ways similar to Kimberlé Crenshaw's exhortation cited in Chapter One, Hilliard explained how rejecting the overtly oppressive systems of the past neither undoes the resulting damage nor stops the ongoing reproduction of that damage. Instead, he advocated for researchers and evaluators to deconstruct dominant ideologies by systematically demonstrating how prevailing narratives "serve the interest of upper class, male, white [*sic*] and middle aged social groups while they simultaneously frame, fragment and distort the perceptions and concerns of more subordinated groups" (p. 9). He carefully listed several ways that researchers and evaluators reproduce racialized difference through the exchange of citations and demonstrated how this served their

social group interests and compromises the interests of racially otherized groups, particularly those of African descent.

Noting that multicultural research teams were one way to counteract this pattern of reproduction and that the power to define reality must be a democratic process, Hilliard said,

> …[S]ocial scientists have an obligation not merely to attempt to be equitable from this point forward, but to reflect carefully upon present models, assumptions, and data in order to deconstruct, where necessary, those structures that maintain dominance in a society that has set democracy as a goal. (p. 9)

Considering representation from ethical and methodological perspectives. During the late 1980s, two of AEA's earliest presidents each addressed racialized difference within the Association from the ethical perspective of ensuring that evaluators from under-represented groups gained access to professional advancement opportunities. Their presidential columns and talks are discussed below.

Covert on increasing access among minorities as an issue of justice. In the President's Corner of what was then *Evaluation Practice* (now the *American Journal of Evaluation*), Robert Covert provided the first documentation of ENet and ERS demographics. According to results of a survey of the combined membership, respondents identified themselves as follows: 82% White, 2% other, 2% Black, 2% Asian, 1% Hispanic, and less than 1% Native American; 10% of respondents did not answer the question (p. 96). Covert referred to this "small number of minority members" as "discomforting" (p. 96) and asked whether it represented the number of racially otherized people who practiced or studied evaluation, or whether it instead reflected AEA's inability to "encourage the participation of more minority people" (p. 97). Although Covert stopped short of examining the structure surrounding evaluation, he raised the possibility that individuals representing racially otherized groups were practicing evaluation at the time and furthermore that the structure of the professional association might have been failing to attract and retain them. Importantly, he made this point about racially otherized groups in the context of representation more generally—international members and graduate students having been named in subsequent

paragraphs of the same column. That he did not connect representation to other aims—construct validity and systemic bias, for example—suggests that he considered equal access to professional opportunities for evaluators representing multiple groups a worthy goal unto itself, from an ethical perspective, rather than as a means toward an end.

Covert on racial bias and the need for reflexivity. Covert's attention to racialized difference continued in his presidential address at AEA's 1987 conference, although he shifted listeners'/readers' gazes from evaluation practitioners to the interaction between evaluators of all racial classifications and program participants representing racially otherized groups. His focus also shifted from ethical concerns to practical concerns of data quality. His address represents an early and rare instance in the evaluation literature and professional association's development in which the words "racist" and "White" appear. He focused first on methodological validity:

> We need to examine our instrumentation and determine whether they are racist and/or sexist. A number of evaluation measures that have been used for a long time have been consistently discriminating against different groups of people, particularly women and minorities…. Why has the New York legislature passed some legislation on the issue of racist and sexist testing? …Why isn't the evaluation community setting those standards? (p. 35)

Recognizing measurement as both quantitative and qualitative, Covert went on to raise the concept of reflexivity:

> These measures not only include tests, but also include questions asked in an interview of a member of a minority or a woman…. Have you looked at yourself as a person, and do you know how you feel about people from different races? …Some of the interpretations of historic results were incredibly interesting to people from the black [*sic*] community because the interpretations were from a white [*sic*] perspective. If you're trying to get the perspective of other people, you've got to be sensitive to that perspective. You've got to be aware of it, and you've got to be sensitive to yourself as the instrument. (p. 35)

Finally, he included programmatic concerns:

> We're not just talking about evaluation, but we're also talking about programmatic concerns…. Program administrators give a set of tests, and they can't figure out why young white [*sic*] men are disproportionately represented in the gifted population. (p. 35)

The evaluation literature has since advanced all three strains of thought—validity, reflexivity, and program theory—to varying degrees. Of these, validity has received the greatest emphasis in relation to racialized difference. Additionally, variations of the word "race," like the word "white," nearly disappeared for the following twenty years. Notably, Covert tied his comments regarding validity directly to discrimination and systemic oppression and called attention to the structures immediately surrounding evaluation as well, including the role that the field could play in potentially influencing legislation to mitigate the negative effects of what he described as "racist and sexist" (p. 36) testing that consistently discriminated against women and minorities (p. 35).

Conner on "people of color" and "diversity." Two years later, in the President's Corner column, AEA President Ross Conner (1989) used the words "people of color" and "diverse" for what appears to be the first time in the evaluation literature to raise membership concerns similar to those of Covert, but specific to leadership development:

> Do these opportunities assure that evaluators from diverse backgrounds, experiences, and viewpoints are represented? AEA has had a fairly good history of equal involvement of men and women, but our record on involvement of people of color in leadership positions has not been as strong. How can we as an organization be sure that diversity is assured? (p. 79)

Like Covert, Conner expressed his concern about the representation of racially otherized groups without connecting it to the composition of program participants. Rather, the concern was explicitly values-based. "Equal involvement" was described as either "fairly good" or "not…as strong." The language used by Covert and Conner suggested concern regarding barriers to the professional advancement of evaluators representing racially otherized groups. Conner's language also represented the beginning of some shifts in the evaluation field's construction of racially otherized groups, however—certainly a shift from the language used during Covert's era. One shift was from the term "minorities," which many found diminishing considering that racially otherized groups are the numeric majority globally (and may have been the numeric majority on this continent were it not for its settler colonization by European countries) to the more affirming (but not uncontested) phrase "people of color." Another shift first evidenced during Conner's presidency was toward the discourse of diversity. "Diverse

representation" and "diversity"—qualities Conner sought to "assure"—were both directly connected to "people of color" and thus racialized.

The production of difference, redistributive justice, and white supremacy. In the 1990s, several authors raised explicitly critical questions about evaluation's construction of difference, including racialized difference. Their work, which placed these questions in the context of a history of systemic oppression and the need for redress, is discussed below.

Richardson on the production of "at-risk." As discussed above, Besag exposed the social structure underlying the production of racial classification as well as the academic structure underlying the continuing motivation in educational research to establish hierarchical differences among groups defined by an inconsistent selection of phenotypic characteristics. Nearly a decade later, Virginia Richardson continued that tradition by utilizing critical theory to expose the structure of the K-12 education system that underlay the production of "at-risk" as a category of students. Richardson (1990) began by explaining that the fields of educational research and policy analysis borrowed the phrase "at-risk" from the medical and psychiatric literature to describe individual children who had the potential to fail or drop out of school. According to Richardson, failing and dropping out were treated as akin to "diseases" in education. "At-risk" bore medical or epidemiological connotations—not unlike the physiological or genetic connotations that racial classification bears. "Certain characteristics may relate to the acquisition or manifestation of the diseases, including poverty, single-parent homes, or personal characteristics such as ethnic or linguistic diversity…" (p. 65).

Also like Besag regarding the construction and reinforcement of racialized difference, Richardson revealed how attempting to predictively classify children as "at-risk" might serve the purposes of macro-level structures. "The decision to employ such a model meets the needs of national policymakers, who require an easily identifiable group of students for purposes of categorical funding…." (p. 65). While Besag did not speculate about the motivation of individual researchers and instead emphasized the incentive structure, Richardson acknowledged that individuals using the epidemiologically

predictive model might seek to serve children they believed need assistance, but that the effect of their approach reinforced existing power differentials:

> It may be the case that individuals who advocate programs that serve these hidden functions do so in the belief that the programs will, in fact, care for children in need. The problems may be conceptualized and programs developed to respond to social problems, but such programs may also respond to the socially defined "realities" of schooling, such as the requirements for accountability. These "realities" end up shaping programs so that they function to maintain the status quo. For example, the need for ensuring that federal funds are spent on students who truly require such support leads to the development of easily applied indicators of needy children. This policy labels children as deficient on the basis of characteristics over which they have no control…. The indicators developed to meet requirements for accountability either ensure that low-income, ethnically and linguistically different children are labeled, as they are in various at-risk programs being developed, or function to do so, as in the case of Special Education and Learning Disabilities programs. (pp. 63-64)

Using nested levels that corresponded with the mechanisms-based model of inequality (Reskin, 2003), Richardson went on to raise questions about the location of problems in individuals rather than in the structures surrounding them:

> Unfortunately…, this model assumes that the problem is inherent in the student, and the search for cause is limited to the characteristics of the students themselves. Characteristics of society and schools are left unexamined. This leads to the labeling of students as inherently deficient and to the assumption that the role of the schools as institutions is to fix the student…. We must look to a model where the focus is not just on students as deficient but on the interaction between the student and the schooling context…. The focus of [the Social Constructivist Model] is not on the child alone but on the interaction between the child and the nested contexts of classroom, school, school district, and state. (pp. 65-66)

Similar to Besag, Richardson did not consider the racial classification of researchers in this system. Without referencing racial classification or color, however, she explicitly tied the "at-risk" designation of children to "minority status" (p. 63), "ethnically and linguistically *different* children (p. 64), "ethnically and linguistically *different* students (p. 73), and "ethnic and linguistic *diversity*" (p. 65, all emphases mine). Like "minority," "different" and "diversity" are inherently relational terms. As "minority" was discussed earlier, "diversity" and "different" are discussed below.

Richardson's use of the word "diversity" differed considerably from Conner's first use in the evaluation literature the previous year and foreshadowed the discourse of diversity that continues to prevail in the current evaluation literature. She wrote, "Certain characteristics may relate to the acquisition or manifestation of the diseases, including poverty, single-parent homes, or *personal characteristics such as ethnic or linguistic diversity*…." (p. 65, emphasis added). "Diversity," however, simply means "the condition or quality of being diverse, different, or varied; difference, unlikeness" (*OED Online*, 2015). Rather than a *personal characteristic*, then, diversity—even ethnic or linguistic diversity—is a condition or state that can only exist *among multiple individuals*. Richardson revealed that while "at-risk" appeared in both popular and academic discourse as an essential category, it was a socially produced category. This is not entirely different from the way that racial categories are socially produced, but have ossified as real.

Davis on the implications of using race as a variable. In 1992, James Davis questioned research designs that compared program outcomes by race, which he argued —not unlike Hilliard—often perpetuated simplistic, misleading, and negative conclusions about programs and program participants. Having grounded the article in the context of Great Society programming and the War on Poverty, Davis—like Covert, House, Kirkhart, and Mertens—raised this concern from the perspective of both validity and justice. Because program evaluation was recognized as an authoritative source of knowledge production and the medium through which the experiences of program participants are conveyed to decision makers, validity was pivotal. Moreover, Davis named evaluation as a social practice, within a macro-level social context, that influenced how evaluators understand and portray the social realities of program participants and specific otherized groups. What set Davis's piece apart was the specificity with which he approached validity. Rather than proffering methodological solutions, however, he cautioned against relying on randomization and tests of statistical significance to compensate for poor theoretical conceptualization of problems, programs, and construct validity of variables. He similarly cautioned against focusing on between-group variation

at the expense of within-group heterogeneity, anticipating considerations of intersectionality—a concept that had first been named just three years prior (even if identified and written about a century earlier) (Crenshaw, 1989).

Presciently stating, "Whether race is conceptualized biogenetically or socio-culturally influences the nature of policy and programmatic responses to evaluation results, but that conceptualization is rarely articulated in evaluations" (p. 58), Davis proceeded to explain both conceptualizations of race as dialectical in that they involved Whites defining other groups as "different." Instead, Davis himself explicitly emphasized the importance of evaluators understanding the experiences and social locations of program participants and relevant groups—as opposed to sharing the same racial classification—to articulate a sound grand theory and program theory to guide the development of measures and evaluation design. He listed several ways to overcome many of the difficulties that he identified. These included swapping out variables to see how the interpretation changes, soliciting feedback on designs and interpretations from affected groups, and recruiting evaluators who represent affected communities and demonstrate cultural sensitivity. Like Covert and Conner, Davis used much of the language—for example, "minority," "culture," "diverse"—seen in prevailing narratives of racialized difference. His focus on the historical and structural context and implications underlying and surrounding the treatment of race, however, revealed his critical orientation. Like Covert, House, and others, Davis encouraged the field of evaluation to become "more self-critical and to redefine itself and its ways of knowing those it seeks to serve" (p. 65).

Madison on primary inclusion. The Davis piece above was part of a *New Directions for Program Evaluation* edition on Minority Issues in Evaluation, edited by Anna Madison. Madison mentioned in the Editor's Notes (1992a) and described more fully in the chapter she contributed to the issue (1992b) the need to involve racially otherized groups in ways that extend beyond their participation in the evaluation of programs that have already been designed in response to problems that have already been defined. She advocated for participants' primary inclusion in evaluation, defining primary

inclusion as "direct participation of program participants in all phases of program development, from the construction of problems to the evaluation and the interpretation of findings" (Madison, 1992b, p. 35). In addition to articulating the definition of social goals as a distinct step within the program planning and evaluation process—one that was subject to cultural dominance—she stated that involvement of racially otherized groups must include their "assessment of social conditions, of the programs that address social conditions, and of interpretations of social programs' and policies' outcomes" (p. 3).

While Madison supported the recruitment and advancement of evaluators representing groups that were racially otherized—evidenced through her leadership of the MIE TIG that is discussed in the archival analysis—she also foreshadowed intersectional analyses, specifically considering the intersection of racial classification and socio-economic class status. Although intersectionality had yet to become a widely used and misused term, Madison acknowledged that many "middle-class minorities…are too removed from the worldviews and experiences of the program participants…." (Madison, 1992b, pp. 35-36). Instead, she asked why evaluation could not serve as a tool for program participants to advocate for their own best interests. In other words, like Davis, she considered shared racial classification insufficient. What Madison called primary inclusion differed from the current prevailing discourse on inclusion in that it centered the experiences and perspectives of those most negatively affected by asymmetrical structures and recognized their potential as agents of change.

In addition to introducing this use of the term "inclusion" to the evaluation discourse, Madison used many of the same terms that are used in currently prevailing discourses as well as others, including "minority" and "non-white," that have fallen out of favor among racially otherized groups in the 25 years since her edited issue. Still, her historicization and contextualization of current conditions, complication of race as an essentialized category, and description of the scope of program participants' involvement in the evaluation process distinguish her language from that of liberal constructions of difference. In referring to "patterns of cultural dominance and sub-ordination in this country" that are not unlike "[t]he dynamics of dominant colonial culture-third-world

culture interactions," she betrayed her theoretical orientation (Madison, 1992b, p. 36). Moreover, her framing of the issue consolidated a small body within the evaluation literature that might now be labeled critical because she problematized the asymmetry of structures within and surrounding evaluation and contemplated the consequences for racially otherized groups.

Stanfield on white supremacy. John Stanfield grounded his AEA keynote, reprinted in the *American Journal of Evaluation*, in "the vices of colonialism and blatantly sexist and racist policies" (1999, p. 416). He voiced a desire to "escap[e] nineteenth-century notions of America and of people of color as 'minorities'" (p. 429). He critiqued the de-politicized notion of diversity that was becoming increasingly popular within evaluation by raising the issue of power:

> We rarely are encouraged to pause and consider that diversifying curriculum and adding color here and there on college campuses and other institutions does not transform segregated institutions into empowered, desegregated, integrated entities unless the diversity, unless the addition of pepper in the salt, is done in the seats of power and privilege of the institution. (p. 427)

He similarly complicated the notion of culture—also increasingly popular within evaluation:

> Post-modern efforts to promote feminist and cultural approaches to research in the sciences and humanities have not…translated into much transformation of the gendered and the racialist nature of who does science or humanities; who controls the grants and contracts; who does the grunt work; and who are the "subjects." Take in particular the case of research on Black people and other people of color and its lucrative funding sources. In this arena, the contradictions between the claims of multiculturalism in the academy and in grant applications, on the one hand, and who really controls the money and the data, on the other, are increasingly paradoxical. (p. 417)

Third, Stanfield added nuance to the idea that stakeholder inclusion, as prescribed by various participatory evaluation approaches such as deliberative democratic evaluation, necessarily addresses racialized power dynamics. He illustrated the "racialized authority relations that dictate who has control of the dollars, the data, and the publishing agenda" (p. 17) that he had observed in community outreach efforts:

> [M]ost of the principal investigators in the community outreach partnership…movement … are White (and usually male). These projects commonly have a person of color, usually Black, functioning on the project as the community liaison. Moreover…, even though community members are supposed to be consulted in all aspects…from grant writing to implementation, they are rarely given real power in project decisions. And they are even more rarely provided significant decision-making roles in the culture of sponsoring universities and colleges. (pp. 417-418)

Stanfield's greater contribution to thinking within evaluation about racialized difference, however, involved his detailed examination of the construct of race and structural inequity. In ways reminiscent of Besag nearly two decades earlier, Stanfield exposed the "fallacious thinking" (p. 423) and "pseudo scientific claims" (p. 423) surrounding racialized difference, which he described as "a mythological construct" (p. 423) rooted in "the false notion that social and cultural characteristics…can be attributed to real or imagined phenotypical traits" (p. 420). He described it as having become accepted, however, "as a fact of life…in academic scientific and humanities communities, in the U.S. Census and other government data-gathering agencies, in popular culture and the mass media, and in private survey and polling companies" (p. 422).

Importantly, Stanfield did attribute characteristics to specific racially otherized groups—for example, he discussed the implications for evaluators of the African American oral tradition. He attributed such characteristics, however, not to race, but rather to "culturally and socially different worlds and realities" (p. 415) and emphasized those "created and transformed by people of color" (p. 415). In doing so, Stanfield understood and portrayed racially otherized groups as actors responding to their environments. Also like Besag, Stanfield connected the construction of racial difference to a macro-level socio-economic structure of white supremacy. Tracing the historical origins and evolution of evaluative traditions in the social sciences back to W.E.B. DuBois and referring to critical race theorist Derek Bell, Stanfield described white supremacy as ordinary in and foundational to U.S. society:

> White supremacy has been a normative form of racial dominance in the United States since its colonial origins. Given its diffusion throughout American society,

> White [*sic*] supremacy touches and shapes every one, every institution, every community, and every ecosystem. No one, collectively or individually, escapes from the dynamics of socialization and stratification or from the impacts of White supremacy…. In this sense, White [*sic*] supremacy consists of more than negative attitudes and behaviors towards racialized out-groups. It involves the very way this society is organized, and significantly influences our views and treatments of self and others. (p. 420)

The mythological construction of racial difference, Stanfield stated, had historically been useful for justifying the institutionalization of economic domination and economic exploitation (p. 422) that is white supremacy. It continued to be "profitable to those who build careers as race experts in the academy and beyond" (p. 423).

In ways similar to Madison's (1992b) emphasis on the implications of who defines social problems and policy goals, Stanfield called attention to the extent to which Eurocentric experiences were considered normative while the experiences of racially otherized individuals and groups were interpreted as deviant or falling short in relation to those racialized as White:

> The media and politicians have conveniently placed the problematics of welfare on the shoulders of undereducated and highly unemployable beneficiaries, rather than on the backs of those sectors which refuse to redistribute resources to the poor and needy, such as medical and agribusiness corporations, corporate slum landlords, and even local colleges and universities…. (p. 427)

> While American social scientists, specifically those who claim expertise on people of color, and their academic institutions and research agencies continue to ground their analyses in assimilationist and reformist perspectives, there is a deepening gap between what we claim and the nature of the non-White world over which we claim intellectual authority. (p. 428)

Stanfield was explicit in distinguishing between what he called racialism—"complex processes of socialization and societal organization" (p. 420) that affect everyone, regardless of racial classification, socialized in a white supremacist society—and racism. The U.S. focus on racism—overt or covert negative attitudes and behaviors against racially otherized individuals and groups—implied that it is possible for individuals to somehow escape the effects of white supremacy. Becoming anti-racialist according to Stanfield, however, involved "constructing a new language…. [B]efore the

language changes, the power and the privilege dynamics, and *the structures that the language produces and institutionalizes*, must change as well…. (p. 422, emphasis added).

Stanfield's prescription for becoming anti-racialist differed from that of liberals and conservatives alike, whom he described as promoting education and exposure to the positive histories and social qualities of those of other so-called races. Such measures corresponded with an aversion to structural explanations regarding the social and economic mobility of racially otherized groups. The preference, he observed, was for psychological explanations in individualized terms, rather than "explanations in terms of structural frameworks of the authority relations of White [*sic*] supremacy" (p. 427).

From critical theories of systemic oppression to diversity and cultural competence. In the early 2000s, four pieces of literature explicitly cited critical theories of systemic oppression. These articles' exploration of critical race theory, feminism, and indigenous studies is discussed below.

Hood on the injustice of U.S. democracy and need for representation from the perspective of validity. In 2000, Stafford Hood echoed Stanfield's questioning of the USA as a democracy, when he boldly named the extent to which "democracy in America has been experientially different for certain groups" (p. 77) in response to the characterization by Ernest House and Kenneth Howe (2000) of deliberative democratic evaluation as an approach that considers the sociopolitical structures surrounding evaluation and that can attend to power imbalances. Among fewer than five peer-reviewed items to reference critical race theory directly, Hood's commentary was the first of a handful of examples within the evaluation literature in which racial inequities were described as foundational to U.S. development. Specifically, he complicated prevailing portrayals of democracy as necessarily fair, just, and protective of minority interests by noting democracy's relationship in the USA to the enslavement of Africans:

> [M]y view on democracy is "tinted" by my racial and cultural heritage. I offer this not as an apology for the views I express in this commentary on House and Howe's deliberative democratic evaluation approach. Rather, I provide it as a "reader's guide." For some it may be more comfortable to frame my arguments within the construct of critical race theory whereby the law and legal traditions

> are analyzed "through the historical and contemporary perspectives of racial minorities in this country" (Delgado, 1995). There is a similarity between my arguments and the legal arguments posed by critical race theorists…. Our concerns are grounded in American history…. In my view, the experience and analysis of African Americans in regard to American democracy is profoundly linked with American slavery. (p. 79)

Hood went on at length to quote Frederick Douglass' exposure of the hypocrisy underlying Fourth of July celebrations, affirming the piece's enduring relevance to those in the United States who continue to be disenfranchised, and continued by quoting W.E.B. DuBois, who is considered among the forerunners of critical race theory. He acknowledged House's previous work on race and policy and commended his writing in educational research and policy arenas about racial formation and structural racism in the United States, but added that "the issue of race must be placed firmly into this discussion" (p. 80). While supporting deliberative democratic evaluation's prescription to include all relevant interests, address power imbalances, and protect the interests of less powerful stakeholders, he lamented the claim that attention to social justice and inclusion alone could reduce bias in evaluation. "I find it impossible to discuss American democracy without considering race" (p. 82).

Hood's call for increasing the number of evaluators from racially otherized groups differs from Madison's call for primary inclusion of program participants earlier in the evaluation process in an intentional attempt to conceptualize problems and establish program outcomes in ways that challenged White cultural dominance nearly a decade prior.[49] Hood called for achieving a critical mass of program evaluators representing racially otherized groups to incorporate "the experiences of people of color into the social and institutional structures" (p. 80). However, unlike Covert and Conner, who similarly advanced the goal of increasing the number of evaluators representing racially otherized groups in AEA, particularly in leadership positions, Hood advocated for increasing the numbers of evaluators representing racially otherized groups *as a*

---

[49] This is not to suggest that Madison did not focus attention on recruiting and mentoring evaluators representing racially otherized groups. She had in fact led that effort decades prior, as discussed in archival analysis.

means to an end rather than as an end in itself. Hood linked the need for more evaluators representing racially otherized groups to presumably more valid evaluation results, marking an important turn in the evaluation literature: Evaluators of color, he stated, would be able to draw from shared life experiences or cultural backgrounds and interpret nonverbal cues from program participants. To make this case, Hood marked racially otherized groups as having culture—culture manifest in nonverbal communication that evaluators from presumably similarly racially otherized groups could recognize and interpret.

> In the case of *some* racial and cultural groups, it is not always what they say but how they say it, as well as the observed nonverbal behaviors. *An evaluator of color* could play an important role as an "interpreter" in the design stages, during the dialogue process, during the implementation of the evaluation, and in the interpretation of evaluative findings. (Hood, 2000, p. 82, emphasis added)

One year later, Hood (2001) called attention to the number of academically trained and published scholars racialized as African American conducting educational evaluation between the 1940s and 1960s, during which time Ralph Tyler was promoting evaluation to assess the extent to which program objectives had been fulfilled. Unlike Tyler, whose racial classification was assumed through its omission to be White, however, these evaluators had been entirely absent from the evaluation canon (most concretely represented in Marv Alkin's 2004 publication of *Evaluation Roots: Tracing Theorists' Views and Influences*) until 2001 when Hood, with Rodney Hopson, embarked on a project entitled "Nobody Knows My Name." This quest documents the otherwise invisible history of scholars racialized as African American who had been conducting and publishing about evaluation for more than 80 years.

As in the 2000 commentary in which he situated deliberative democratic evaluation within the macro-level structure of U.S. democracy, exposing African Americans' experience therein, Hood situated educational evaluation within macro-level U.S. social and economic structures by noting the Tyler period as significant from the perspective of increased access to education among segments of the U.S. population racialized as African American through several policy-level changes. These included desegregation of the armed services, the GI Bill, and the 1954 *Brown v. Board of*

Education of Topeka decision. Furthermore, he referred to the field of evaluation (beyond its professional association) as a body that had chosen to center the voices of scholars who were classified as White men. Finally, he continued to emphasize the *practical* need to increase the number of evaluators representing racially otherized groups. Again, while Hood acknowledged the moral dimension of this effort, his emphasis on the practical need was a departure from writing by Covert and Conner about the justice-oriented need to recruit more evaluators representing racially otherized groups into AEA. It was also a precursor to the discourses of culture and diversity that soon became more fully entrenched within the evaluation literature in the 2000s.

> [I]n the early years of educational evaluation, there were trained African American scholars who conducted educational evaluations and published in scholarly journals—yet went unnoticed. I further argue that *the increased participation of African Americans and other evaluators of color is a pragmatic necessity for evaluation*, as well as a sensible and morally decent thing to do. The reader is advised that my view enlarges the present boundaries of responsive evaluation. For this, I ask your thoughtful consideration—but offer no apology. As an African American man I neither seek nor expect affirmation in a field that has only recently included white (sic) women in its pantheon. (pp. 32-33)

Hood's assertion of what he referred to earlier as his "racial and cultural heritage" (2000, p. 79) and later as his "lived experience within [the African American] community" (2001, p. 32) as strengthening rather than compromising his insights regarding race (in both pieces, Hood explicitly refused to apologize for his perspective) represented one of the earliest examples within the peer-reviewed evaluation literature in which the author wrote in the first-person as a racially otherized human being about race. In both pieces, race was presented as fundamental to both the USA and to the field of evaluation.

> Race is the foremost social issue of our generation and has been since the inception of our nation. For the matter at hand, I aver that race influences who is awarded evaluation contracts, who is awarded professorial positions, and who is listened to by evaluation clients. (p. 35)

Hood's 2001 article was among the earliest to articulate the theory and practice of what would later be called culturally responsive evaluation, which is distinct from both cross-cultural and culturally competent evaluation through its focus on action in the form

of responsiveness. He devoted the bulk of the article to this legacy that evaluators racialized as African American offer the field in general and offer the evaluation of educational programs intended to benefit racially otherized individuals and groups in particular. Hood divided this legacy into the following three related areas: (1) their use of qualitative methods prior to their wide acceptance within the field, (2) their shared lived experience, and (3) their focus on issues, including separate and unequal education. Unlike in his 2000 commentary, however, Hood for the most part described the evaluation practice of evaluators racialized as African American as arising out of the context of structural racialization—specifically, a shared understanding among the program evaluators, program participants, and presumably teachers, administrators, and other school staff—of segregated education as well as a shared investment in the education of children racialized as African American.

Importantly, Hood's detailed description of the culturally responsive theory and practice of educational evaluators racialized as African American was largely to serve his greater point about the field's omission of their contributions to evaluation, despite their publication in scholarly journals like the *Journal of Negro Education*, whose White readership Hood surmised was likely small, based on the extent to which it was cited.

> There are no simple answers to questions about why the work of these "leaders" has gone unnoticed, or similarly, why people of color do not sit in the seats of power in educational evaluation today. (p. 40)

Hood made even this larger point about the field, however, to advocate ultimately for a concerted effort to increase the number of evaluators representing racially otherized groups, which he argued would increase not only the accuracy of results that he discussed in his 2000 piece, but also the efficiency of the evaluation process.

> [B]ut it will not do merely to add the names of African American educational evaluators to a reference list here, a bibliography there. We must additionally and aggressively cultivate the rich potential of young scholars of color that lies fallow in the field of educational evaluation as well as others who await admission to our programs of higher education…. Evaluators of color, when evaluating educational programs in settings where our particular cultural groups are intended to be the primary beneficiaries, bring different experiences to the evaluation than do our white [*sic*] evaluator counterparts. Establishing the necessary rapport with principals, teachers, and students in these settings is likely to take us less time

> because of the probability that a "shared lived experience" exists. Less time will be required in translating the cultural nuances and nonverbal communications associated with observations, conducting interviews, and interpretations that go beyond quantitative indicators of what the program appears to be. (2001, p. 41)

Hood's arguments supporting efforts to increase the number of evaluators representing racially otherized groups did not stop at accuracy and efficiency. He closed by emphasizing the potential for such an increase to influence the field's approach to racism.

> Well beyond efficiency and effectiveness, the evaluation community has a moral obligation to embrace many more evaluators of color. Early African American evaluators exemplified moral leadership and courage when they used their work to engage the profound political chasms of racism in the mid-twentieth-century United States. Our future as an evaluation community and the importance of our work for our stakeholders depends on our ability to embody similar courage and commitment. (2001, p. 41)

As such, Hood linked an increase in evaluators representing racially otherized groups not only to culture, but also to justice. Based on his research about early educational evaluators racialized as African American, he hypothesized that the practice of responsive evaluation conducted by evaluators representing racially otherized groups could contribute to the understanding and reduction of racial disparities. Responsive evaluation was described as that in which evaluators responded to the situations of program participants representing racially otherized groups by drawing upon their shared lived experience, the larger issues framing the program and its evaluation, as well as qualitative methods that could illuminate the results experienced by program participants in ways that quantitative methods alone could not. While Hood's writing in 2000 and 2001 was largely devoid of references to diversity, it was replete with references to culture, particularly in its conceptualization of shared lived experience.

Ward Hood and Cassaro on white supremacy and intersectionality. In their article for a 2002 *New Directions for Evaluation* edition devoted to Feminist Evaluation (Seigart & Brisolara, 2002), Denice Ward Hood and Denice Cassaro used the phrases "white supremacy" and "intersectionality" for the first time from within the evaluation canon. (White supremacy had appeared before, but previously it was by guest speakers at the

annual conference, whose speeches were later printed in the *American Journal of Evaluation*.) In ways similar to Hood, this piece—like many within the feminist tradition—was simultaneously personal and political in that the authors wrote in the first person and named their multiple social locations—including the systems by which they were super-ordinated and those by which they were sub-ordinated—but did not limit their analysis to their personal experiences of oppression at the internalized or interpersonal level. Rather, they explicitly connected personal, political, and professional work and encouraged readers to do the same.

> In formulating our thoughts on the importance of incorporating feminist theories that explicitly address issues of difference into the evaluation enterprise, we inevitably found ourselves reflecting on a range of emotions related to our experiences as members of both oppressed and oppressing (dominant) groups. We share some experiences and perspectives as women and as people who have moved from one socioeconomic class to another. Our other identities, however, in combination with the two already mentioned, move us both into very different locations with our experiences and understandings, based on our access to power and privileges through our affiliation with dominant group identities. In our case, one of us as a white woman has access and enjoys privileges that her African American colleague does not have, based on the systemic racism (*more correctly*, white supremacy) that exists in our society. An additional shift occurs between us as we consider another identity. One of us as an African American heterosexual woman enjoys privileges that her white lesbian colleague does not have, based on the systemic heterosexism (another form of supremacy embedded in male supremacy, or *what is more gently referred to* as sexism) that exists in our society. (Ward Hood & Cassaro, 2002, p. 27, emphasis added)

The Ward Hood and Cassaro piece was also rare in the way that it called attention to language in the context of systemic oppression, as illustrated by the italicized phrases above and the quotation below:

> No discourse belongs solely to its author. Readers, too, participate in the creation of meaning. A definition sets up a dialogue, and it does that best when the author gives a clear and unambiguous account of what feminism is in the context of her own discourse…. Definitions are tentative, open to challenge, must be argued for and substantiated, and can always be modified. A definition is not the essential and only true meaning. (p. 28, citing Thompson, 2001, p. 6).

In ways very similar to Hilliard, Ward Hood and Cassaro asserted the power that evaluation holds in terms of controlling the means of knowledge production—specifically, "the *control* and manipulation of meaning" (p. 30, citing Thompson, 2001, p. 31, emphasis added). This control was not separate from—indeed, it is integral to—the control of nature and people.

After illustrating the production of difference—in terms of sex, gender, sexuality, race, and socio-economic class status, as well as more generally acknowledging difference as inherently relational—Ward Hood and Cassaro discussed feminist evaluation. Perhaps their greatest contribution to the evaluation literature was their thorough treatment of intersectionality, the academic and activist work of women of color, and Black feminist thought—including Black feminist epistemology.

Hood on the tradition of responsive evaluation among scholars classified as African American in relation to cultural competence. In the 2004 *New Directions for Evaluation* edition devoted to Cultural Competence in Evaluation (Thompson-Robinson, Hopson, & SenGupta, 2004), Hood advanced understanding of responsiveness within the context of educational evaluation and educational evaluation within the context of education in the USA more broadly. The responsive approach of the evaluators whose contributions Hood had sought to document grew out of the interest in emancipation that they shared with the participants—generally also classified as African American—of the programs that they tended to evaluate in the era of legal segregation. As in Hood's previous work, shared "culture" and "lived experience" were often used to convey this shared *interest*, which arose from a shared *experience* of *oppression*. Also similar to his previous work, Hood proffered this shared culture and lived experience as the basis of his argument—rooted in social justice and service to benefit "the less powerful" (p. 22)—for the recruitment and training of more evaluators representing racially otherized groups in the current era, in which urban schools were more racially and economically segregated than before the *Brown v. Board of Education* decision.

What was somewhat, but not entirely, new about this piece by Hood was the diasporic connection he drew between the struggle of evaluators representing racially

otherized groups within the USA and that of the African Evaluation Association as well as the historic connection he drew between the former and the struggle waged by Africans abducted and aboard the Amistad slave ship. In the film, in the book, and in history, the fate of the Africans engaged in the resistance effort lay in the hands of those who benefited from their capture and enslavement, maneuvering within a system of justice that was logistically and philosophically foreign to them. The African defendants' victory, facilitated by American settlers classified as White, left the multiple structures underlying their capture fundamentally intact.

LaFrance on changing knowledge creation, components, and ramifications. In the same issue as Hood's piece above, Joan LaFrance discussed culturally competent evaluation in Indian Country and indigenous communities by flipping the script (2004). While her piece was not necessarily the first time that the term "indigenous" had appeared in the evaluation literature, it was the first time that indigeneity, de/colonization, and sovereignty were unpacked at length, in the first person. LaFrance described Indian Country and indigenous communities as those whose lands had been taken from them by outsiders.

LaFrance flipped the script by foregrounding Indigenous Evaluation Frameworks and discussing considerations for culturally competent evaluation in Indian Country only after asserting the value of, and grounding readers in, indigenous knowledge systems and epistemologies. She quoted Deloria (1999) as saying, "I believe firmly that tribal ways represent a complete and logical alternative to Western science" (LaFrance, 2004, p. 41, citing Deloria, 1999, p. 66). While she introduced her piece by noting the presumptuousness of generalizing across indigenous groups and assuming that such a level of awareness could equip any evaluator, including from within Indian country, to practice in Indian country, LaFrance did distill indigenous knowledge systems and epistemology into a few broad brushstrokes that underlay an emerging Indigenous Evaluation Framework. Such a framework values holism more than linearity and hierarchy; it values connectedness—to people and nature—more than individual mobility, and thus tends to be place-based; it values the tradition in which elders pass

along knowledge through relations rooted in respect and reciprocity; and it fundamentally values self-determination.

LaFrance noted that such systems of knowledge could only be considered valid, and thus wield influence from the perspective of sovereignty and self-determination, when those with access to dominant knowledge production systems create safe places for them to thrive. Instead of imposing their training onto indigenous communities, she invited evaluators to take advantage of the applied nature of their job by serving a role in changing the dominant structures and processes of knowledge production:

> [E]valuators can take liberties to explore cultural epistemologies that differ from those taught in the academy if such exploration contributes to the validity and usefulness of evaluation in the context of program operations. Those evaluators who belong to the academy should also be able to bring the fruits of their explorations into the academic discourse. (p. 42)

Expanding the boundaries of time and space. Literature looking critically at racialized difference in the way that Hood (2000, 2001, 2004) and Ward Hood and Cassaro (2002) did at the start of the century was largely absent for nearly a decade. However, four articles in the 2010s did connect the evaluation enterprise to macro-level systems, including the industries surrounding evaluation, settler colonialism, and the capitalist development of the USA, as discussed below.

Johnston-Goodstar on decolonizing evaluation. LaFrance's focus was ensuring that the practice of evaluation honored and protected indigenous sovereignty. In 2012, Katie Johnston-Goodstar extended LaFrance's idea that evaluators could play a role in broadening the ways of knowing taught and sanctioned in the academy to decolonizing the evaluation enterprise itself. Having named research and evaluation as political acts, Johnston-Goodstar further stated that they are "intricately tied" to colonization, much as Hilliard had. Until indigenous researchers and communities have full control over their own evaluation projects, according to Johnston-Goodstar, evaluation advisory groups would be essential. She cautioned evaluators to consider and attend to the ways that "their practices may replicate and/or be seen to replicate these colonial patterns" (Johnston-Goodstar, 2012, p. 10). As such, she not only acknowledged colonization as a

dynamic rather than an event, but she also distinguished motivation or intention from mechanism or impact by differentiating *replicating* from *being seen as replicating*.

Evaluation advisory groups might challenge business as usual, if the communities that they represent grew to realize that business conducted as such counteracted their own values or decolonization and liberation efforts. They might provide a space for community members to share and deliberate on differences in their opinions, priorities, and ultimate aims for knowledge production and dissemination as well as their concerns and ideas about the evaluation process. They might provide opportunities for communities to build their own capacity, including relationships, to conduct or genuinely collaborate on future evaluations. Finally, Johnston-Goodstar argued that evaluation advisory groups might allow community members to wrestle with knowledge paradigms and speak back to the institutions and curricula that train evaluators.

Lee and Gilbert on evaluation's ecosystem. Lee and Gilbert were the first in the U.S. peer-reviewed evaluation literature to acknowledge and in fact problematize the system—and its racially stratified nature—surrounding evaluation from the perspective of building a pool of evaluators representing racially otherized groups and building cultural competence among all evaluators with the ultimate aim of serving the interests of groups that are disproportionately represented in evaluated programs. Broadening the unit of analysis to include organizational policies and practices as well as institutional norms and lengthening the period of analysis to include generations (Lee & Gilbert, 2014, pp. 106-107), they explicitly shifted the focus from the individuals representing racially otherized groups recruited through the GEDI program to the system surrounding them. Lee and Gilbert referred to these organizational policies and practices and institutional norms as having made programs like GEDI necessary "in the first place"—contextualizing and historicizing the relative lack of evaluators representing racially otherized groups as a legacy resulting from institutional and structural decisions rather than naturalizing it as an inevitable result of individual or community deficits. Their use of the word "threatening" suggests that leaders of organizations, agencies, and institutions hiring evaluators and presumably GEDI interns had a stake in maintaining

longstanding policies, practices, and norms. And further, that stake was not compromised simply by hiring evaluators—certainly at the intern level—representing racially otherized groups.

Hood on reclaiming the lineage of evaluators classified as African American for the benefit of future generations of evaluators. In 2017, Hood continued the collective exploration of early evaluators classified as African American by documenting the contributions of Ambrose Caliver in the U.S. Office of Education. Hood and others' work in this area was noteworthy for connecting the insightful and innovative work of evaluation practitioners directly with the surrounding context, including but not limited to overt oppression—and collective action against oppression. In this case, it was delimited by the *Brown v. Board of Education* decision regarding educational segregation. Not only did Hood describe how Caliver's work was influenced by U.S. educational policy, but he also described how Caliver's work influenced U.S. educational policy.

Moreover, Hood continued the pattern described in his previous articles, of reflecting in the first person about the significance of time and space underlying his determination to collectively amend and ultimately re-think the canon of evaluation scholars. While that work had been ongoing since the 2004 discussion with Hopson that Hood described, what was new about Hood's piece on Caliver was that it was printed in the *American Journal of Evaluation*, as opposed to a journal of educational research or a themed edition of *New Directions for Evaluation*. What was also new was the solidarity that Hood expressed in honoring existing members representing multiple otherized groups and calling for additional such members to add to and re-shape the canon. Notably, Hood closed this piece by naming emerging evaluation scholars and practitioners and the CREA community as opposed to that of the American Evaluation Association.

House on racial narratives. In 2017, Ernie House published an article that addressed racialized difference in relation to macro-level U.S. policies and structures and to evaluation in ways that had not appeared in the U.S. evaluation literature since 2002. He had published a similarly thorough piece in the educational research literature nearly

20 years prior (House, 1999b). Having rooted the article in an explanation of Feagin's concept of the white racial frame (2013), House used personal narrative and painstaking detail about the racialized origins of the USA's wealth—to the extent that it was accumulated through enslaved African labor in the capitalist production of commodities—to raise readers' awareness about possibly, if not inevitably, biased definitions of problems, development of measures, and interpretation of results. He focused on the prevailing Black-White binary without incorporating an analysis of U.S. extraction of indigenous land or military aggression in—and anti-immigrant policy toward—Latin America and Asia.

Writing in the first person, House illustrated the white racial frame even as he potentially reinforced it by providing an unbroken, detailed narrative account of an incident—the violent intrusion into their home by a man classified as Black—that his family experienced and that largely reflected and reinforced white supremacist narratives. Rather than disrupting the narrative in progress, House's rhetorical strategy involved providing complicating details and deconstructing the narrative's coherence only afterwards. Because the account remained whole—a self-contained story that was only interrogated afterwards—and precisely because it tapped into the white racial frame of narratives that readers had necessarily internalized, his personal story may be the most memorable and easily repeatable part of the article.

In contrast to House's personal story, his expository account of the U.S. accumulation of wealth—while well-documented and thoroughly-reasoned—was less self-contained and subsequently less easily internalized and repeated, precisely because it countered the white racial frame. Its reliance on voluminous data and logic might make it more convincing than the story to evaluators, whose livelihood and professional identity lie in evidence and logic. However, part of House's point seemed to be that evaluators used story as much as they used "data," not only in the sense of qualitative research, but perhaps more importantly in the sense of making meaning out of data. From the stage of problem definition to the stage of outcomes analysis, evaluators both contribute to and rely on macro-level narratives when they frame problems, develop measures, and

interpret findings. House's presentation of voluminous data and logic challenging the white racial frame seemed incapable of counterbalancing the powerful narrative that he also presents, which reinforces it. As such, his rhetorical strategy seemed miscalculated.

Interpretation and use of findings. The first finding is that evaluation has a long history of literature recognizing and looking critically at racialized difference. Many such items included in the textual analysis for having contributed to evaluation's construction of difference were transcriptions of keynote speeches, often by guest speakers, at AEA's annual conference; many others were published in themed editions of *New Directions for Evaluation* intended to create space for interests that were more typically otherized: Minority Issues, Feminist Evaluation, Cultural Competence, and Indigenous Evaluation.

The second finding is that the above literature does not form a coherent body. Most of the authors included—certainly those addressing evaluation generally or from the nonprofit perspective—had written critically about racialized difference within the U.S. peer-reviewed evaluation literature only once, and much of the literature they wrote was published before the turn of the 21st century. Aside from House and Hood—both of whom come from education contexts—nearly all authors whose work was included in the textual analysis for having contributed to movement in evaluation's construction of racialized difference did not continue to build that body of writing within evaluation.

Authors whose work was included in the textual analysis did not necessarily cite one another, build on each other's work, or comprise a community. House has shepherded a critical strain of thinking about justice, including with respect to racialized difference, in relation to macro-level social, economic, and political systems through evaluation from the 1980s to 2017 and has been frequently cited. Evaluation scholars and practitioners did not expand upon or deepen the topics of justice and race that he raised, however, as of December 2017. Similarly, Hood almost single-handedly centered evaluation's exclusion of scholars and practitioners from racially otherized groups since the late 1990s (2000, 2001, 2004, 2014, 2017). Outside education, only Conner and LaFrance, representing health, international development, and indigenous contexts,

continued to publish articles (as opposed to presidential columns and speeches), to varying degrees, within U.S. peer-reviewed evaluation literature between 2002 and 2017.

After 2003, literature critically addressing racialized difference from the perspective of indigeneity and colonization, often written in the first person, however, began to grow after it first appeared. This is the third finding. LaFrance was part of this tradition—referred to as a tradition because these authors did build on each other's work and cited each other. Their scholarship cohered and they formed a community.

The above findings informed the archival analysis as well as the CST strand's purposive sample of interviewees (see Appendices A and B) and interview questions (see Appendix D). They also elicited nine phrases that were identified *in vivo* as having been used to signify racialized difference in U.S. evaluation literature. The last finding consists of the nine phrases used to signify racialized difference that were generated f rom a close reading of the 20 items. They are listed below in alphabetical order.

- Colonization
- Culture
- Diversity
- Inclusion
- Indigenous
- Minority
- Of color
- Race
- White

Like the three findings described earlier, the phrases above constituted the search terms for the diachronic analysis, guided the search for professional association documents and additional journal content from within evaluation for the archival analysis, and informed the CST strand's purposive sample of interviewees (see Appendices A and B) and interview questions (see Appendix D). Finally, they guided the search for literature outside evaluation to interpret the overall results (Kelle, 2007).

Conclusion of Chapter Four

The textual analysis focused on the construction of racialized difference *within individual* pieces of peer-reviewed literature written by individuals or small groups of individuals. Literature included in the textual analysis was purposively selected from peer-reviewed articles from U.S. evaluation journals that both problematized racialized difference and represented a change in language or introduced language that had not previously appeared in U.S. peer-reviewed evaluation journals. While the search yielded a sample of 20 such items, these items did not constitute a coherent body of empirical or theoretical scholarship on difference, identity, or systemic oppression. The authors did not necessarily cite one another, build on each other's work, or comprise a community. Indeed, many items included in the textual analysis for having contributed to evaluation's construction of difference were transcriptions of speeches by guests invited to AEA's annual conference. Others were published in themed editions of *New Directions for Evaluation*.

Most of the authors addressing evaluation generally or from the nonprofit perspective had written critically about racialized difference within the U.S. peer-reviewed evaluation literature only once, typically before 2000. Only House and Hood—both of whom come out of educational evaluation—continued to write about racialized difference. An important exception to this first finding is the second finding: that evaluation literature critically addressing racialized difference from the perspective of indigeneity and colonization, often written in the first person, actually grew after it first appeared in a themed journal edition in the early 2000s. These authors did build on each other's work and cited each other. Their scholarship cohered and they formed a community.

The last finding consists of the nine phrases used to signify racialized difference that were generated from a close reading of the 20 items: (1) colonization, (2) culture, (3), diversity, (4) inclusion, (5) indigenous, (6) minority, (7) of color, (8) race, and (9) white. These nine phrases—how the U.S. scholarly evaluation literature has constructed racialized difference—and the first two findings informed the diachronic analysis and

archival analysis of the CDA strand, the CST strand's purposive sample of interviewees and interview questions, as well as the search for literature outside evaluation to contextualize findings.

CHAPTER 5: RESULTS OF DIACHRONIC ANALYSIS

How has the Construction of Racialized Difference Changed since Evaluation Began Formalizing as a Field in the Early 1970s?

Critical Discourse Analysis (CDA): Diachronic Analysis

The source of data that was used to measure diachronic change in language consisted of the entire population of material (aside from book reviews) published in peer-reviewed journals in the USA, Canada, and the UK. This data source corresponds with the interpersonal level of the social-ecological model in that it identifies relationships intertextually.

Data collection. The sample for the diachronic analysis included 9,114 items other than book reviews published between 1977 and 2018 in evaluation journals identified through a search in peer-reviewed journals from the UK, Canada, and USA that had "evaluation" in their title and were not specific to public health, education, or international development. Of these, 896 were published in UK journals; 704 were published in Canadian journals; and 7,514 were published in U.S. journals—both affiliated and unaffiliated with AEA. A total of 4,112 items were published in U.S. evaluation journals unaffiliated with AEA: 2,437 in *Evaluation and Program Planning*, 1,372 in *Evaluation Review*, and 303 in the *Journal of Multi-disciplinary Evaluation*. A total of 3,402 items were published in U.S. journals affiliated with AEA: 1,963 in the *American Journal of Evaluation*, formerly *Evaluation Practice*, and 1,439 in *New Directions for Evaluation* (formerly *New Directions for Program Evaluation*). Because the earliest issues of the *American Journal of Evaluation* and *New Directions for Evaluation*, in particular, included items first published in the newsletters of AEA's precursors (Evaluation Research Society and Evaluation Network), the above counts included items beyond research articles, but did not include book reviews. Table 5 lists the journals with their counts of items published by geographic region, country, AEA affiliation, and publication.

Table 5. *Count of items published in evaluation journals by publication, region, country, and AEA affiliation*

	Counts			
	Journal	**Geographic Region**	**Country**	**AEA affiliation**
Europe		896		
UK			896	
Evaluation	896			
North America		8,218		
Canada			704	
Canadian Journal of Program Evaluation	704			
USA			7,514	
Unaffiliated with AEA				4,112
Evaluation & Program Planning	2,437			
Evaluation Review	1,372			
Journal of Multi-disciplinary Evaluation	303			
Affiliated with AEA				3,402
American Journal of Evaluation	1,963			
New Directions for Evaluation	1,439			
Total	9,114	9,114	9,114	7,514

Notes. Includes items in peer-reviewed journals alone
* Excludes book reviews, front matter, back matter

Analysis of the trajectory for each term identified through the textual analysis as potentially signifying racialized difference was conducted by counting the number of items that contained each search term each year and dividing it by a count of total items published that year. Resulting relative annual frequencies were then compared across search terms, over time, and between journals affiliated with AEA and those unaffiliated with AEA. Thus, the number of items published each year was important from the perspective of arriving at the relative annual frequency for each term. Within U.S. journals—both affiliated and unaffiliated with AEA—an average of 179 items (185 median and 189 mode) were published each year, with a low of 32 in 1977 and a high of 250 in 1984. On average, journals unaffiliated with AEA published 98 (both mean and median) items each year, with a low of 32 in 1977 and a high of 162 in 2017. AEA-affiliated journals published an average of 83 items (median 81) each year, ranging from a low of 35 in 1978, 1979, and 1980 (the first three years of their existence) to a high of 141 in 1984. Table 6 lists the overall count, minimum annual count, and maximum annual count of items published in U.S. evaluation journals, along with the year(s) of their minimum and maximum annual counts, by AEA affiliation.

Table 6. *Overall, minimum annual, and maximum annual count of items published in U.S. evaluation journals*

Counts and Years	**Total: AEA-affiliated and unaffiliated**	**AEA-unaffiliated**	**AEA-affiliated**
Overall count of published items	7,514	4,112	3,402
Minimum annual count of published items	32	32	35
Year(s) during which minimum annual count occurred	1977	1977	1978, 1979, 1980
Maximum annual count of published items	250	162	141
Year(s) during which maximum annual count occurred	1984	2017	1984

Data processing and sample. Among the 9,114 UK, Canada, and U.S. peer-reviewed journals, a total of 5,664 journal items published between 1977 and 2018 were identified as containing variations of the nine search terms (colonization, culture, diversity, inclusion, indigenous, minority, of color, race, and white) identified through the textual analysis as potentially signifying racialized difference. Of these items, 697 were in the UK journal *Evaluation*, 73 were in the *Canadian Journal of Program Evaluation*, and 4,894 were from U.S. journals. Of those published in the USA, 1,861 were from journals affiliated with AEA (*American Journal of Evaluation* and *New Directions for Evaluation*). Again, the counts include items beyond research articles, but do not include book reviews. Table 7 lists the journals with their counts of items containing the search terms by publication, AEA affiliation, country, and geographic region.

Table 7. *Items in evaluation journals containing search terms identified through textual analysis*

	Counts			
	Journal	**Geographic Region**	**Country**	**AEA affiliation**
Europe		697		
UK			697	
Evaluation	697			
North America		4,967		
Canada			73	
Canadian Journal of Program Evaluation	73			
USA			4,894	
Unaffiliated with AEA				3,033
Evaluation & Program Planning	1,838			
Evaluation Review	1,105			
Journal of Multi-disciplinary Evaluation	90			
Affiliated with AEA				1,861
American Journal of Evaluation	1,174			
New Directions for Evaluation	687			
Total	5,664	5,664	5,664	4,894

Notes. Excludes book reviews, front matter, back matter

Sampling criteria. Many if not all the search terms identified through the textual analysis as potential signifiers of racialized difference also have definitions or can be used in ways that are entirely unrelated to racialized difference. For example, "race" can refer to a race track. Items that contained variations of the three terms that have widely disparate meanings and usages beyond dimensions of difference—specifically, "colonization," "race," and "white"—were coded for relevance (whether the search term they contained was used in relation to racialized difference) and substantiveness (whether the search term they contained was in the substance of the text as opposed to in the citations or biographies of the authors) as described in the next section on data processing and summarized in Table 8.

While findings from the textual analysis suggest that authors consistently used variations of "culture," "diversity," "inclusion," and "minority" to signify racialized difference, these four terms in particular can potentially signify dimensions of difference beyond racialized difference and are thus not directly comparable with variations of "of color," "race," or "white." Additionally, in preliminary results of the diachronic analysis—before items were coded for relevance or substantiveness—annual frequencies for variations of "culture," in particular, eclipsed the frequencies for variations of all other search terms, including "diversity" and "inclusion." As a result, items that contained variations of "culture" were coded for relevance and substantiveness as summarized in Table 8 and explained in the following section. Only items containing variations of the term "culture" that were coded as related to racialized difference were included in the final diachronic analysis.

Items containing variations of "diversity" and "inclusion," however, were coded neither for the extent to which the search term they contained was related to racialized difference nor for the extent to which the search term was in the body of the text. This decision was made for two reasons: (1) because the number of items identified as containing variations of these terms was too voluminous, and (2) because the textual analysis suggested that variations of "diversity" and "inclusion" had a range of variation in meaning that was somewhat narrower than that of "culture": they had become nearly

synonymous in the U.S. evaluation literature with difference, albeit not necessarily racialized difference.

Findings from the textual analysis also suggest that "minority" had a range of variation in meaning that was somewhat narrower than that of "culture." It, too, had become nearly synonymous in the U.S. evaluation literature with difference (and interview results discussed in the last segment of this chapter suggest it had become synonymous with racialized difference—Blackness, in particular). As a result, items containing variations of "minority" were not coded for relevance to racialized difference or substantiveness. Items containing variations of "indigenous" and "of color" were also not coded for relevance to racialized difference or substantiveness.

Because some interviewees affiliated with AEA (discussed in Chapter Seven) said that they deliberately used "culture," "diversity," and "inclusion" to avoid focusing exclusively on racialized difference, a search for variations of "intersectionality"—potentially more comparable than "race" in that it addressed multiple dimensions of difference—was conducted. To be clear, only one item (the 2002 piece by Ward Hood and Cassaro) in the textual analysis used the term "intersectionality"; the calculation of frequencies for it in regard to the diachronic analysis was motivated by an interest in understanding the relative frequency of "culture," "diversity," and "inclusion." The search for variations of "intersectionality" in U.S. peer-reviewed evaluation literature yielded only 102 items total from the field's inception through 2018—less than that of colonization. This was approximately 1% of the U.S. peer-reviewed evaluation literature as a whole and far less in terms of annual frequencies. Almost all the items containing variations of "intersectionality" were published in journals unaffiliated with AEA, and one-fifth were published in 2018. As a result, the final diachronic analysis did not include annual frequencies for variations of "intersectionality."

Data processing. Items found to contain the terms identified through the textual analysis were coded for relevance and the extent to which use of the search term that they contained was substantive as follows. Those in which use of the search term was relevant, but not in the title, abstract, or body of the item were coded as 1, and those in which use

of the search term was irrelevant (regardless of where in the item) were coded as 2. Items were also coded according to whether their use of the search term represented prevailing, liberal constructions of race as a static, natural category attached to individuals (coded as 3) or as a critical construction (coded as 4). Again, "critical" was defined broadly. Any item that included any acknowledgment of racialized difference as being (re)produced; relational or structural; asymmetrical, hierarchical, or stratified; or tied to history and time was coded as 4. Many items contained multiple uses of the search terms, in multiple ways. In such cases, the entire item was coded according to the most critical use of the search term that it contained, as long as it was in the substance of the item. Table 8 summarizes these coding criteria.

Table 8. *Criteria for coding search results for inclusion in sample for diachronic analysis*

Search term	**Relevant, Not substantive**		**Irrelevant**		**Relevant & Substantive: Liberal construction**		**Relevant & Substantive: Critical construction**	
	Example	**Code**	**Example**	**Code**	**Example**	**Code**	**Example**	**Code**
Coloni*	In a quotation but not elaborated upon	1	Colony of bacteria	2	During the Colonial era	3	Colonial forest policy in India	4
Cultur*	In citations	1	Agriculture	2	Cultural competence	3	Cultural hegemony	4
Rac*	In footnote	1	Space Race	2	Race, class, and gender	3	Racial discrimination	4
White*	In author biography	1	White paper	2	White, Black, Asian	3	Whiteness	4
Divers*	Items containing variations of these terms were not coded for the extent to which their use of the search term was relevant (related to racialized difference) or the extent to which it was substantive (in the body of the text)							
Inclusi*								
Indigen*								
Minorit*								
Of color								

In 47% of the items that contained variations of "culture," the term was unrelated to racialized difference (for example, agriculture or organizational culture) or outside the body of the text (for example, only in the references or the author's biography). In more than 60 percent of those remaining, a variation of "culture" was used uncritically as a dimension of difference associated with, if not a proxy for, race—sometimes as one among many marginalized group identities—or geographic origin. This was consistent across journals regardless of geographic location, country, or affiliation with AEA. These 60 percent were coded as 3, and the remaining 40 percent were coded as 4.

In 23% of the items that contained variations of "colonization," the term was unrelated to racialized difference (for example, colonies of bacteria) or outside the body of the text (for example, in the references or the author's biography). The remaining 98 items were loosely categorized as critical (69 or nearly 70%, coded as 4) or uncritical (29 or nearly 30%, coded as 3). Any item containing variations of "colony" that included acknowledgment of European settler and non-settler colonization of Africa, Asia, or the Americas—regardless of whether it was indifferent or concerned about it—was coded as a 4. Those containing variations of "colony" that referred to a period of U.S. or other countries' history, for example, were coded as 3.

In 24% of the items that contained variations of "race," the term was unrelated to racialized difference (for example, referring to a race track) or outside the body of the text (for example, in the references or the author's biography). In more than three-quarters of those remaining, a variation of "race" was used as a demographic variable—typically in a table or list of dimensions of "diversity"—and thus coded as 3. This was least true of items in the *American Journal of Evaluation*, in nearly half of which a variation of "race" was accompanied by some consideration of complexities regarding racialized difference.

In 45% of the items that contained variations of "white," the term was unrelated to racialized difference (for example, as a proper noun) or outside the body of the text (for example, in the references or the author's biography). In more than 80 percent of those remaining, a variation of "white" was used as a demographic variable—often in a

table or in racially disaggregated reporting of results—and thus coded as 3. This differed widely among items in U.S. journals unaffiliated with AEA, those in U.S. journals affiliated with AEA, and those in journals from the UK and Canada. Items in U.S. journals unaffiliated with AEA were the most likely to contain variations of "white" unaccompanied by any consideration of its produced nature.

Only items coded as both relevant and substantive (3 and 4) were ultimately included in the diachronic analysis. They were distinguished from each other to provide some indication of the level of criticality with which terms identified through the textual analysis as potentially signifying racialized difference were used in the U.S. evaluation literature and the extent to which that might vary by journal and AEA affiliation. Table 9 lists the counts of items containing relevant and substantive uses of each search term by journal. Additionally, because of differences in language across contexts even within North America—for example use of "aboriginal" and "First Nations" in Canada as opposed to "indigenous" in the USA—only items from U.S. journals (both affiliated and unaffiliated with AEA) were ultimately included in the analysis.

Table 9. *Count of items in U.S. evaluation journals containing relevant and substantive uses of search terms*

	Colonization	**Culture**	**Diversity**	**Inclusion**	**Indigenous**	**Minority**	**Of color**	**Race**	**White**
Evaluation (UK)	11	174	429	263	6	71	8	47	14
Canadian Journal of Program Evaluation (Canada)	2	20	47	38	0	12	1	4	1
American Journal of Evaluation (U.S.)	19	326	733	418	87	220	111	188	103
Evaluation and Program Planning (U.S.)	27	559	1,025	742	95	412	186	487	376
Evaluation Review (U.S.)	7	134	435	407	54	285	183	358	312
Journal of Multi-disciplinary Evaluation (U.S.)	5	20	36	26	7	21	22	17	12
New Directions for Evaluation (U.S.)	27	249	195	87	67	134	83	121	60
Total	**98**	**1,482**	**2,900**	**1,981**	**316**	**1,155**	**594**	**1,222**	**878**

Findings. Unlike the results of the textual analysis that were reported with details about—and excerpts illustrating—discursive strategies and linguistic features, results of the diachronic analysis are reported as frequencies alone. The diachronic analysis attended to language use and context only to the extent that it excluded articles in which the search term was outside the body of the text or unrelated to racialized difference (described above, under data processing and sampling criteria).

Analysis of the overall and annual frequencies for variations of each search term derived from the textual analysis (colonization, culture, diversity, inclusion, indigenous, minority, people of color, race, and white) from the field's inception through 2018 echoed the findings from the textual analysis. Multiple strands of thinking regarding racialized difference were found to have existed since evaluation's earliest years. Table 10 lists the overall, minimum annual, and maximum annual frequencies—as well as the years in which they occurred—for variations of each search term identified through the textual analysis as potential signifiers of racialized difference in the U.S. evaluation literature. Listing them as a whole and in terms of affiliation with AEA from 1977 to 2018 illustrates the extent to which AEA affiliation correlated with the trajectory of language signifying racialized difference in U.S. evaluation journals.

The results are also described in narrative form below, again by affiliation with AEA for the same reasons. The trajectories for each potential signifier of racialized difference are first presented and described together. Then, trajectories are presented and described for specific pairs and groups to facilitate comparison among variations of: (1) "minority," "people of color," and "white"; (2) "culture" and "race"; and (3) "indigenous" and "colonization." As noted several times in the textual analysis, the use or lack of use of specific terms does not, in isolation, determine whether any individual item from the literature reflected a critical analysis. Rather, the comparisons simply provide some indication of when alternative constructions of difference fell in and out of favor in an effort to inform the archival analysis and CST interviews. When was racialized difference constructed as natural as opposed to produced? When was it constructed as individual as opposed to relational? When was it constructed as lateral as opposed to asymmetrical? When was the asymmetry historicized and contextualized structurally?

Table 10. *Overall, minimum annual, and maximum annual frequencies for search terms in U.S. evaluation publications by affiliation with AEA*

	Overall Frequency			**Minimum Annual Frequency**			**Year(s)**			**Maximum Annual Frequency**			**Year(s)**		
	Total	non-AEA	AEA	**Total**	non-AEA	AEA	**Total**	non-AEA	AEA	**Total**	non-AEA	AEA	**Total**	non-AEA	AEA
Diversity	**32%**	36%	27%	**9%**	12%	0%	**1980**	1980	1978 1979	**65%**	65%	66%	**2016**	2016	2016
Inclusion	**22%**	29%	15%	**5%**	7%	0%	**1978**	1978	4 years	**51%**	51%	52%	**2016**	2016	2016
Culture	**17%**	17%	17%	**3%**	3%	0%	**1977** **1984**	1977	4 years	**36%**	33%	51%	**2015** **2018**	2010	2017
Race	**16%**	21%	9%	**5%**	8%	0%	**1989**	1989	4 years	**34%**	34%	24%	**1977**	1977 2004	2004
Minority	**14%**	17%	10%	**4%**	5%	0%	**1985**	1985	1978 1979	**24%**	31%	21%	**2001**	2001	2014 2018
White	**11%**	17%	5%	**4%**	6%	0%	**1982**	1980	6 years	**25%**	34%	15%	**1977**	2001	2014
Of color	**8%**	10%	6%	**1%**	2%	0%	**1993**	1993	5 years	**15%**	18%	17%	**2014**	2009	2014
Indigenous	**4%**	4%	5%	**0%**	0%	0%	**1978** **1986**	5 years	10 years	**17%**	10%	27%	**2018**	2018	2018
Colonization	**1%**	1%	1%	**0%**	0%	0%	**Most years**	Most years	21 years	**6%**	3%	12%	**2018**	1977 1998 2010 2015 2018	2018

U.S. evaluation literature: Frequencies for variations of "diversity" and "inclusion" increased until 2016. Again, variations of "diversity" and "inclusion" potentially signify dimensions of difference beyond racialized difference, and items containing those search terms were not coded with regard to their specificity to race. Of the U.S. items that contained variations of "diversity," 53% also contained variations of "culture," and 29% also contained variations of "race." A full 100% of the items that contained variations of "diversity" also contained variations of "inclusion," however—regardless of publications' geographic location or AEA affiliation—although the reverse was not true. Only approximately half (54%) of the items that contained variations of "inclusion" in U.S. literature as a whole also contained variations of "diversity." In items in U.S. publications (from journals both affiliated and unaffiliated with AEA), the overall frequencies for variations of "diversity" (32%) and "inclusion" (22%) increased between 1977 and 2016, at which point they decreased. Annual frequencies for variations of "diversity" ranged from a low of 9% in 1980 to a high of 65% in 2016, and annual frequencies for variations of "inclusion" ranged from a low of 5% in 1978 to a high of 51% in 2016. Figure 8 illustrates this trajectory.

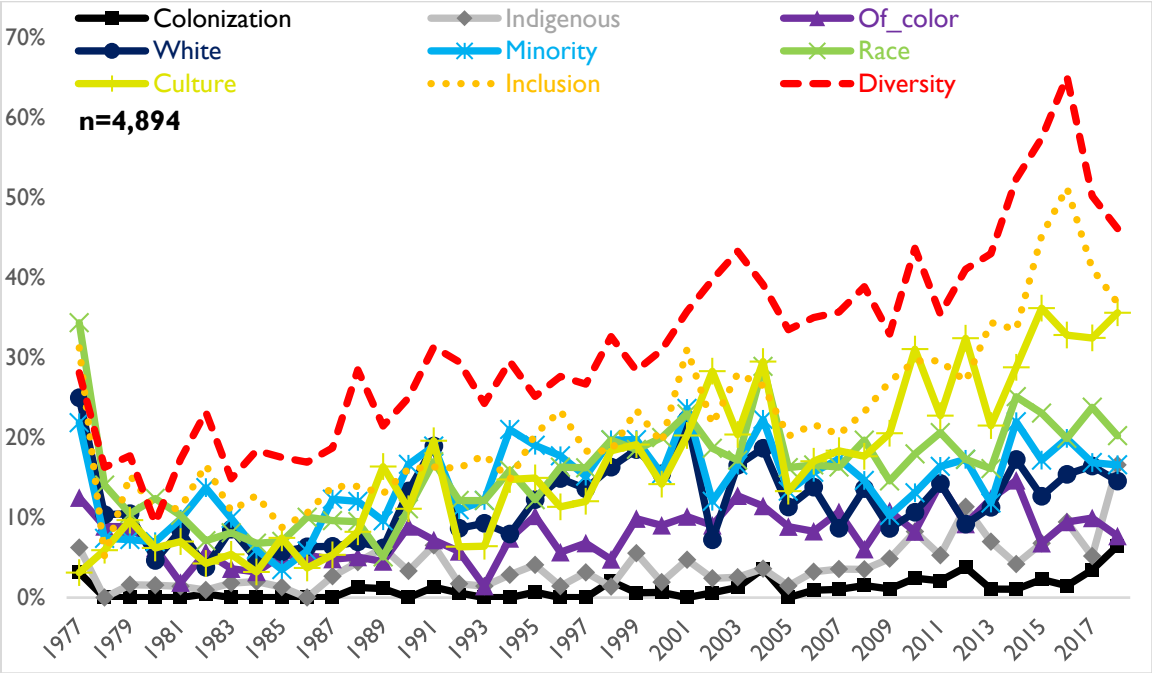

Figure 8. Trajectory in U.S. evaluation publications for all terms identified through textual analysis

U.S. evaluation literature: Variations of "minority" and "white" had a higher overall frequency than that for variations of the phrase "of color." Of U.S. literature that contained variations of "minority," approximately 43% also contained variations of "white" and the phrase "of color" each. Overall frequencies in U.S. evaluation literature for variations of "minority" and "white" (14% and 11%, respectively) were higher than those for variations of the phrase "of color" (8%), although annual frequencies for all three terms increased at the same rate. It is important to note—because items containing variations of "minority" and the phrase "of color" were not coded for relevance and substantiveness, whereas items containing variations of "white" were—that the gap between these trajectories persisted even when all items containing variations of "white" (as opposed to only those coded as relevant and substantive) were included in the analysis. Annual frequencies for variations of the phrase "of color" ranged from a low of 2% in 1993 to a high of 18% in 2009. Annual frequencies for variations of "minority" ranged from a low of 4% in 1985 to a high of 24% in 2001. Similar to variations of "race," annual frequencies for variations of "white" ranged from a low of 4% in 1982 to a high of 25% in 1977. Figure 9 illustrates the trajectory for variations of "minority" and "white" in relation to variations of the phrase "of color" in U.S. publications.

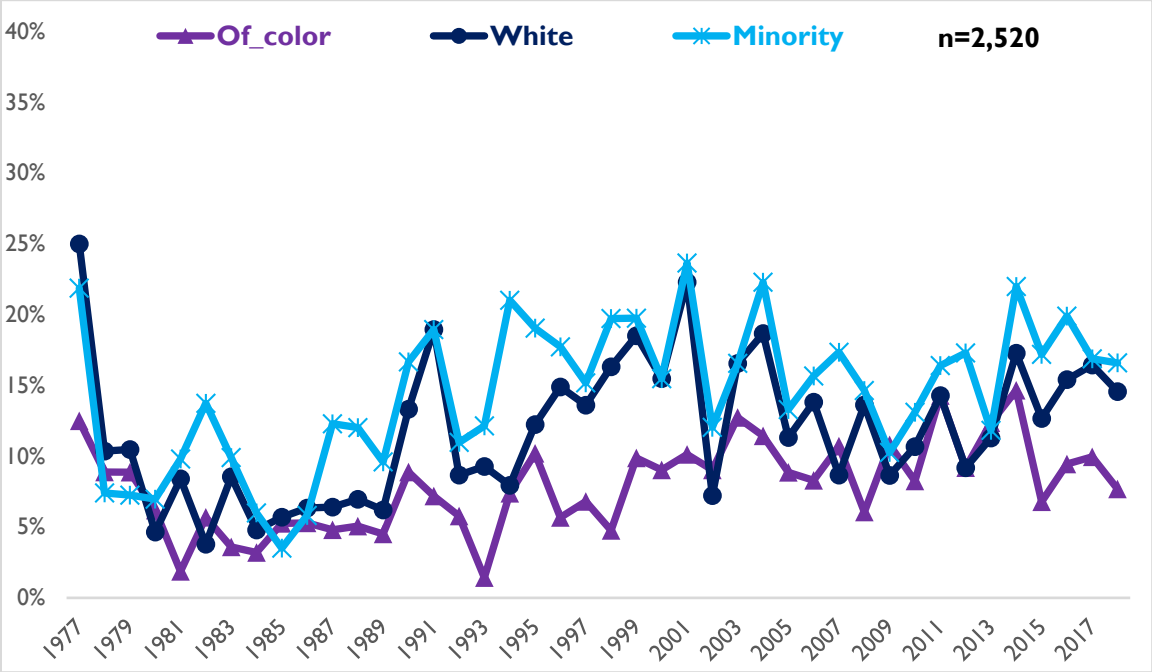

Figure 9. Annual frequencies for variations of "minority," "white," and "of color" in U.S. evaluation publications

U.S. evaluation literature: Annual frequencies for variations of "culture" surpassed that for variations of "race." Again, only items using variations of "culture" to signify racialized difference were included in the diachronic analysis. The overall frequency for variations of "culture" (17%) in the U.S. evaluation literature was close to the overall frequency for variations of "race" (16%). However, annual frequencies for variations of "culture" ranged from a low of 3% in 1977 and 1984 to a high of 36% in 2015 and 2018. In almost direct contrast, annual frequencies for variations of "race" ranged from a low of 5% in 1989 to a high of 34% in 1977. Similarly, the annual frequency for variations of "culture" increased from 3% in 1977 to 36% in 2018 while the annual frequency for "race" decreased from 34% in 1977 to 20% in 2018. Unlike annual frequencies for variations of "diversity" and "inclusion," which decreased from 2016 to 2018, the annual frequency for "culture" increased from 33% to 36%. Figure 10 illustrates the trajectory for variations of "culture" and "race" in U.S. evaluation publications as a whole.

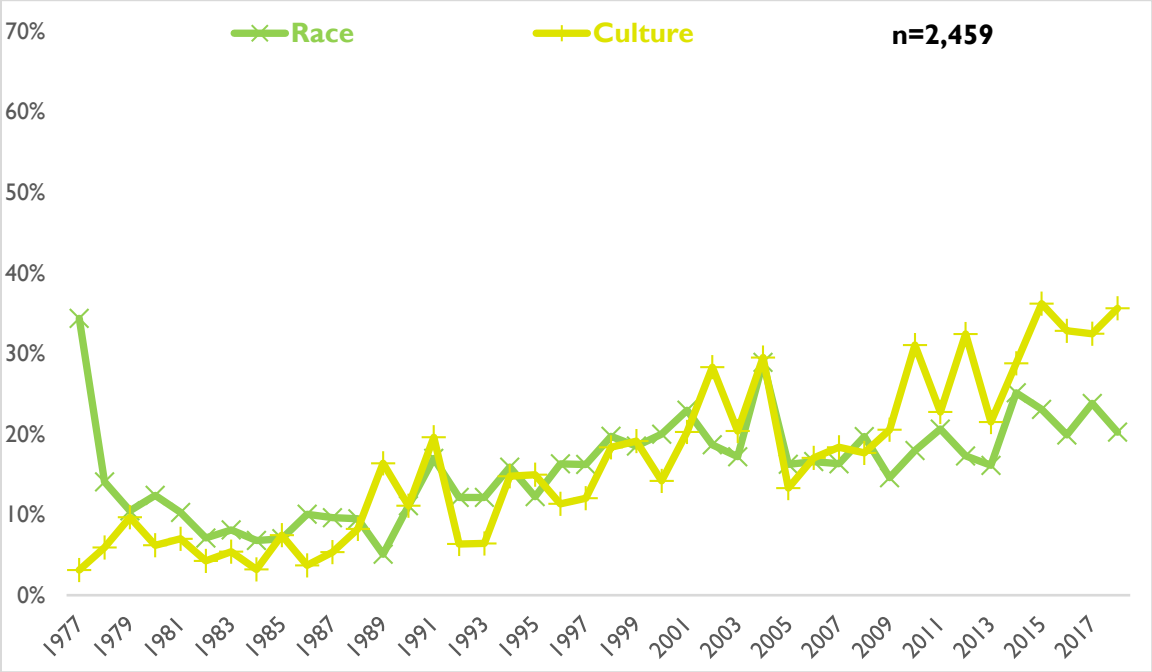

Figure 10. Annual frequencies for variations of "culture" and "race" in U.S. evaluation publications

U.S. evaluation literature: Variations of "indigenous" had a higher overall frequency than that for variations of "colonization." Only 15% of U.S. evaluation literature that contained variations of "indigenous" also contained variations of "colonization." Forty-seven percent of U.S. items that contained variations of "colonization" also contained variations of "indigenous." The overall frequency for variations of "indigenous" (4%) was higher in the U.S. evaluation literature than that for variations of "colonization" (1%). It is important to note—because items containing variations of "colonization" were coded for relevance and substantiveness whereas items containing variations of "indigenous" were not—that the gap between the trajectories for variations of "indigenous" and that of "colonization" persisted even when all items that contained variations of "colonization" (as opposed to only those coded as relevant and substantive) were included in the analysis.

Annual frequencies for variations of "indigenous" ranged from a low of 0%, which characterized most years between 1977 and 2018, to a high of 17% in 2018. Annual frequencies for variations of "colonization" ranged from a low of 0%, which also characterized most years between 1977 and 2018, to a high of 6% in 2018. The annual

frequency for variations of "indigenous" increased—from 6% to 17% between 1977 and 2018—as did the annual frequency for variations of "colonization," but at a slower rate (from 3% to 6% during the same period). Annual frequencies for variations of both increased after 2005, at which point the annual frequency for variations of "indigenous" was 1% and that for variations of "colonization" was 0%. Figure 11 illustrates the trajectory for variations of "indigenous" and "colonization" in U.S. publications.

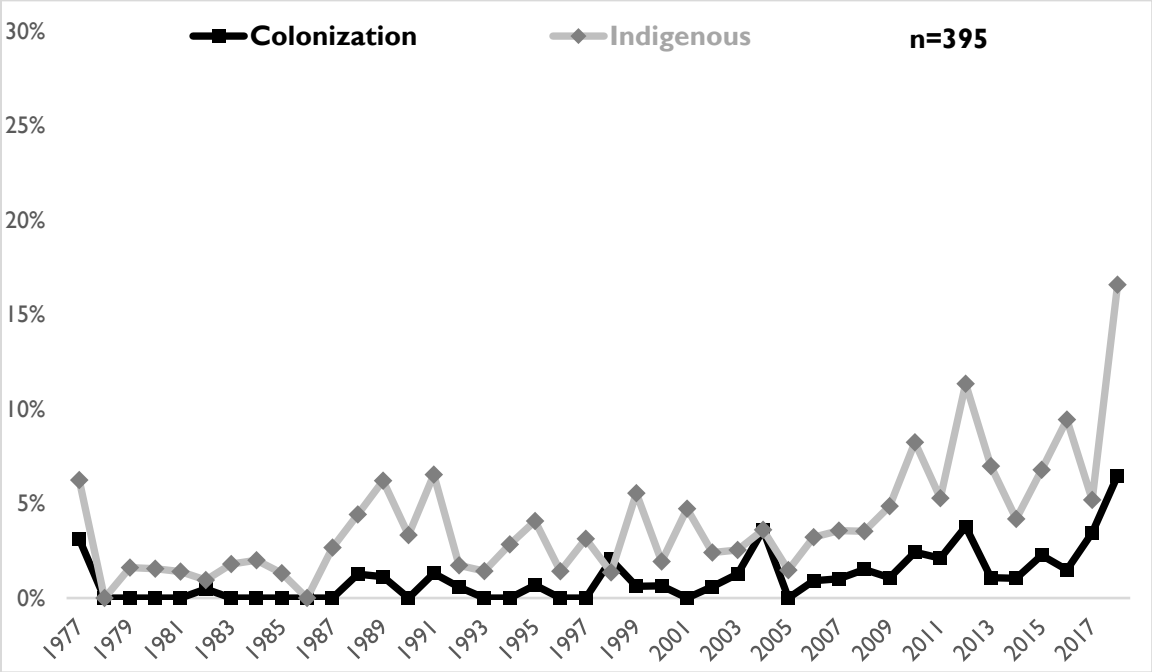

Figure 11. Annual frequencies for variations of "indigenous" and "colonization" in U.S. evaluation publications

U.S. evaluation publications unaffiliated with AEA: Annual frequencies for variations of "diversity" and "inclusion" increased until 2016. Again, variations of "diversity" and "inclusion" signify dimensions of difference beyond racialized difference and items containing those search terms were not coded with regard to their specificity to race. In U.S. journals unaffiliated with AEA, the overall frequency for variations of "diversity" (36%) and "inclusion" (29%) both increased between 1977 and 2016, at which point they decreased. Annual frequencies for variations of "diversity" ranged from a low of 12% in 1980 to a high of 65% in 2016. Annual frequencies for variations of "inclusion" ranged from a low of 7% in 1978 to a high of 51% in 2016. Figure 12 illustrates this trajectory.

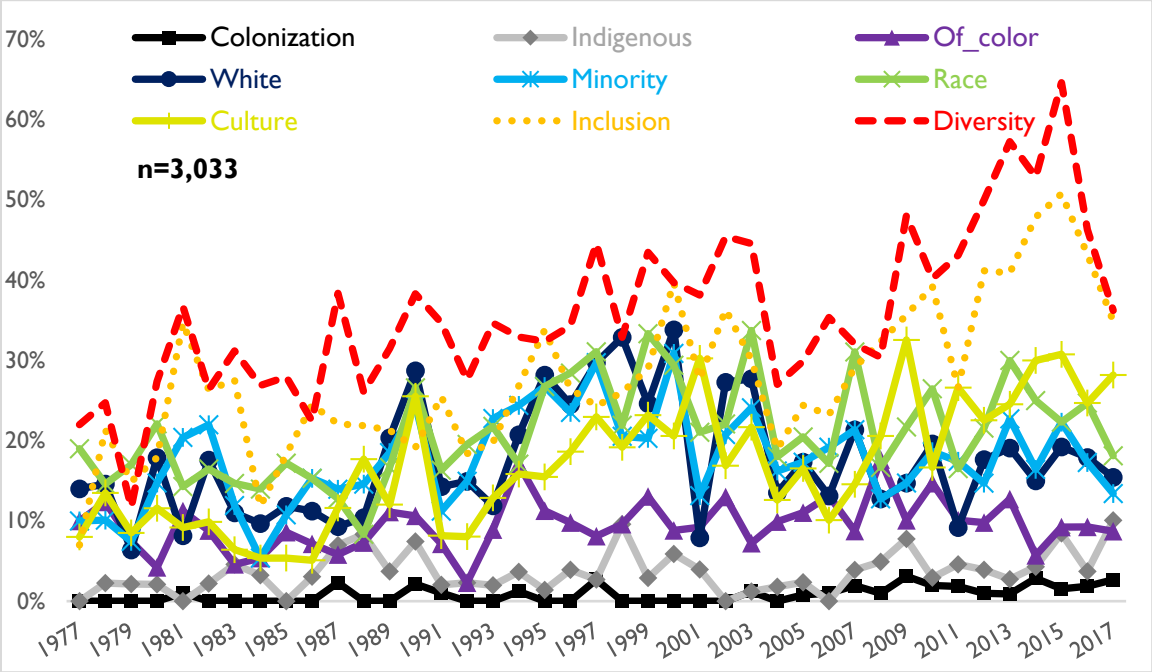

Figure 12. Trajectory in U.S. evaluation publications unaffiliated with AEA for terms identified through textual analysis

U.S. evaluation publications unaffiliated with AEA: Variations of "minority" and "white" had a higher overall frequency than that for variations of the phrase "of color." In U.S. evaluation items published in journals unaffiliated with AEA, overall frequencies for variations of "minority" and "white" (both 17%) were higher than those for variations of the phrase "of color" (10%). Annual frequencies for variations of "white" and "minority" increased in the late 1980s and mid- to late-1990s. Annual frequencies for variations of "minority" ranged from a low of 5% in 1985 to a high of 31% in 2001, and that for variations of "white" comparably ranged from a low of 6% in 1980 to a high of 34% in 2001, at which point both began to decrease—albeit erratically. Annual frequencies for variations of the phrase "of color" ranged from a low of 2% in 1993 to a high of 18% in 2009. Figure 13 illustrates the trajectory for variations of "minority" and "white" in relation to variations of the phrase "of color" in U.S. publications unaffiliated with AEA.

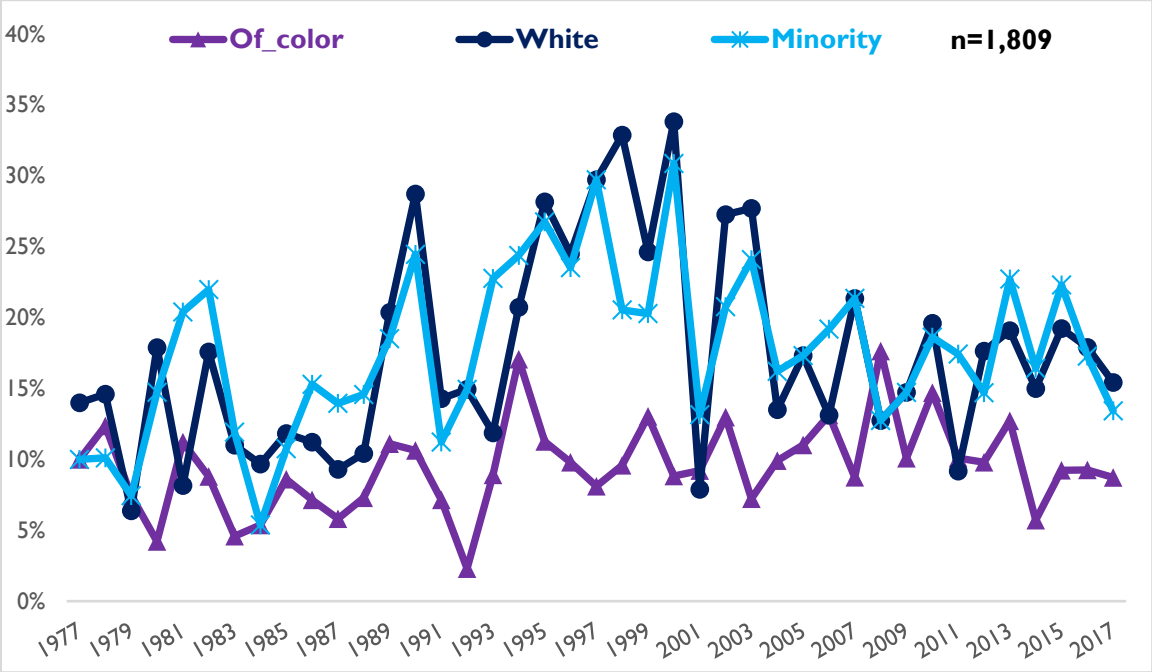

Figure 13. Annual frequencies for variations of "minority," "white," and "of color" in U.S. evaluation publications unaffiliated with AEA

U.S. evaluation publications unaffiliated with AEA: Variations of "race" had a higher overall frequency than that for variations of "culture." Again, only items using variations of "culture" to signify racialized difference were included in the diachronic analysis. Unlike in the U.S. evaluation literature as a whole, in evaluation publications unaffiliated with AEA the overall frequency for variations of "culture" (17%) was slightly lower than that for variations of "race" (21%). Annual frequencies for variations of "culture" ranged from a low of 3% in 1977 to a high of 33% in 2010. Annual frequencies for variations of "race" ranged from a low of 8% in 1989 to a high of 34% in 1977 and 2004. However, the annual frequency for variations of "race" decreased slightly—from 19% to 18% between 1977 and 2018—whereas the annual frequency for variations of "culture" increased—from 8% to 28% during the same period. The annual frequency for variations of "culture" surpassed the annual frequency for variations of "race" in 2014 and increased from 25% to 28% between 2017 and 2018. Figure 14 illustrates the trajectory for variations of "culture" and "race" in U.S. publications unaffiliated with AEA.

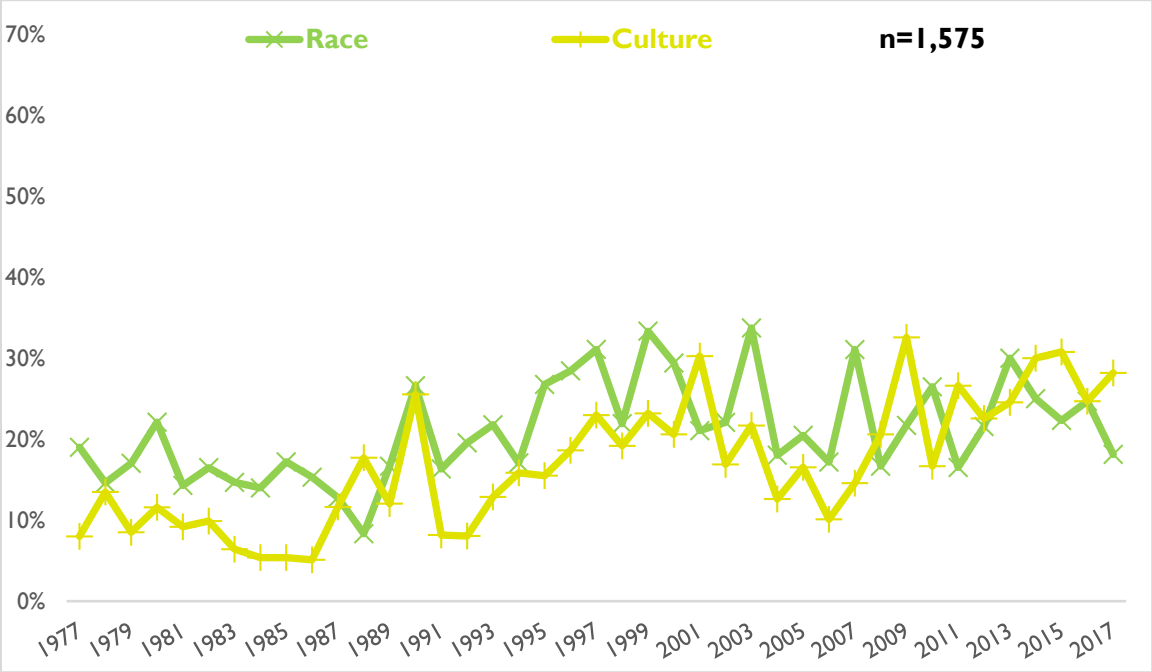

Figure 14. Annual frequencies for variations of "culture" and "race" in U.S. evaluation publications unaffiliated with AEA

U.S. evaluation publications unaffiliated with AEA: Variations of "indigenous" had a higher overall frequency than that for those of "colonization." While approximately half of U.S. items that contained variations of "colonization" also contained variations of "indigenous," that percentage differed dramatically depending on AEA affiliation. Only 35% of AEA-unaffiliated publications that contained variations of "colonization" also contained variations of "indigenous." In U.S. publications unaffiliated with AEA, the overall frequency for variations of "indigenous" (4%) was higher than that for variations of "colonization" (1%). It is important to note—because items containing variations of "colonization" were coded for relevance and substantiveness whereas items containing variations of "indigenous" were not—that the gap between the trajectories for variations of "indigenous" and that of "colonization" persisted even when all items that contained variations of "colonization" (as opposed to only those coded as relevant and substantive) were included in the analysis. Annual frequencies for variations of "indigenous" ranged from a low of 0%, which characterized most years between 1977 and 2018, to a high of 10% in 2018. Annual frequencies for variations of "colonization"

ranged from a low of 0%, which also characterized most years between 1977 and 2018, to a high of 3% in 1977, 1998, 2010, 2015, and 2018. Annual frequencies for variations of both "indigenous" and "colonization" in U.S. evaluation publications unaffiliated with AEA increased slightly and at comparable rates of increase between 1977 and 2018. Figure 15 illustrates the trajectory for variations of "indigenous" in relation to "colonization" in U.S. publications unaffiliated with AEA.

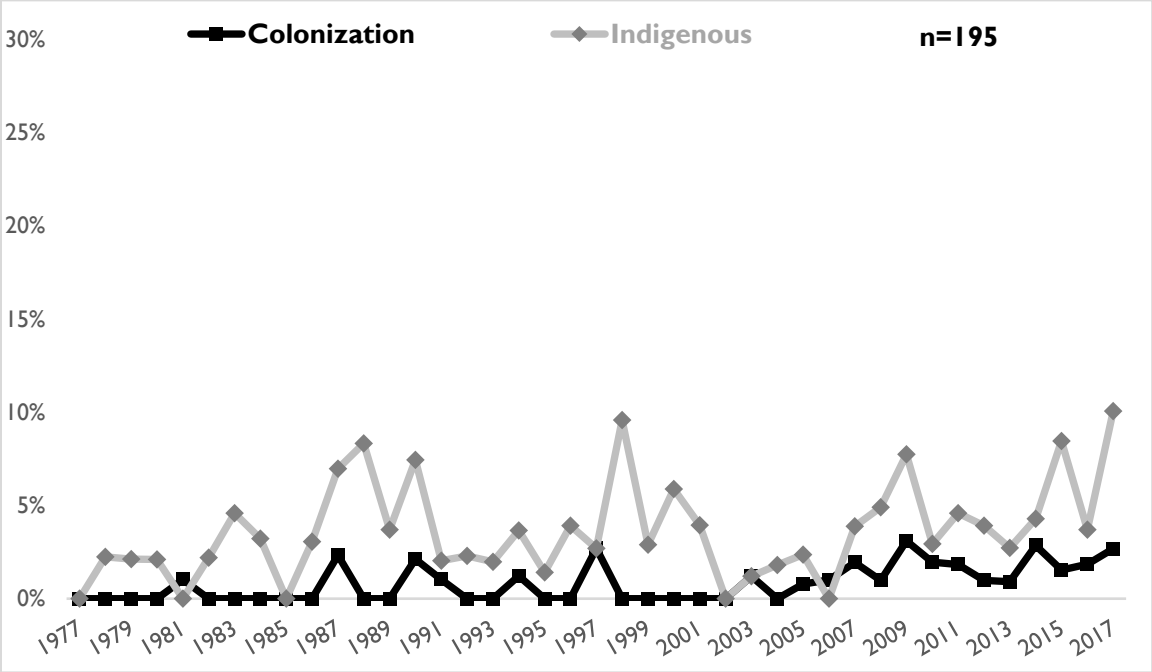

Figure 15. Annual frequencies for variations of "colonization" and "indigenous" in U.S. evaluation publications unaffiliated with AEA

U.S. evaluation publications affiliated with AEA: Annual frequencies for variations of "diversity" and "inclusion" continued to increase. Again, variations of "diversity" and "inclusion" can potentially signify dimensions of difference beyond racialized difference and items containing those search terms were not coded with regard to their specificity to race. Of the AEA-affiliated items that contained variations of "diversity," 59% also contained variations of "culture" and 25% also contained variations of "race." A full 100% of the items that contained variations of "diversity" also contained variations of "inclusion," regardless of publications' geographic location or AEA affiliation. Not all items that contained variations of "inclusion" also contained variations of "diversity," however; only 61% of such items in AEA-affiliated literature also

contained variations of "diversity." In AEA-affiliated journals, overall frequencies for variations of "diversity" (27%) and "inclusion" (15%) increased between 1977 and 2018, and—unlike in U.S. evaluation literature unaffiliated with AEA and U.S. evaluation literature as a whole—continued to do so between 2017 and 2018. Annual frequencies for variations of "diversity" ranged from a low of 0% in 1978 and 1979 to a high of 66% in 2016. Annual frequencies for variations of "inclusion" range from a low of 0%, which characterized four different years between 1977 and 2018, to a high of 52% in 2016. Figure 16 illustrates this trajectory.

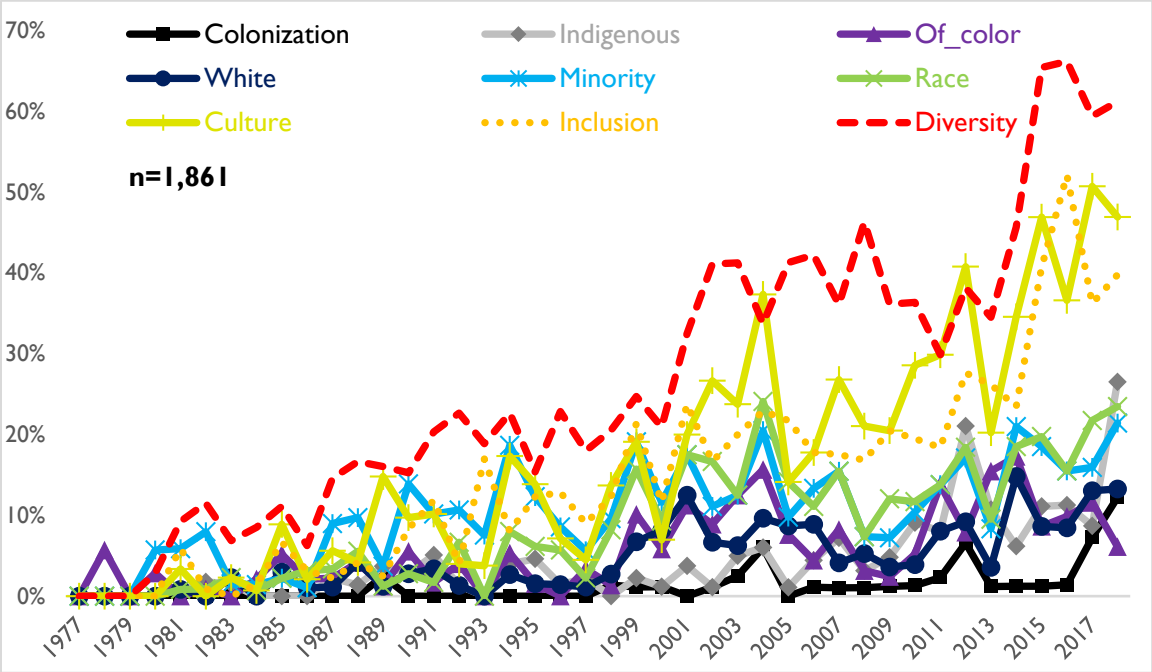

Figure 16. Trajectory in U.S. evaluation publications affiliated with AEA for terms identified through textual analysis

U.S. evaluation publications affiliated with AEA: Variations of "minority" had a higher overall frequency than that for those of "white" and the phrase "of color." Thirty percent of AEA-affiliated literature that contained variations of "minority" also contained variations of "white" and 37% of such literature also contained variations of the phrase "of color." The overall frequency for variations of "minority" (10%) in AEA-affiliated publication items was higher than that for variations of both "white" (5%) and the phrase "of color" (6%). It is important to note—because items containing variations of "minority" and the phrase "of color" were not coded for relevance and substantiveness,

whereas items containing variations of “white” were—that the gap between these trajectories persisted even when all items containing variations of “white” (as opposed to only those coded as relevant and substantive) were included in the analysis. Annual frequencies for variations of “minority” ranged from a low of 0% in 1978 and 1979 to a high of 21% in 2014 and 2018. Annual frequencies for variations of “white” ranged from a low of 0%, which characterized six different years between 1977 and 2018, to a high of 15% in 2014. Annual frequencies for variations of the phrase “of color” ranged from a low of 0%, which characterized five different years between 1977 and 2018, to a high of 17% in 2014. Items in AEA-affiliated and unaffiliated publications differed in their overall frequencies for variations of “white” and the phrase “of color.” The overall frequency for variations of “white” in journals affiliated with AEA was less than one-third that in journals unaffiliated with AEA. In AEA-affiliated journals, the overall frequency for variations of the phrase “of color” was similarly less than half that in journals unaffiliated with AEA. However, the annual frequency for variations of “white” increased between 2016 and 2018, although at a slower rate than it did for variations of “minority.” Figure 17 illustrates the trajectory for variations of “minority” relative to variations of “white” and the phrase “of color” in U.S. publications affiliated with AEA.

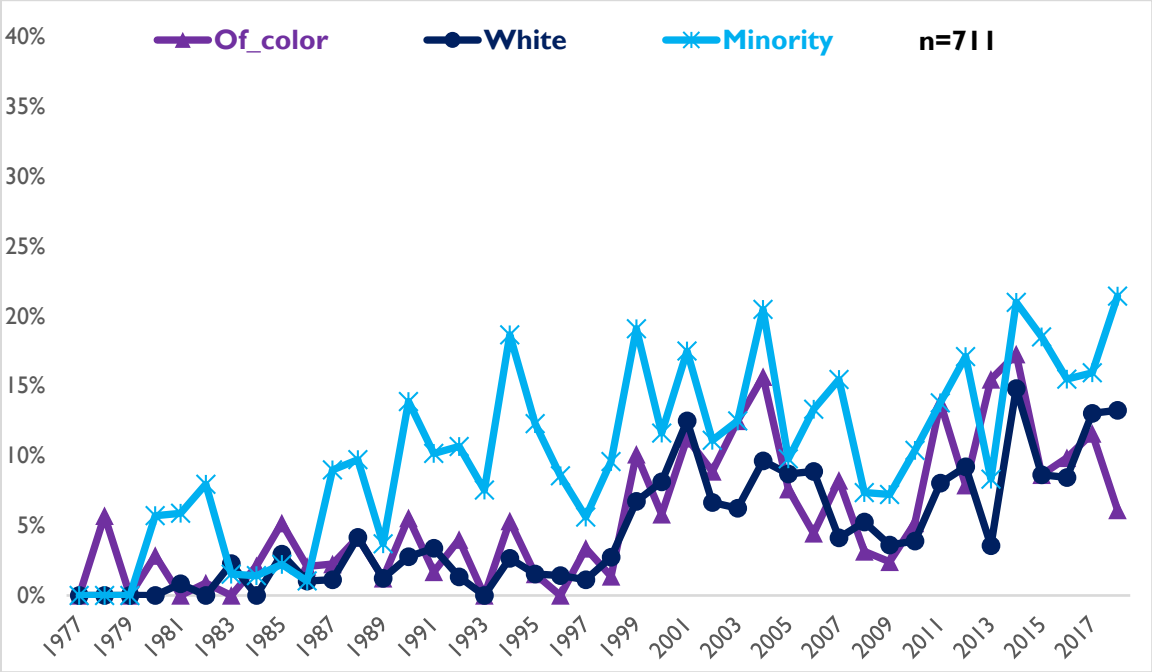

Figure 17. Annual frequencies for variations of "minority," "white," and "of color" in U.S. evaluation publications affiliated with AEA

U.S. evaluation publications affiliated with AEA: Variations of "culture" had a higher annual frequency than that for those of "race." Again, only items using variations of "culture" to signify racialized difference were included in the diachronic analysis. In direct contrast with U.S. evaluation publications unaffiliated with AEA, in items in AEA-affiliated publications the overall frequency for variations of "culture" (17%) was lower than that for variations of "race" (9%). Annual frequencies for variations of "culture" ranged from a low of 0%, which characterized four different years between 1977 and 2018, to a high of 52%—in 2016. Annual frequencies for variations of "race" ranged from a low of 0%, which characterized four different years between 1977 and 2018, to a high of 24% in 2004. The highest annual frequency for variations of "culture" within AEA-affiliated evaluation journal items was 50 percent greater than that in non-affiliated journal items.

Similarly, the overall frequency for variations of "race" in journals affiliated with AEA was less than half that in those unaffiliated with AEA. Annual frequencies for variations of both "culture" and "race" increased between 1977 and 2018, from 0% to 47% for variations of "culture" and from 0% to 23% for variations of "race," with a

higher rate of increase taking place for both after 2001 and especially between 2010 and 2017. From 2017 to 2018, however, annual frequencies for variations of "culture" decreased slightly—from 51% to 47%—while the annual frequency for variations of "race" increased even more slightly, from 22% to 23%. Figure 18 illustrates this trajectory for variations of "culture" in relation to that for variations of "race" in U.S. publications affiliated with AEA.

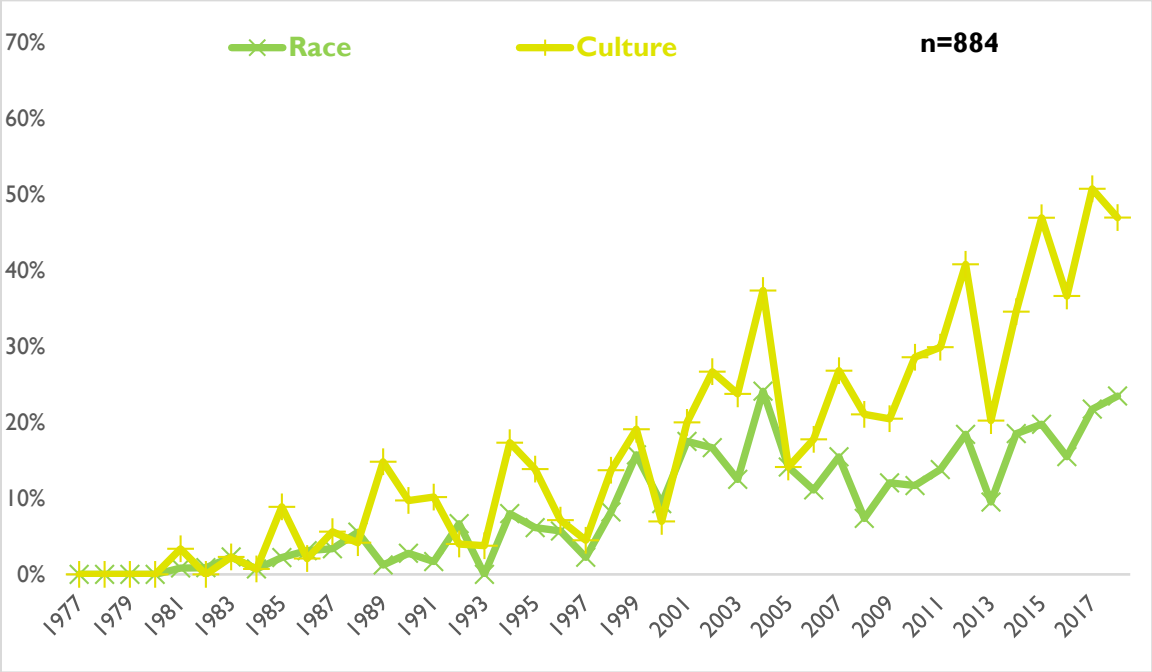

Figure 18. Annual frequencies for variations of "culture" and "race" in U.S. evaluation publications affiliated with AEA

U.S. evaluation publications affiliated with AEA: Variations of "indigenous" had a higher annual frequency than that for variations of "colonization." Only 22% of AEA-affiliated literature that contained variations of "indigenous" also contained variations of "colonization." While approximately half of U.S. items that contained variations of "colonization" also contained variations of "indigenous," that percentage differed dramatically depending on AEA affiliation. Nearly three-quarters (74%) of AEA-affiliated publications that contained variations of "colonization" also contained variations of "indigenous." In AEA-affiliated publications, the overall frequency for variations of "indigenous" (5%) was higher than that for variations of "colonization" (1%). It is important to note—because items containing variations of "colonization" were

coded for relevance and substantiveness whereas items containing variations of "indigenous" were not—that the gap between the trajectories for variations of "indigenous" and that of "colonization" persisted even when all items that contained variations of "colonization" (as opposed to only those coded as relevant and substantive) were included in the analysis.

Annual frequencies for variations of "indigenous" ranged from a low of 0%, which characterized ten different years between 1977 and 2018, to a high of 27% in 2018. Annual frequencies for variations of "colonization" ranged from a low of 0%, which characterized 21 different years between 1977 and 2018, to a high of 12%—also in 2018. The annual frequency for variations of "indigenous" in publications affiliated with AEA increased between 1977 and 2018 and especially between 2005 and 2018. The overall frequency for variations of "indigenous" in AEA-affiliated journals was almost three times that in U.S. journals that were un-affiliated with AEA. Similarly, the overall frequency for variations of "colonization" in AEA-affiliated journals was four times that in U.S. journals that were unaffiliated, and the rate at which variations of both terms were increasing was faster in AEA-affiliated journals, as well. Figure 19 shows the trajectory for variations of "indigenous" in relation to "colonization" in U.S. publications affiliated with AEA.

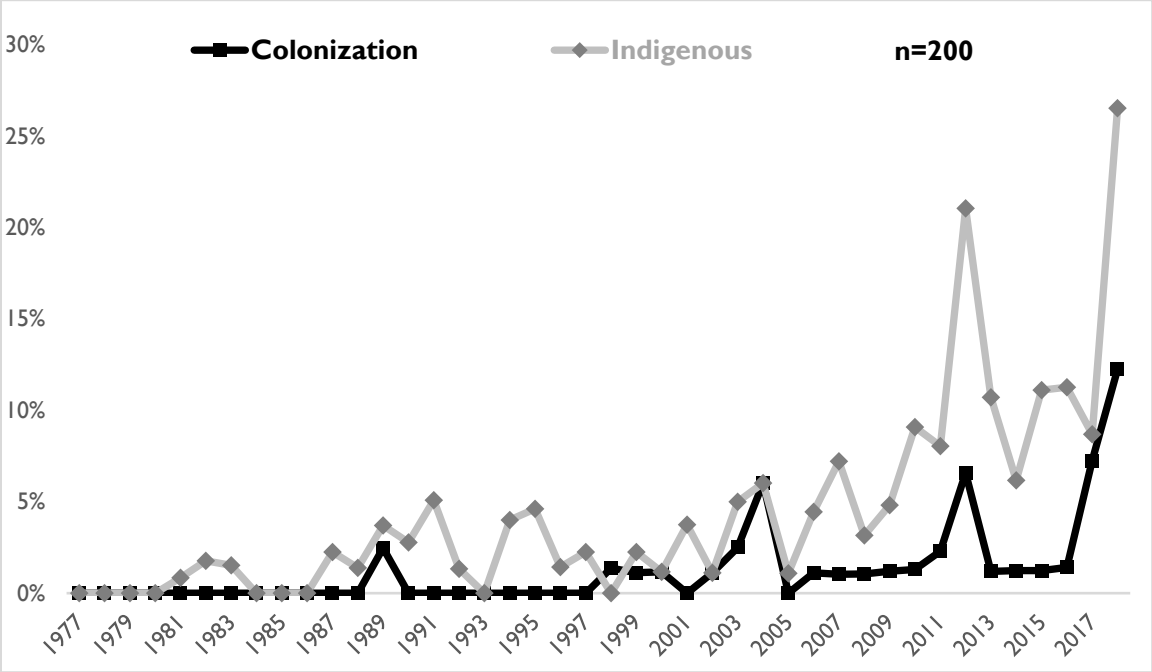

Figure 19. Annual frequencies for variations of "indigenous" and "colonization" in U.S. evaluation publications affiliated with AEA

Interpretation and use of findings. The above findings point to one constant and three areas of change in the U.S. evaluation literature between its inception and 2018. The constant is described below and descriptions of the three areas of change follow it in roughly chronological order:

- Consistent use of variations of "minority" relative to "of color": U.S. evaluation literature tended to signify racialized difference in terms of minority and majority status. This was fairly consistent between 1987 and 2018 regardless of AEA affiliation, but it was more characteristic of items from journals unaffiliated with AEA than it was of those affiliated with AEA. Specification of the "white" majority was inconsistent, however, over time and AEA-affiliation.
- Rise of variations of "diversity" and "inclusion": While the annual frequencies for variations of "diversity" and "inclusion" eclipsed nearly all other potential signifiers of racialized difference between approximately 2001 and 2016, they declined between 2017 and 2018 in literature unaffiliated with AEA and in the U.S. literature as a whole. In contrast, they increased between 2017 and 2018 in literature affiliated with AEA.

- Rise of variations of "indigeneity" and "colonization": In U.S. evaluation literature, regardless of AEA affiliation, indigeneity was generally decoupled from colonization. However, the annual frequencies for both increased—most dramatically in literature affiliated with AEA, especially between 2005 and 2018.
- Replacement of variations of "race" with variations of "culture": In U.S. evaluation literature, the annual frequency for variations of "culture" surpassed that for variations of "race" as a signifier of racialized difference. This was consistent regardless of AEA affiliation. However, it was most characteristic of items from journals affiliated with AEA. Relative to each other, annual frequencies for variations of "race" were consistently higher than those for variations of "culture" between 1977 and 1988, at which point the annual frequencies for variations of "culture" and "race" began to fluctuate; those of "race" were generally, but not consistently higher until 2008. In 2008, the annual frequency for variations of "culture" surpassed those of "race." The greatest variance in annual frequencies between variations of "culture" and "race"—with frequencies for variations of "culture" being higher than that for variations of "race"—occurred in 1989, 2002, the period between 2010 and 2012, and the period between 2015 and 2018. Annual frequencies for variations of "race" were higher than that of "culture" from 1985 to 1988, 1992 to 1994, and 1999 to 2002.

Conclusion of Chapter Five

The diachronic analysis focused on the construction of racialized difference *among individual* pieces of peer-reviewed literature and among individuals or small groups of individuals. The search and subsequent sample of peer-reviewed journal content, aside from book reviews, for the diachronic analysis were driven by the nine terms generated through the textual analysis. The frequencies of items containing substantive and relevant use(s) of these nine terms, relative to the total number of items published, by publication and by year, were calculated and plotted graphically to illustrate how evaluation's construction of racialized difference had changed between 1977 and 2017. Findings consisted of four over-arching trajectories: (1) the consistent use of variations of "minority" and inconsistent use of variations of "white" relative to "of

color"; (2) the rise of variations of "diversity" and "inclusion"; (3) the rise of variations of "indigeneity" and "colonization"; (4) the replacement of variations of "race" with variations of "culture."

The variation between AEA-affiliated and -unaffiliated journal items suggests that journals' affiliation with AEA might have influenced the language that their content contained. This influence could have stemmed from the authors' membership with AEA and socialization into the professional association's norms—including its discourse—or into the affiliated journals' language—by reading their content, access to which is complimentary to dues-paying members. It could also have stemmed from the intentional emulation or citation of authors already published in the journals by authors seeking to publish in them. Finally, it could have stemmed from decisions made about the journals' format, special issues, and reviewers by the journals' editorial board as well as the reviewers' decisions about individual items. All are examples of the discursive production of racialized difference. Pivotal dates, events, and scholars identified through the textual analysis and diachronic analysis informed the selection of documents, interviewees, interview questions, and outside literature to contextualize these findings through the archival analysis and CST interviews.

CHAPTER 6: RESULTS OF ARCHIVAL ANALYSIS

How Does the Trajectory of Language Regarding Racialized Difference Relate to the Systems Surrounding Evaluation?

Critical Discourse Analysis (CDA): Archival Analysis

Documents published by AEA comprised the third source of data for the CDA. This data source corresponds with the institutional and structural levels of the social-ecological model in that it identifies relationships contextually.

Data collection. Periods identified through the textual analysis and diachronic analysis as those during which shifts occurred in the discursive strategies related to racialized difference and annual frequencies of potential signifiers of racialized difference guided the search for additional material that could contextualize the shifts. These searches were for material in AEA-affiliated journals as well as documents published by AEA and were conducted from the *American Journal of Evaluation* and *New Directions for Evaluation* websites and AEA's website, respectively. Additionally, a search for documents containing variations of the nine terms that potentially signify racialized difference identified through the textual analysis was conducted on AEA's website.

The identities of journals' individual editors-in-chief might have borne some influence on the content that was published within their journals and could therefore be considered worthy contextual data. From the perspective of systems thinking and critical theories of systemic oppression described in Chapter Two, however, identity is neither inherent nor fixed; in other words, critical theories of systemic oppression would challenge the assumption that editors-in-chief who self-identify as members of a racially otherized group would necessarily advance critical constructions of difference. Furthermore, institutional and structural mechanisms amplify and counteract the potential of any individual to influence a system, regardless of their intentions. These mechanisms include the publication-related decision-making processes, socio-economic relations that underlie which group interests are represented among those who submit manuscripts and those who comprise editorial boards, the training both receive, the standards and

principles that they help develop and to which they professionally subscribe, and the editorial review process that they follow. These decision-making processes and socio-economic relations shape the potential for individual editors-in-chief who identify with or are classified under particular racial groups, or—more importantly—who have an interest in advancing particular constructions of difference (liberal or critical), to exert influence on the language used within the journal.

Moreover, the conventions of both academia and evaluation disfavor authors' disclosure of racial group identification; such disclosure within the evaluation literature, as noted in the textual analysis, has taken place disproportionately among those who (1) identified with a racially otherized group and (2) challenged one or more characteristic of liberal constructions of difference. Assuming that the racial group identification of those who did not disclose it was White would reinforce the normativity of whiteness. As a result, the racial group identification of editors, authors, AEA presidents, and other *individuals* was not included among the data considered in attempting to answer the third research question, about how the trajectory related to surrounding systems. In keeping with critical theories of systemic oppression, only the explicit articulation of *group interests*—for example, that of the Multi-ethnic Issues in Evaluation or Indigenous Peoples in Evaluation Topical Interest Group—were considered.

Data sampling and processing. Searches were conducted for literature and documents disseminated during and immediately surrounding the years during which one or more shift in discourse was identified through the diachronic analysis in relation to the textual analysis. The shifts observed include: (1) consistent use of variations of "minority" and inconsistent use of variations of "white" relative to "of color" over time and journal affiliation; (2) the rise of variations of "diversity" and "inclusion" over time and journal affiliation; (3) the rise of and decoupled relationship between variations of "indigeneity" and "colonization" over time and journal affiliation; (4) the replacement of variations of "race" with variations of "culture" over time and journal affiliation.

The time periods identified as potentially pivotal in the above shifts included 1985-1994, 1999-2008, and 2016-2018. The periods delimiting these searches were

selected with consideration for the time delays (distal effects) among discursive events, between them and the meso-level, institutional context of the fields surrounding them, and between them and the macro-level social, political, and historical context surrounding the field. Additionally, a search for variations of the nine search terms identified through the textual analysis was conducted on AEA's website, including information about the association's history and structure, Topical Interest Groups, and conference program history, as well as in its library and community discussion fora. Documents internal to AEA were not included in the archival analysis. The search yielded the meso-level, institutional discursive events that are described by year and by U.S. president (as an indication of the macro-level social, political, and historical context) in Table 11. The years during which the diachronic analysis revealed a potential discourse shift (1985-1994, 1999-2008, and 2016-2018) are indicated in bold type as are the potential signifiers of racialized difference (colonization, culture, diversity, inclusion, indigenous, minority, of color, race, and white) that each discursive event contains.

Table 11. *Meso-level, institutional discursive events signifying racialized difference in the U.S. field of evaluation, by year and macro-level/structural context of USA*

U.S. socio-political context	Years	Meso-level, Institutional Discursive Events
Reagan	1984	• **Culture** was discussed in relation to demographic changes in USA in international panels at Evaluation Research Society/Evaluation Network's (precursors to AEA) joint annual conference
	1985	• *New Directions for Program Evaluation* edition on **Culture** & Evaluation (Patton) was published
	1987	• **Minority** Task Force, led by Madison, was established within AEA
	1988	• Enslaved labor, **colonization** of **indigenous** Africans, people **of color**, **culture** as well as **race** and **white** as produced was addressed at AEA's annual conference through keynote speech (Hilliard) under Patton presidency
G. H. W. Bush	**1989**	• Cross-**cultural** evaluation was included in AEA's annual conference theme by president Conner
	1992	• *New Directions for Program Evaluation* edition on **Minority** Issues in Evaluation (Madison) was published
Clinton	**1994**	• **Multicultural** validity was addressed at AEA's annual conference, whose theme focused on social justice, through keynote speech by president Kirkhart
	1998	• **White** supremacy, **race**, **colonization, indigenous**, people **of color**, **culture, diversity, inclusion, and minorities** was addressed at AEA's annual conference, whose theme was Transforming Society through Evaluation, through keynote speech (Stanfield) under president Mertens
	1999-2001	• Building **Diversity** Initiative committee was established, study was conducted, and results were reported
G. W. Bush	**2001**	• **Diversity** Committee was established on AEA's Board of Directors
	2004	• **Minority** Issues in Evaluation TIG was re-named Multi-ethnic Issues in Evaluation TIG
		• **Cultural** Reading of Joint Committee on Standards for Educational Evaluation's (JCSEE) Standards for Program Evaluation was undertaken
		• Inaugural AEA/Duquesne University Internship (later Graduate Education **Diversity** Internship) cohort was launched

U.S. socio-political context	Years	Meso-level, Institutional Discursive Events
G. W. Bush continued		• *New Directions for Evaluation* edition on **Cultural** Competence (Thompson-Robinson, Hopson, and SenGupta) was published
	2005	• **Indigenous** Peoples in Evaluation TIG was established within AEA
	2007	• Tracking Transformation evaluation of Phase I of AEA's Building **Diversity** Initiative (Peak, et al.), which proffered the term **multiculturalism** as opposed to **cultural** competence, was conducted
Obama	2011	• Third edition of The Standards for Program Evaluation, which calls for an **inclusive** orientation that accounts for the **cultural** contexts of clients and other stakeholders, was published
		• Statement on **Cultural** Competence was approved by AEA's membership
	2012	• **Indigenous** inhabitants of land were honored in the opening ceremony of AEA's annual conference under president Hopson
		• **Decolonizing** methodologies addressed at AEA's annual conference through keynote speech (Tuhiwai Smith) under president Hopson
	2013	• First **Culturally** Responsive Evaluation & Assessment conference took place at University of Illinois-Urbana Champaign under Hood
	2014	• Latinx Research and Evaluation Discourse whose mission involves theory, research, and practice of **culturally** responsive evaluation, was co-founded by Lisa Aponte-Soto and others within AEA
	2015	• **Multiculturalism** was included in AEA's annual conference theme by president Donaldson
	2016	• AEA Dialogues on **Race** & Class took place under president Newcomer
Trump	**2017**	• Graduate Education **Diversity** Internship TIG was established
	2018	• Evaluator Competencies, which include **cultural** competence, were approved by AEA's membership
		• *New Directions for Evaluation* edition on **Indigenous** Evaluation (Cram, Tibbetts, and LaFrance) was published
		• *American Journal of Evaluation* section on **Race** and Evaluation (Rallis) was published
		• Revised Guiding Principles for evaluators—featuring **diversity**, **cultural** competence, **race**—were approved by AEA's membership

Findings. Ways that the search terms' annual frequencies shifted over time or varied by AEA affiliation were detailed in the findings from the diachronic analysis. Specific comparisons (minority, of color, and white; culture and race; indigenous and colonization; and diversity and inclusion) illuminated when and under what conditions U.S. evaluation literature focused on: (1) minority status or indigeneity without acknowledging whiteness in the historical and structural contexts of, for example, exclusionary immigration policies and settler colonization; (2) culture without acknowledging social group interests and identification developed in response to structurally mediated relations, such as that between property and property owner, which produced distinctions in race; or (3) diversity and inclusion, which imply random assortments and lateral relations without acknowledging highly codified systems of stratification. Visually plotting the meso-level, institutional discursive events that took place within AEA on to these comparative trajectories pointed to pivotal periods, events, and actors to include for study in the interview strand. After a short introduction to the origins of AEA, findings from the archival analysis are described below. Discursive events that were discussed at length in the findings from the textual analysis—largely items from AEA-affiliated journals—are noted only briefly to historicize shifts in annual frequencies, whereas those that represent AEA decisions and documents are analyzed in greater detail.

Origins of the American Evaluation Association. What is currently the American Evaluation Association's predecessors arose when federal spending on social science research flourished as part of Great Society Programming and the War on Poverty in particular. They include the May 12 Group, Council for Applied Social Research (CFASR), Evaluation Network (ENet), and Evaluation Research Society (ERS). Each of these is described below.

In the late 1960s, education researchers from academia started creating a professional community in the nascent field of evaluation through an annual gathering that was intentionally invitation-only. In fact, it selected the May 12 Group as its name, after the first date on which they met so that it would

> … have absolutely zero cachet, and so no one will be able to argue that they were entitled to join the May 12th group because it's called something generic. And so the idea was you got invited to the May 12th group, and if you weren't invited, then you weren't in, and so there was no official stuff." (Oral History Project Team, 2005)

The group—for which there is no comprehensive list, let alone demographic data—met informally in someone's house or university. While Lincoln and Guba recollected that "the program evaluation community was so small in the late 1960s and early 1970s that virtually everyone who was writing about evaluation could be invited [to the May 12 Group] and accommodated comfortably" (Lincoln & Guba, 2004, p. 237), Hood has often noted that African American educational researchers had been publishing since the 1940s in publications to which they had access, like the *Journal of Negro Education* (Hood, 2001). None of the educational researchers documented by the Nobody Knows My Name project has been listed among various members' recollection of May 12 Group attendees.

The May 12 meetings continued until some members felt that the need to organize something more official than "an intellectual elite group…[to push] for reform in the way in which educational movements were being appraised" (Oral History Project Team, 2005). With funding from Phi Delta Kappa, the education honorary society, they held the first meeting of 10 to 15 individuals in 1974 or 1975 in Snowmass, Colorado, to talk about the need to develop a network of evaluators that would become the Evaluation Network (ENet). They created a newsletter and held an annual conference. Unlike the May 12 Group, ENet's membership consisted of both academics and nonacademic practitioners. Based in the Midwest, its membership grew to 600-700 nationwide. They were overwhelmingly evaluators who had an academic public-school function.

Also in the mid- to late 1970s, Clark Abt, founder of Abt Associates—a large, privately-held contract research firm—founded the Council for Applied Social Research (CFASR). Unlike ENet, which focused on small-scale evaluations and local practice, CFASR focused on large-scale evaluations of national programs (Oral History Project Team, 2004, p. 248). Cronbach and others noted that "the firms that compete for contracts have begun to work together, particularly through the Council for Applied

Social Research" to influence funders' administrative practices and to publicly honor officials whose planning and contracting performances the members admired (, Chronbach, 1980, p. 355 as cited in Gargani, 2011). "The council thus serves to express industry views on matters of common interest" (Chronbach, 1980, p. 355 as cited in Gargani, 2011). The first annual meeting, held in Washington, DC, in 1977, was attended by academics, professional consultants, and high-level governmental policy makers deeply involved with federally funded research.

CFASR's inaugural meeting was attended by Marcia Guttentag, Lois-ellin Datta, and Carol Weiss. They "noticed that with few exceptions the speakers were White men. This did not sit well with Guttentag who, believing that she could not make diversity a high priority for CFASR, organized [the Evaluation Research Society in 1976] as an alternative that would promote inclusiveness and social justice" (Gargani, 2011, p. 430). While less "entrenched in the machinery of government contracting" than CFASR, the resulting ESR was more academic and government-affiliated than ENet (Gargani, 2011, p. 430). Based on the East Coast, its membership consisted of policy makers, evaluation researchers in universities, evaluation practitioners, and human service providers. It was still closely affiliated with academic and government power structures, and its most visible members were described as "powerful male academics and policy makers," but not classified racially by Gargani (2011, p. 430). As such, the extent to which the goal of developing an alternative to the existing evaluation fora for White men was achieved is unclear, but Guttentag's initial conceptualization of ERS as an alternative represents an early awareness—not shown in Campbell's 1971 piece (reprinted in 1991)—that evaluators who were not classified as White men were practicing at the time, despite their absence at CFASR, as well as a value judgment—like that expressed by Covert and Conner approximately a decade later—that their absence at such fora was of concern. It also represented an early attempt to remedy that absence through the structure of a professional association. ERS devoted considerable effort to issues of training, work standards, and employment opportunities in evaluation. It published a Directory of Evaluation Training Programs and a set of Standards for Program Evaluation and

established a job information and exchange service. It also published a membership directory.

A merger between CFASR and ERS was approved by both memberships in 1980, resulting in a combined membership of about 1,700. During the early years of Reagan's presidency in the 1980s, when government funding for evaluation decreased dramatically, ENet and ERS—which had held joint conferences for some time—also started to discuss merging. ENet held its last joint annual conference with ERS still as a separate organization in 1985 and transferred its assets to the new, combined association—the American Evaluation Association—in 1986. Sage Publications renewed its contract with the ENet Board, and *Evaluation News* became *Evaluation Practice*, which later became the *American Journal of Evaluation*.

The discursive attention that AEA directed to "minority issues" coincided with shifts in the annual frequencies for variations of the term "minority" that were not limited to AEA-affiliated literature. As described above, the origins of AEA as a professional association lay in: (1) the May 12 Group, which was explicitly and intentionally organized to be exclusive as opposed to inclusive; (2) CFASR, which crystallized the close association among policy makers, nonprofit organizations, and researchers that was detailed in the segment on the nonprofit industry in Chapter Two; and (3a) ENet and (3b) ESR, which tried and succeeded to varying but limited degrees in creating broader alternatives to existing associations of evaluation scholars and practitioners. Furthermore, AEA was born nearly one generation after the boom in government-funded research and evaluation accompanying Great Society programming had birthed its predecessors. Its birth was at a time when government funding was rapidly decreasing, in accordance with racialized neoliberalization's backlash against and devolution of redistributive policies and programs including Affirmative Action. The language of early efforts within AEA to focus attention on under-represented groups can be understood within this context.

Soon after ERS and ENet merged, Robert Covert became the new Association's second president by the flip of a coin. During his presidency in 1987, the Minority Task

Force was established to increase the number of "minority" members in AEA and the number of "minorities" who participated in the annual meeting. It was expected to "look at ways in which the needs of minority members could better be served in the organization, particularly in terms of increased participation in the annual meeting" (Covert, 1987a, p. 101). This and other proximate and potentially related meso-level, institutional discursive events that took place within AEA are shown in Figure 20 and Figure 21 in relation to the trajectory for variations of "minority"—compared with "of color" and "white" in U.S. literature as a whole and AEA-affiliated literature, respectively. The most salient of these events are discussed below.

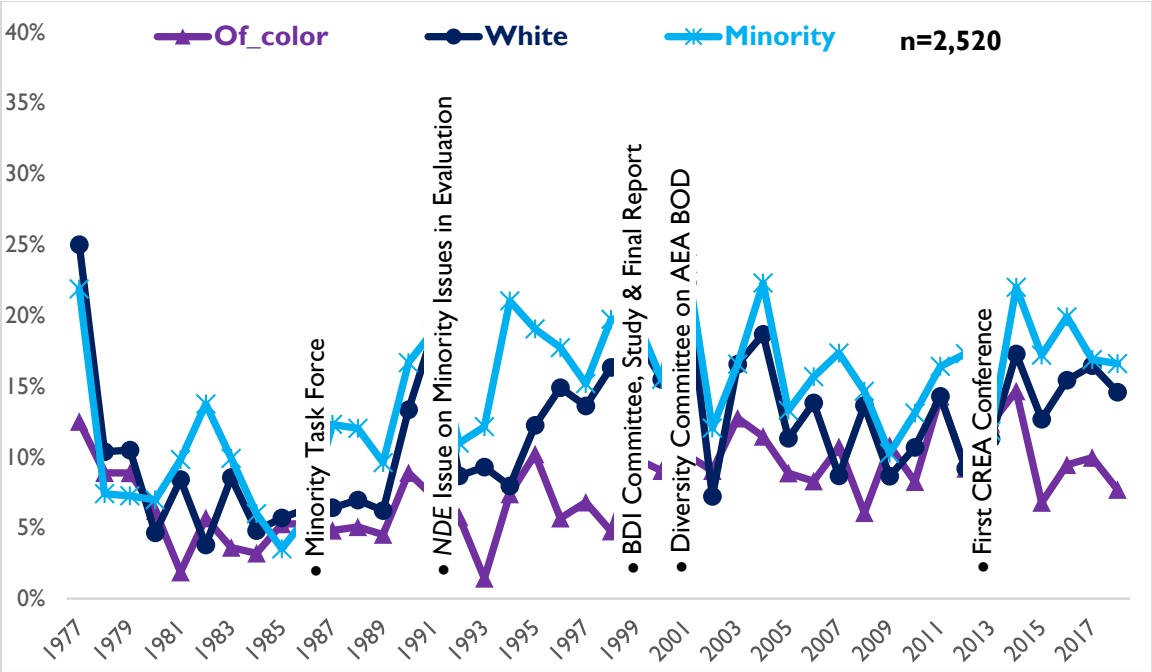

Figure 20. Meso-level, institutional discursive events in relation to annual frequencies for variations of "minority," "white," and "of color" in U.S. evaluation publications

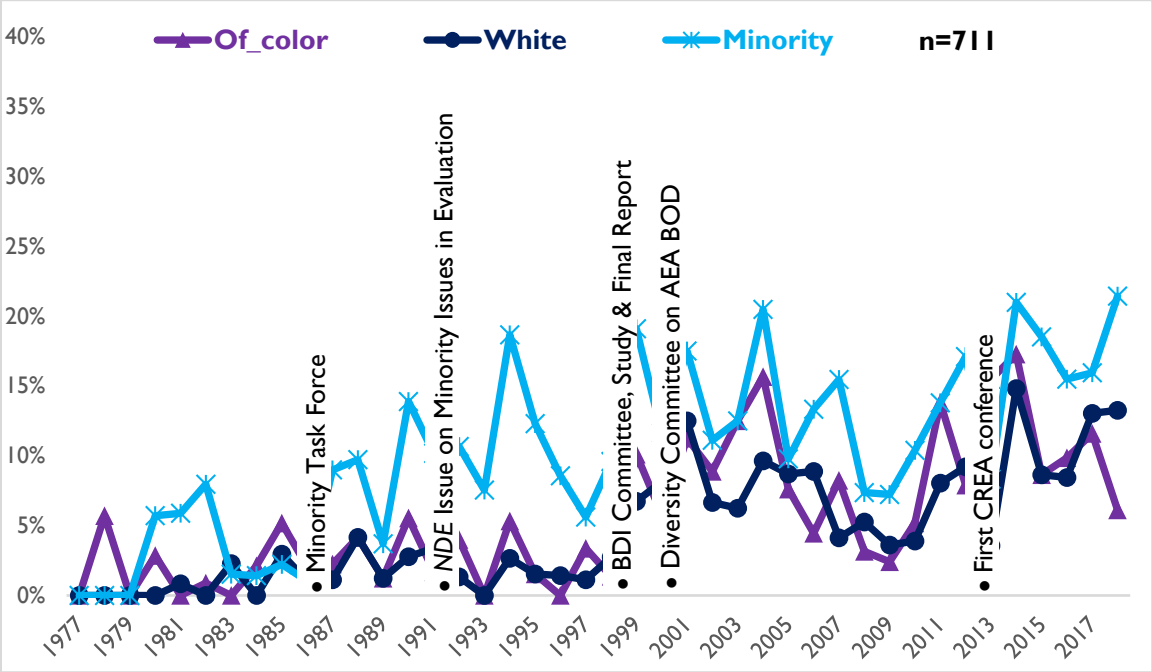

Figure 21. Meso-level, institutional discursive events in relation to annual frequencies for variations of "minority," "white," and "of color" in U.S. evaluation publications affiliated with AEA

Establishment of the Minority Issues in Evaluation TIG and AEA-affiliated journal theme. Annual frequencies for variations of "minority" increased immediately following both the establishment of the Minority Issues in Evaluation TIG and publication of a *New Directions for Program Evaluation* edition focused on Minority Issues in Evaluation. Led by Anna-Marie Madison, the Minority Task Force eventually became the Minority Issues in Evaluation (MIE) Topical Interest Group (TIG), which focused on recruiting racially otherized members and mentoring them once they had joined AEA (King, Nielsen, & Colby, 2004). In 2004, the TIG changed its name to the Multi-ethnic Issues in Evaluation (also MIE) TIG, suggesting a shift away from the term "minority," although the term has still not entirely fallen out of favor in the evaluation literature. Indeed, annual frequencies for variations of "minority" increased slightly (from 20% to 21%) between 2004 and 2018 in AEA-affiliated publications. Additionally, the term remained part of the MIE TIG's mission as of 2018: "to raise the level of discourse on the role of people of color in the improvement of the theory, practice, and methods of

evaluation and increase the participation of members of racial and ethnic minority groups in the evaluation profession" (AEA, 2018c).

In 1992, Minority Task Force leader and subsequent Minority Issues in Evaluation (now the Multi-ethnic Issues in Evaluation) TIG founder and president Anna-Marie Madison assembled a *New Directions for Program Evaluation* edition focused on minority issues in program evaluation. Somewhat like Covert's presidential column and address, the issue's use of "minority" and its variations, which were typical at that time, belay the extent to which Madison questioned power in framing this first compendium that examined how evaluation as an enterprise affects racially otherized groups (Madison, 1992a).

In the Editor's Notes, Madison interrogated the structures surrounding evaluation, first by naming "minority-majority social and economic disparities" (p. 3). Second, she named the larger purpose of evaluated programs as being redistributive and "racial and ethnic minorities and poor people" (p. 1) as disproportionately among the intended beneficiaries of redistributive policy. Third, Madison portrayed racially otherized groups participating in social programs not just as passive consumers of services, but as potential producers of knowledge and agents of change and problematized the limited opportunities for them to exercise that potential within current structures in and around evaluation. According to Madison, although they "have the highest stake in the attainment of social policy and program goals, …they have little input into defining social goals or interpreting the impact of social policy on their lives" (1992a, p. 1). Fourth, she questioned whether redistributive policies and programs were substantive or merely symbolic and whether evaluators should judge the aims of the programs and the institutions that design and implement them. Finally, she pointed to the possible self-interest of the field of evaluation to bypass such questions in an attempt to both avoid political consequences and assert itself as a legitimate form of scientific inquiry.

This issue remains one of few within the evaluation literature in which the disproportionate representation of racially otherized groups in evaluated programs is explicitly tied to macro-level structures—both the "social and economic inequities" (p. 1)

as well as the redistributive policies that undergird the programs that evaluators study. Having framed the issue as such, Madison (1992a) was able to call particular attention to the incongruence between program participants and program evaluators without racially homogenizing either group and to contextualize rather than naturalize the power differential between them:

> It is generally acknowledged that evaluators and the minority populations that are overrepresented as beneficiaries of social policy are, for the most part, dichotomous groups. These groups differ culturally, socially, and economically, yet *people affected by the programs evaluated depend on evaluators* to validate the degree to which desired social outcomes are achieved. In this relationship, *evaluators are in a position of power* in that their critical judgments and interpretations of the truth about the relative quality of social programs may have *serious consequences for the lives of minorities*…. [E]valuators should be concerned about whether issues of social justice are addressed in social policy and programs. Such issues are those that are fundamentally embedded in the *majority-minority imbalance in the distribution of power and resources and in the social outcomes of this imbalance.* (p. 1)

While the dichotomy and differences between program participants and program evaluators—also named by Yvonna Lincoln (1994) and Ernie House (1990, 1995a)—might seem overstated nearly thirty years later, and while the extent to which program evaluators depended on program participants and redistributive social programs in general for their livelihood might be understated in the paragraph above, Madison was careful throughout the Editor's Notes (1992a) to consistently ground the relations between program participants—who disproportionately represent otherized groups—and program evaluators—whose racial classification she did not prioritize as much as their social location and corresponding worldview—in asymmetric socio-economic structures. This nuanced treatment of the incongruence challenged prevailing discourses that conflated racial classification, socio-economic class status, and human service program participation.

Likewise, while Madison described society as culturally diverse and noted cultural differences between program participants and program evaluators, her language differed from Merryfield and Hilliard (both of whom also tied culture to macro-level systems of super-ordination) in that throughout the Editor's Notes, she never described

the program evaluation that took place in such contexts as "cross-cultural." Rather, she characterized it as *cultural dominance*—which conveyed an asymmetric dynamic akin to Young's use of "cultural imperialism" to describe one face of oppression (Young, 2011). Cultural dominance similarly contrasts sharply with the lateral relations more commonly conveyed by phrases like "cultural diversity" and "multiculturalism":

> The aim of this volume is to begin discussion of some minority concerns about the impact of cultural dominance on definitions of social goals and on the measurement of their outcomes in a culturally diverse society, and about the political consequences for minorities of cultural dominance in the selection of evaluation methods. (Madison, 1992a, p. 2)

Madison thus unequivocally characterized the "culturally diverse society" as hierarchically organized or stratified.

The second two periods during which annual frequencies for variations of the term "minority" increased slightly relative to the phrase "of color" were at the turn from the 20th to the 21st century, during which several discursive events addressing racialized difference occurred. These are discussed in depth with respect to the finding regarding diversity and inclusion (below). Another increase took place in 2013, coinciding with the first CREA conference.

The increase in annual frequencies for variations of "diversity" and "inclusion" as potential signifiers of racialized difference between approximately 2001 and 2017 and the decline between 2017 and 2018 in U.S. evaluation literature coincided with several meso-level, institutional discursive events within the field of evaluation that addressed racialized difference. Discursive events in AEA that occurred during the periods of greatest change in the annual frequencies for variations of "diversity" and "inclusion" relative to that for variations of nearly all other potential signifiers of racialized difference between approximately 2001 and 2017—especially between 1999 and 2007 in literature affiliated with AEA—are illustrated in Figure 22 and Figure 23 and described below.

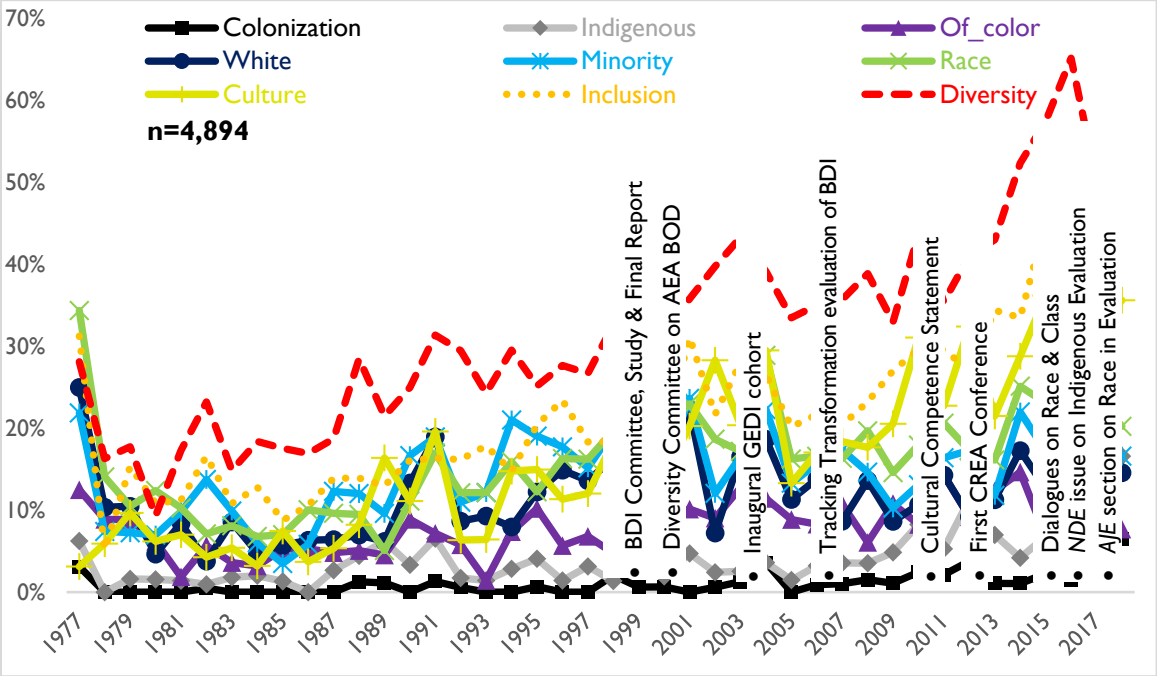

Figure 22. Meso-level, institutional discursive events in relation to annual frequencies for variations of "diversity" and "inclusion" in U.S. evaluation publications

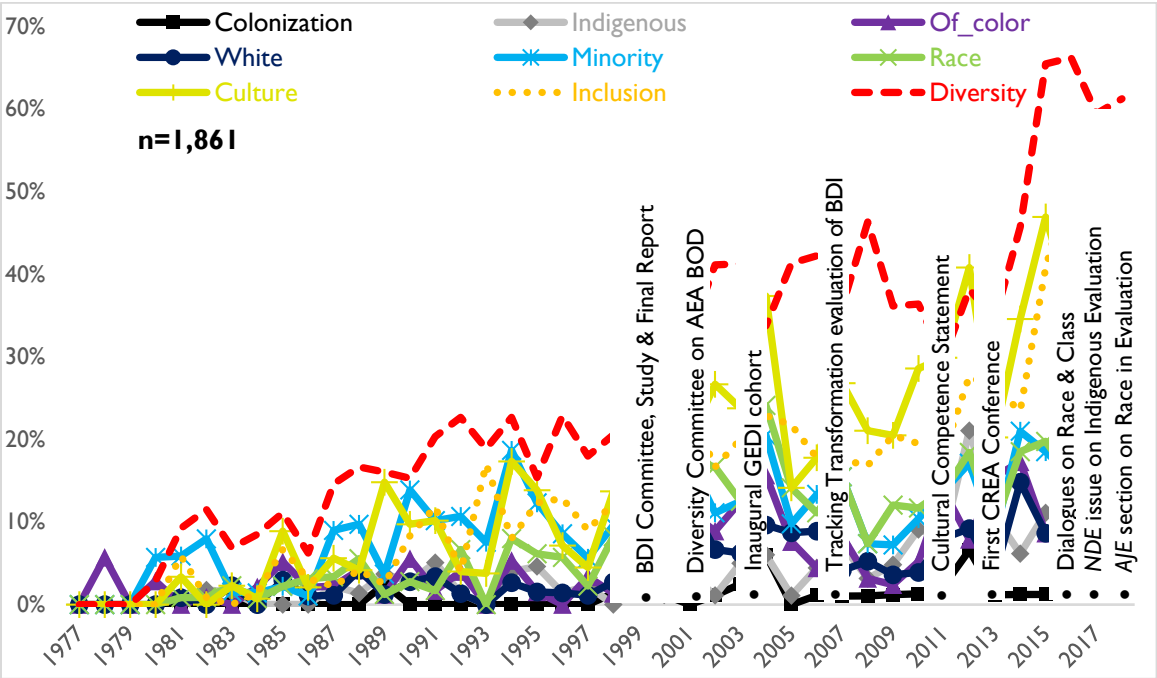

Figure 23. Meso-level, institutional discursive events in relation to annual frequencies for variations of "diversity" and "inclusion" in U.S. evaluation publications affiliated with AEA

Building Diversity Initiative study and implementation. According to the Building Diversity Initiative (BDI) report prepared by the Association for the Study and Development of Community (now Community Science), the BDI was an effort by AEA and the W. K. Kellogg Foundation "to address the complexity of needs and expectations concerning evaluators working across cultures and in diverse communities" (2002, p. 1). The purpose of the Initiative was "to improve the quality and effectiveness of evaluation by increasing the number of racially and ethnically diverse evaluators in the evaluation profession, and…to improve the capacity of all evaluators to work across cultures" (p. 1). The contraposition of "racially and ethnically diverse evaluators" and "all evaluators" in this account of BDI's purpose immediately reveals the connotation of "diverse" not as describing a group of evaluators characterized by internal differences. Rather, "diverse" was used to describe the racially otherized evaluators ("racially and ethnically diverse evaluators"), signifying them as different *from* an implicit standard. The BDI report specifies clearly that while it recognized and respected varying types of diversity,

> … [the] Building Diversity Initiative was developed in response to a request from the W. K. Kellogg Foundation. The foundation is committed to working in a way that is sensitive to diverse communities and has made an effort to identify diverse evaluators to work on their community building initiatives. Representatives of the foundation approached members of AEA to draft a proposal to address perceived *disparities in the evaluation field as they relate to ethnically diverse professionals*. Thus, the Building Diversity Initiative was conceived and developed in response to the expressed interests of the Foundation. The Initiative has principally worked to *understand and address diversity issues as they relate to ethnicity* (p. 11, emphasis added).

The initiative explored issues in four areas:

- Pipeline issues, barriers, and potential strategies to increase the number of people of color in middle school through graduate school who are exposed to and engaged in evaluation, including people who did evaluation work without identifying themselves as evaluators
- Professional development programs or strategies that would provide evaluators of color and evaluators who worked cross-culturally with targeted training opportunities

- Work access strategies that would increase access to evaluation opportunities for evaluators of color and evaluators with cross-cultural experience
- Recruitment barriers for evaluators of color and culturally competent evaluators in joining AEA and in the evaluation field

The resulting action plan focused on the institutionalization of racial and ethnic diversity and cultural competence into the policies and practices of AEA. Recommendations were broadly organized into four categories:

- Training
- Public education
- Policies
- Relationship building

Each BDI recommendation included existing AEA resources that could help meet the overall goal; potential partners that could provide additional information, resources, or support; possible first steps to creating change as outlined in the proposed recommendation; and resources needed to move forward on the recommendation. AEA's Board took initial action on six of the 14 recommendations (numbers 1, 4, 6, 8, 11, and 14), some of which are depicted above in Figure 22 and Figure 23 or below in Figure 27 and Figure 28: the Diversity Committee on AEA's board, the inaugural GEDI cohort, the Cultural Reading and subsequent revision of the Program Evaluation Standards, and the Statement on Cultural Competence. The BDI Task Force then transferred responsibility for the remaining recommendations, and AEA's Diversity Committee took responsibility for monitoring BDI efforts. Figure 24 lists the 14 recommendations (directly from the BDI report) by level of intervention and expected immediate outcome, according to the ecological model of the systems surrounding evaluation portrayed in Figure 3 in Chapter Two.

BDI Recommendation	Level of	
	Intervention	Immediate Outcome
1. Create a graduate education fellowship program targeted to students of color.	Individual	Interpersonal*
2. Tap into existing educational pipeline programs to expose students of color to evaluation as a career choice.	Institutional	Individual
3. Work with historically Black colleges and universities (HBCUs), Hispanic-serving institutions (HSIs), and Tribal institutions (TIs) to (a) increase the profile of evaluation as a profession and (b) support the creation of evaluation training courses and programs.	Individual*	Interpersonal & Institutional
4. Create "guaranteed" training sessions at the annual AEA conference to address the professional development needs of evaluators of color and cross-cultural evaluators.	Institutional	Individual
5. Create nontraditional training opportunities for people doing evaluation work but who do not identify themselves as evaluators.	Institutional	Individual
6. Organize small business development training for evaluators of color who want to start evaluation-consulting firms.	Institutional	Institutional
7. Provide financial incentives for evaluators of color and all cross-cultural evaluators to participate in training and professional development.	Individual	Individual
8. Create a Council of Evaluation Training Programs (CETP) to serve as a forum to discuss issues of diversity and cultural competence as they relate to training and evaluation.	Structural	Institutional
9. Create and promote a "What Is Evaluation?" campaign targeting students and other potential professionals.	Individual	Individual
10. Engage in a public education campaign to emphasize the importance of cultural context and diversity in evaluation for evaluation-seeking institutions.	Institutional	Institutional
11. Incorporate diversity issues into the review of the Program Evaluation Standards.	Institutional	Institutional
12. Advocate for the creation of an affirmative hiring policy (e.g., Small Business Administration Section 8(a) Business Development Program) for foundations and state and local governments.	Structural	Institutional
13. Encourage mentoring for evaluators of color and those seeking cross-cultural evaluation experience and skills.	Interpersonal	Individual
14. Work with diverse organizations to develop a method of publicizing job opportunities to evaluators of color.	Institutional	Individual
* The original recommendation was to offer individual fellowships, but its implementation utilized a cohort model		

Figure 24. BDI Recommendations by level of intervention and expected immediate outcome

Many of the immediately individual- and interpersonal-level outcomes could plausibly be expected to lead long-term to institutional- or structural level outcomes with enough time and critical mass or critical connections (Symonette, Mertens, & Hopson, 2014). Recommendations that could lead long-term to institutional- and structural-level outcomes could potentially shift the culture, organization of labor, and decision-making processes within evaluation to racially otherized groups. Lee and Gilbert (2014) argued that individual level interventions cannot necessarily be expected to lead to sustainable institutional- and structural-level outcomes, however, unless they are supported by corresponding institutional and structural-level interventions—whether those recommended by BDI or others.

Most of these discursive events can be better understood as processes and decisions—which serve as institutional and structural mechanisms—rather than incidents. For example, action on the AEA/Duquesne University Internship started in 2002, at which time Hazel Symonette (co-chair of the MIE TIG), Rodney Hopson (former chair of the MIE TIG), Donna Mertens (former president of AEA), Kien Lee (author of the BDI report), Prisca Collins (inaugural coordinator), Teri Behrens (director of evaluation at the Kellogg Foundation), and AEA's board of directors engaged in decisions regarding sites, funding, and staffing. The internship was renamed the Graduate Education Diversity Internship or GEDI in 2009, and the Graduate Education Diversity Internship TIG was established in 2017. Each discursive event involved multiple types and iterations of communication within AEA, within the evaluation field more broadly, and—especially in the case of efforts that were intended to recruit members and build pipelines—outside the field of evaluation. Also—importantly from the perspective of enabling conceptualizations of justice—the recommendations did not necessarily distinguish among current and prospective "evaluators of color" as opposed to current and prospective evaluators who worked "cross-culturally," although the interests of each of these groups, potentially representing different social groups, might differ.

The rise in frequencies for variations of "diversity" and "inclusion" in the U.S. peer-reviewed literature continued long after BDI—regardless of AEA affiliation—

through 2016, at which time they began to decline, at least in the U.S. literature more broadly. It is possible that this decline was related to changes in national discourse around difference, which had become more explicit, since the 2016 election campaign. It is also possible that this decline was related to the Dialogues on Race and Class (which were potentially also related to the campaign) initiated under—and limited to the tenure of—AEA President Newcomer; the *American Journal of Evaluation*'s special section on race (Bledsoe, 2018); or *New Directions for Evaluation*'s issue on Indigenous Evaluation (Cram, Tibbets, & LaFrance, 2018)—all of which explored difference with greater levels of analytical specificity than variations of "diversity" and "inclusion" offered.

In U.S. evaluation literature, regardless of AEA affiliation, the largest increases in annual frequencies for variations of "indigeneity" and "colonization," and decreases in the gap between them, coincided with meso-level, institutional discursive events that addressed racialized difference within the field of evaluation. Especially for journals affiliated with AEA, increases in annual frequencies for "indigeneity" and "colonization" took place between 1987 and 1991, 1998 and 2002, and especially 2003 and 2005 as well as 2012 and 2018. Only those discursive events most directly tied to indigenous peoples are shown in Figure 22 and Figure 23 and discussed in the section below; the others are addressed under discussions of minority, culture and race, and diversity and inclusion. Discursive events that took place between 1987 and 1991 included Covert and Conner's work with respect to recruiting racially otherized members and Hilliard's keynote speech, all of which were discussed earlier. They also included those that took place between 1998 and 2002; Stanfield's keynote, the BDI, and establishment of the Diversity Task Force on AEA's board are discussed under the finding regarding diversity and inclusion. Finally, the increases that took place between 2003 and 2005 and at 2012 coincided with many of the discursive events discussed under the finding regarding culture and race.

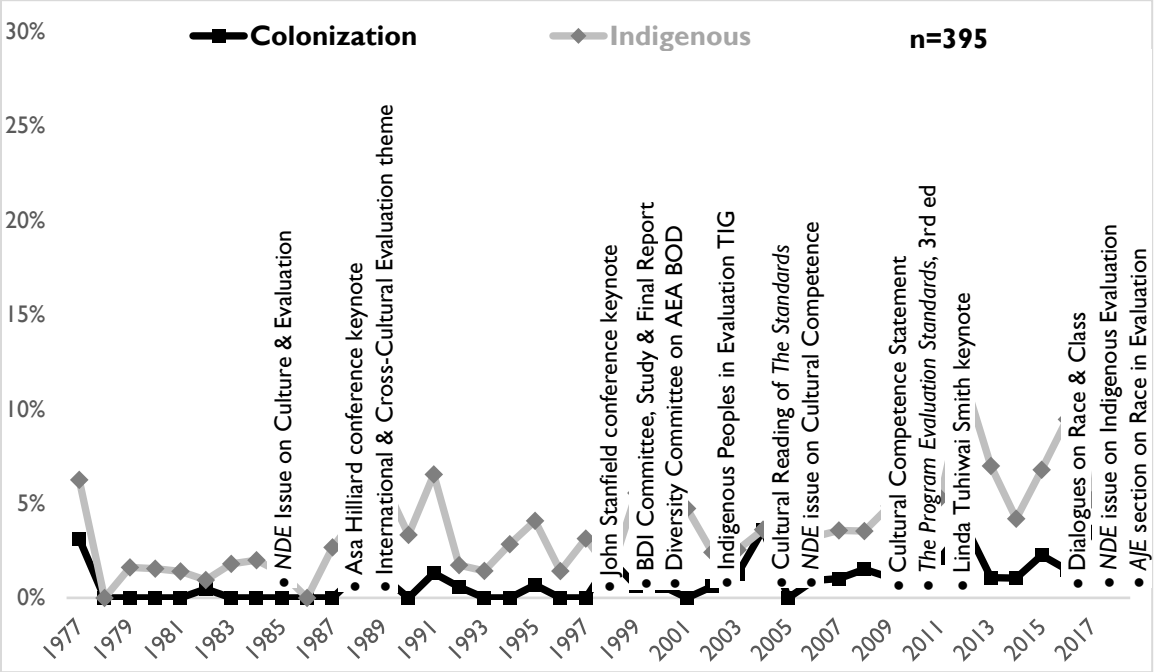

Figure 25. Meso-level, institutional discursive events in relation to annual frequencies for variations of "indigenous" and "colonization" in U.S. evaluation publications

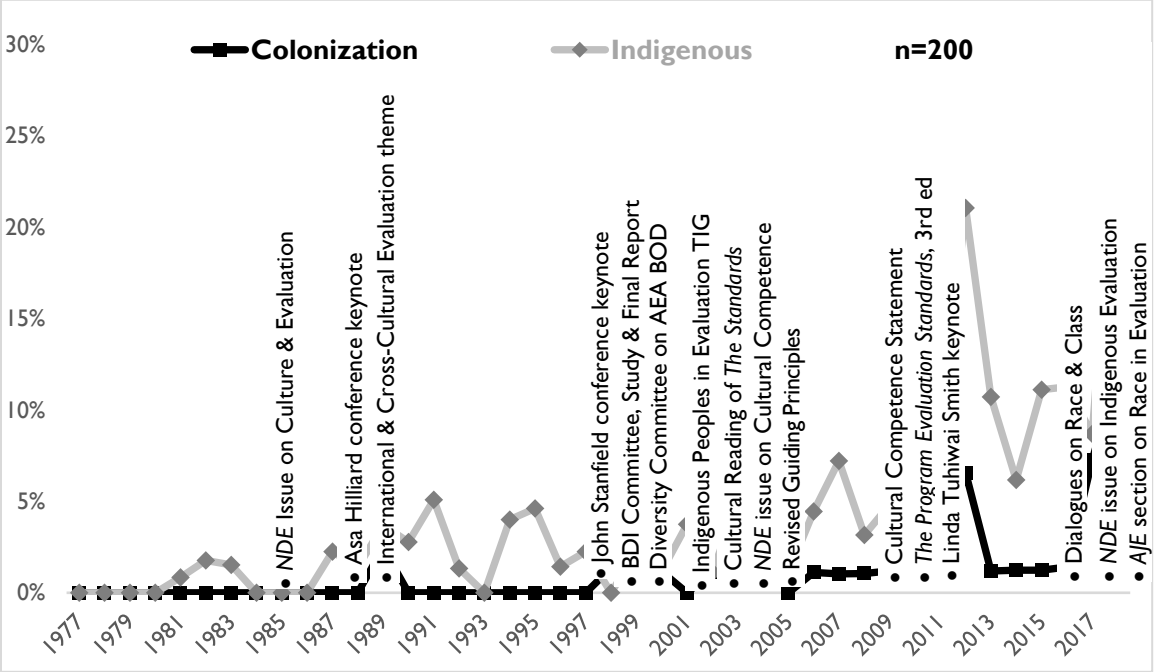

Figure 26. Meso-level, institutional discursive events in relation to annual frequencies for variations of "indigenous" and "colonization" in U.S. evaluation publications affiliated with AEA

Establishment of the Indigenous Peoples in Evaluation TIG. In 2003, a small group of AEA members who identified as indigenous conceptualized a topical interest group (TIG) focused on issues of relevance to indigenous peoples in evaluation. Indigenous members had until then generally joined what was at the time still called the Minority Issues in Evaluation (MIE) TIG.[50] As the number of indigenous members of AEA grew, they held some special meetings to determine their goals within AEA, which they concluded were somewhat distinct from those of the MIE TIG.

> For some in the MIE TIG, this was not welcome news as they thought by forming another "minority" TIG, we would weaken the voice of minority people in the organization. Of course, there were also MIE TIG members who welcomed the idea (LaFrance, personal communication, 4/19/2015).

The Indigenous Peoples in Evaluation (IPE) TIG officially formed in 2005. Its goals involve: (1) developing and disseminating knowledge to assure that evaluations in which indigenous people are among the major stakeholders are culturally responsive and respectful of their interests and rights; (2) creating a venue for indigenous evaluators and others working in indigenous contexts to participate in discourse about evaluation models and methods that support indigenous values, practices, and ways of knowing; (3) and mentoring emerging evaluators interested in evaluation in various indigenous contexts.

Smith keynote at AEA's annual conference. In 2012, under Rodney Hopson's presidency, Linda Tuhiwai Smith was featured as a keynote speaker at AEA's annual conference. Her talk was entitled "Taking a Walk on the Wild Side: Some Indigenous Perspectives on Valuing Complexity, Sustaining Relationships, Being Accountable for Responsibilities and Making Things Relevant." Addressing what it meant to be indigenous, Smith described the rich heritage indigenous peoples have of being descended from a land's first inhabitants as having also been rocked by experiences of colonization, dispossession, and marginalization:

> Both diverse and unequal…cultural, social, and economic inequities mark the "gap" that has to be negotiated…indigenous peoples are constantly vulnerable to

[50] Personal communication with LaFrance, founding member and past president of the Indigenous Peoples in Evaluation TIG, 4/19/15

> the attitudes, perceptions, judgements [*sic*], and moral panic of those in power. (AEA, 2012)

Smith explicitly invited evaluators to be part of the *movement* to decolonize evaluation so that it works *for* and *with* indigenous peoples in transformative ways and noted that her messages might also resonate in other evaluation contexts where the goal is social justice and transformation. Through her use of the word "movement," Smith's invitation corresponded with enabling conceptualizations of justice and efforts toward self-determination.

Also in 2012, Katie Johnston-Goodstar contributed to evaluation's understanding of decolonization through her piece about the necessity of evaluation advisory groups in indigenous evaluation, which was discussed in the textual analysis above. In addition to contextualizing evaluation in relation to research and historicizing both within the structural asymmetry of colonization, she expanded on decolonizing evaluation as distinct from but related to evaluation that honors or serves sovereignty. As much as evaluation can be used in indigenous efforts to maintain sovereignty, she wrote, evaluation as an enterprise must be decolonized.

AEA-affiliated journal theme. In 2018, Fiona Cram, Katherine Tibbetts, and Joan LaFrance edited a *New Directions for Evaluation* edition focused on Indigenous Evaluation that consisted of pieces informed by first-person experience representing multiple continents. The 2018 annual frequencies for both indigenous and colonization increased dramatically from the previous year in the U.S. literature, whether affiliated with AEA or not. In the former, the annual frequency for variations of "indigeneity" increased from 9% to 27%, and that for those of "colonization" increased from 7% to 12%. In U.S. journals more broadly, the annual frequency for variations of "indigeneity" increased from 5% to 17%, and that for those of "colonization" increased from 3% to 6%. These increases may be temporary. While both previous increases in the annual frequencies for variations of "indigeneity" raised annual frequencies, less dramatically, for years thereafter, both previous increases in the annual frequencies for variations of "colonization" were very temporary.

While the articles in the journal edition were excluded from the textual analysis, which included items published only through 2017, the Editor's Notes from this issue explained the relationship between indigeneity and colonization:

> Academic writings about cultural responsiveness and Indigenous peoples can overlook the history of colonization experienced by many Indigenous peoples, as well as ignoring indigenous sovereignty and the challenges of racism (Castagno & Brayboy, 2008; Crazy Bull, 1997). However, since the publication of *Decolonizing Methodologies: Research and Indigenous Peoples* by Linda Smith (2012), Indigenous peoples have been more hopeful about being able to express both their concerns about and aspirations for research and evaluation in ways that others will understand. Integral to these expressions are talk of colonization and decolonization, the relief of being able to speak about past grievances and future aspirations, and the joy of being able to start a conversation that begins at a different place than before; that is, a conversation that sets Linda Smith's seminal work as the backstop and reference point for anyone wanting to know the "back-story." (Cram, Tibbets, & LaFrance, 2018, p. 9)

As pivotal as colonization and decolonization were to the indigenous experience, however, neither defined indigeneity—particularly when authors increasingly represented indigenous groups. The refusal to be defined by colonization might help explain the decoupling of indigeneity and colonization that appeared to continue despite the increases in first-person writing within evaluation that addressed indigeneity.

> "Who is Indigenous?" There are countless definitions that purport to define Indigenous peoples or the state of being Indigenous. However, many of these definitions inexorably link being Indigenous to experiences of colonization, minoritization, and marginalization. Yet decolonization will not make Indigenous peoples any less Indigenous. Nor are those Indigenous people who have had no or only a partial colonial experience more or less authentically Indigenous than others (Smith, 2012 as cited in Cram, Tibbets, & LaFrance, 2018). (p. 7)

The discursive attention that AEA directed to culture as opposed to race coincided with respective shifts in the annual frequencies for variations of those terms in the literature beyond journals affiliated with AEA. Again, annual frequencies for variations of "race" had been consistently higher than those for variations of "culture" until 1988, at which time the two fluctuated until 2008, when variations of "culture" surpassed those of "race." Discursive events in AEA that occurred during the periods of

greatest variance and change in the U.S. literature as a whole and in AEA-affiliated journals, specifically, are illustrated in Figure 27 and Figure 28 and described below.

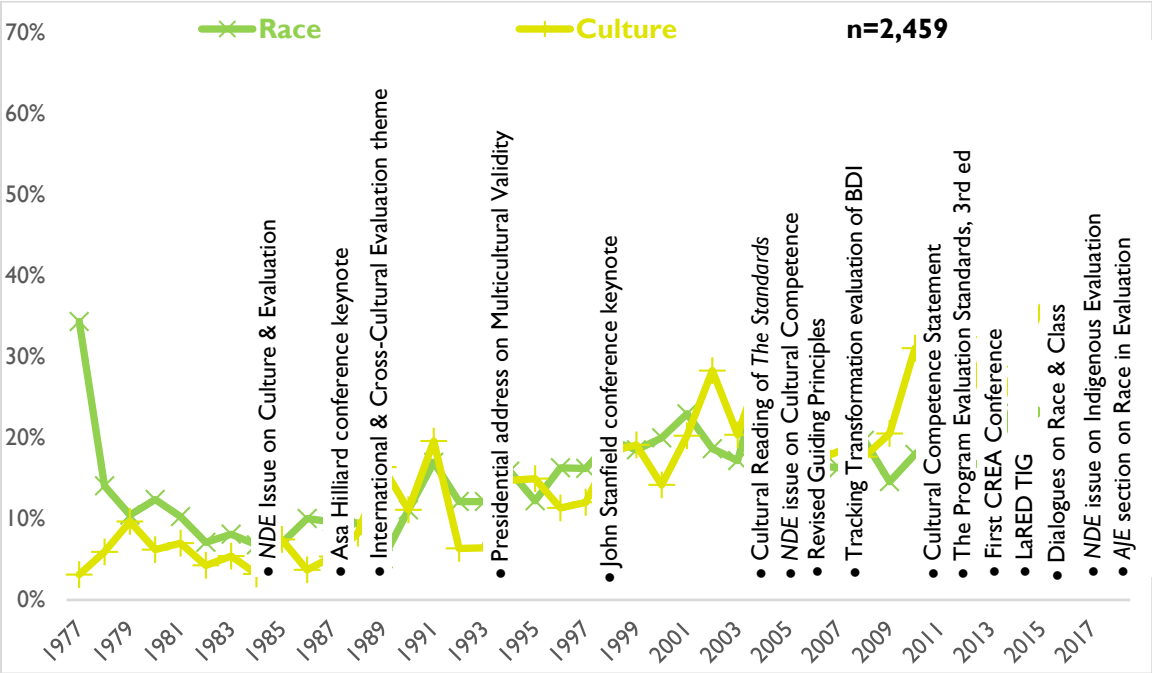

Figure 27. Meso-level, institutional discursive events in relation to annual frequencies for variations of "culture" and "race" in U.S. evaluation publications

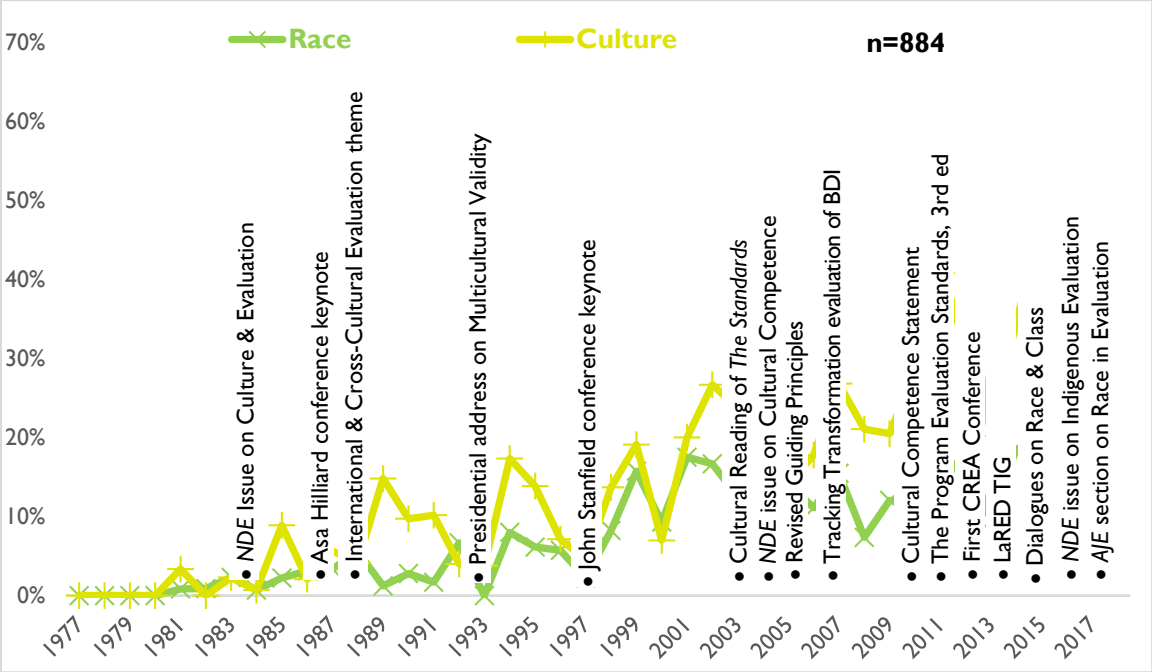

Figure 28. Meso-level, institutional discursive events in relation to annual frequencies for variations of "culture" and "race" in U.S. evaluation publications affiliated with AEA

AEA-affiliated journal theme, AEA presidential attention, and Hilliard keynote at AEA's annual conference. During the 1980s, three of AEA's earliest presidents each addressed racialized difference directly through their role as president and through their leadership in developing the annual conference. In 1984, Michael Quinn Patton issued a call for submissions for a *New Directions for Program Evaluation* edition that would be devoted to culture and evaluation. This call was the first documented focus on culture within the literature, and the bulk of the resulting publication focused on evaluation in international contexts (Patton, 1985). Aside from Merryfield's piece (discussed in the textual analysis), which was USA-facing, titles in the issue referred to Canada, the Caribbean, Egypt, Israel, and the Netherlands.

Soon after the publication, Bob Covert and Ross Conner raised concerns about the representation of what they referred to as racial minorities and people of color, respectively, through their presidencies (Covert, 1987a; Conner, 1989a, 1989b, 1989c). In other words, both approached the topic of minority representation and diversity from a

distinctly racialized perspective (although both also approached it from the perspective of gender, which they distinguished from racialized difference). The textual evidence for these concerns was discussed at greater length in the section on textual analysis.

Between Covert's and Connor's presidencies, Michael Quinn Patton served as AEA president. While he did not necessarily address racialized difference directly through his own presidential address or columns, he did invite Asa Hilliard as a keynote speaker at AEA's annual conference in 1988. Hilliard's speech, the transcript of which was printed in 1989 and discussed at greater length in the textual analysis, linked the often-amorphous notion of culture with racialized difference through a detailed explication of white supremacy as a system of oppression. The following year, Ross Conner selected the theme of International & Cross-Cultural Evaluation for AEA's annual conference.[51] The transcript of his address, also discussed at greater length in the findings from the textual analysis, was printed in 1990. Thereafter, the annual frequencies for variations of culture surpassed race until 1992.

In 1992, the *New Directions for Evaluation* edition focused on Minority Issues in Evaluation was published, as described earlier with respect to the increase in variations of the term "minority" (Madison, 1992a). However, articles within the issue, some of which were included in the textual analysis, approached minority issues with explicit—and critical—attention to racialized difference. Annual frequencies for variations of the term "race" were higher than those for variations of the term "culture" until 1994. In 1994, Karen Kirkhart introduced the term "multicultural validity" into the evaluation lexicon through her presidential address at AEA's annual conference, the theme for which was Social Justice. Annual frequencies for variations of the term "culture" stayed higher than those of "race" until 1999. At that time, they were temporarily surpassed by those of "race."

---

[51]An earthquake in San Francisco the day before the conference began in 1989 led to the unavoidable cancellation of the conference where Conner's address would have been presented at a plenary session. The small group who had made it to San Francisco before the earthquake did present their papers to each other, but the impact of the presidential address was reduced (personal communication with King, AEA 1989 Annual Conference Presidential Address, 4/26/17). As is customary, the address was later published in the *American Journal of Evaluation* (Conner, 1990).

Stanfield keynote at AEA's annual conference. In 1998, Donna Mertens served as president of AEA and invited John Stanfield to deliver a keynote at AEA's annual conference, the theme of which was Transforming Society through Evaluation. The transcript of Stanfield's speech, printed in 1999, was discussed at greater length in the findings from the textual analysis. Much like Hilliard's keynote a decade prior, Stanfield related terms like "minority," "people of color," and "culture" to "race" and "white" supremacy in ways that did not rise to the surface again until 2017. However, annual frequencies for variations of the term "race" remained comparable to the annual frequencies for variations of "culture" through 2008.

During the ten-year period between 1998 and 2008, several discursive events took place within AEA that addressed racialized difference. For example, AEA launched the BDI study in 1999 and its draft report was issued in 2001. The 14 recommendations in the report can broadly be divided into those pertaining to the cultural competence of all evaluators and those pertaining to increasing the pool of racially otherized evaluators. Discursive events most salient to notions of race and culture are discussed below. The BDI report itself was discussed under the finding regarding diversity and inclusion.

AEA-affiliated journal theme, the Guiding Principles, and the Cultural Reading of the Standards. In 2004, three discursive events took place that signified racialized difference using variations of the term "culture." *New Directions for Evaluation* published an issue on Cultural Competence in Evaluation (SenGupta, Hopson, & Thompson-Robinson, 2004). Additionally, the Diversity Committee of AEA's board of directors (discussed with respect to the finding on diversity and inclusion) launched a Cultural Reading of The Program Evaluation Standards. Finally, AEA revised its Guiding Principles for the first time—producing the second edition.[52]

The *New Directions for Evaluation* edition devoted to cultural competence was noteworthy in that in it, SenGupta, Hopson, and Thompson-Robinson equated notions of culture with geographic region by distinguishing "Eastern" and "Western," onto which they further conflated economic system ("agricultural" and "industrial") and time

[52] The Guiding Principles would subsequently be revised again in 2013 and 2018.

("traditional" and "modern"). The Preface of the 2004 Guiding Principles in Evaluation similarly contextualized the principles in terms of culture—conflating it with geographic region and nationality—by stating that the principles

> … were developed in the context of Western cultures, particularly the United States, and so may reflect the experiences of that context. The relevance of these principles may vary across other cultures, and across subcultures within the United States. (AEA, 2004e)

Under the principle of competence, the Principles stated that "[t]o ensure recognition, accurate interpretation, and respect for diversity, evaluators should ensure that the members of the evaluation team collectively demonstrate cultural competence" (AEA, 2004c). Fourteen years later, in the 2018 edition of the Guiding Principles, this principle and others (quoted at length in Chapter Two) were modified in ways that attended for the first time to structural oppression.

Not unlike the Guiding Principles, the Program Evaluation Standards have undergone revisions. Toward the end of AEA's development and ratification by membership of the 2004 edition of the Guiding Principles, the second edition of The Program Evaluation Standards (hereafter, The Standards) was subjected to a Cultural Reading (AEA, 2004a) initiated by AEA Diversity Committee member Karen Kirkhart. During this process, the results of which were shared with the AEA membership in 2004, a designated "group of professional evaluators who share expertise in and concern for issues of cultural diversity and cultural context in evaluation" (AEA, 2004a) reviewed the Standards "with respect to coverage of cultural diversity, treatment of cultural concerns, and attention to cultural competence" (AEA, 2004a). These three discursive events—the 2004 AEA-affiliated journal edition devoted to cultural competence, the 2004 revision of the Guiding Principles, and the 2004 Cultural Reading of the Program Evaluation Standards—reflected the discursive strategies most commonly associated with use of variations of the term "culture" in the U.S. field of evaluation, wherein it has typically been decoupled from notions of hegemony, imperialism, and dominance.

Tracking Transformation. In 2005, the Diversity Committee recommended that AEA take stock of implementation of the BDI action plan. AEA contracted Geri Lynn

Peak to conduct an evaluation of Phase I of AEA's Building Diversity Initiative entitled "Tracking Transformation" between April and October, 2007. Tracking Transformation is noteworthy because it named the two ways in which stakeholders interviewed as part of the evaluation talked about the BDI, retrospectively, as "the Excellence Imperative" and "the Justice Imperative," which correspond with the two strains of literature recognizing racialized difference in evaluation that were described in Chapters One and Two: validity and justice. They also correspond with the two categories of participatory evaluation: practical and transformative (King, Cousins, & Whitmore, 2007).

The report authors defined the excellence imperative as "raising the standard of excellence in evaluation practice" and reflecting "individual and collective desires to build upon, broaden and strengthen—through responsiveness and connection—our theories, methods, approaches and practices with wisdom and excellence" (Peak, Peters, & Fishman, 2007, p. 4). BDI aimed to

> … increase attention to multiculturalism. In retrospect, this went beyond addressing issues of cultural competence and appropriateness of evaluations and evaluators to address multiculturalism in all aspects of evaluation. (p. 4)

Peak, Peters, and Fishman described these aspects of evaluation as encompassing theory building and teaching and training as well as who conducts evaluations and subsequently has access to assessing, judging, and assigning value to the efforts.

They defined the justice imperative as "promoting justice and equity in evaluation practice" (p. 4) Specifically, BDI aimed to

> … increase involvement of people of color at all levels of association involvement and evaluation practice, working to produce more evaluators of color, attract more AEA members of color and, as the initiative progressed, expand the definition of diversity beyond color to reflect more and more of the diversity existent in the human family. (Peak, Peters, & Fishman, 2007, p. 4)

They elaborated on the justice imperative as motivated by "individual and collective commitment to justice (or, social justice) of equity, fairness, respect, inclusiveness and acceptance, promoting unity among evaluators while *redressing past harms*" (p. 4,

emphasis added). As in Covert's columns and speeches two decades prior, this involved increasing awareness and access to AEA's evaluation training resources and supports.

Tracking Transformation was significant in that it affirmed multiple identities without flattening asymmetries and hierarchies. Moreover, like Anna Madison's work in 1992, Asa Hilliard's speech in 1988, John Stanfield's speech in 1998, and the work of several indigenous scholars and practitioners, it tied "culture" to macro-level systems of oppression and to notions of redress. Proffering an alternative to the phrase "cultural competence," Peak, Peters, and Fishman defined multiculturalism as:

> … concerning the process of recognizing, understanding and appreciating the cultural background of others as well as one's own. It stresses an appreciation of the impact of difference *in social location* based on the variety of demographic characteristics that describe our differences and our similarities, including race/ethnicity, gender, class/level, age, sexual orientation, religion, physical/mental ability, immigration status, language, and military experience. (p. 38, emphasis added)

They further defined a multicultural lens as a practice that "accounts for the impact of differences at the personal, interpersonal, institutional, and cultural levels" (p. 38). Peak, Peters, and Fishman not only described difference in terms of *social location based on demographic characteristics* as opposed to essentializing difference by rooting it in the demographic characteristics themselves. They also broadened the unit of analysis beyond individuals by considering the effects of difference at the multiple, nested levels of analysis associated with the social-ecological model.

Figure 29 summarizes the evaluation results by BDI recommendation (directly from the Tracking Transformation report) and by intervention level as of October 2007. It also parenthetically indicates three that AEA documents suggest may have been fulfilled afterwards.

> The most successful actions fell into two main categories: (1) those aimed at increasing awareness and access to AEA's evaluation training resources and supports by people of color through internships, exploration of evaluation career options and training opportunities for students, new professionals and faculty at Minority Serving Institutions and (2) those aimed at transforming evaluation practice guidelines and standards to promote cultural competency as part of quality evaluation practices. (Peak, Peters, & Fishman, 2007, p. 1)

Recommendations #5 through #14 are shown as largely (aside from #11) unfulfilled as of 2007. The Tracking Transformation results thus showed a pattern in which the recommendations that had been fulfilled tended to involve interventions directed at individuals and institutions, whereas those that remained unfulfilled involved institutional- and structural-level interventions. As such, the pattern of fulfillment corresponds with the individualism characteristic of liberal constructions of difference.

Furthermore, Tracking Transformation documented the presence, albeit not pervasive, of tension between the excellence and justice imperatives—a tension that parallels tension between the discourses of diversity and inclusion as opposed to Affirmative Action and Equal Opportunity, as well as those of culture as opposed to race: "Contrary attitudes regarding the importance of redressing racism and prejudice as part of promoting the justice imperative are not welcome for debate" (Peak, Peters, & Fishman, 2007, p. 29). The authors further noted that

> … [f]or some, the shift towards an excellence imperative threatens progress made or to be made in support of the justice imperative. There is a perceived tension between the two diversity agendas—justice and excellence. This is in part a concern that any focus away from justice will diminish momentum and, in part, it is a cultural dilemma faced by people of color—one that is often generational. Another perceived tension, a concern that promoting diverse perspectives and multiculturalism in the theory and methods of evaluation will diminish rigor, must also be considered. (2007, pp. 29-30)

BDI Recommendation	Level of Intervention	Fulfilled?
1. Create a graduate education fellowship program targeted to students of color.	Individual	✓
2. Tap into existing educational pipeline programs to expose students of color to evaluation as a career choice.	Institutional	✓
3. Work with historically Black colleges and universities (HBCUs), Hispanic-serving institutions (HSIs), and Tribal institutions (TIs) to (a) increase the profile of evaluation as a profession and (b) support the creation of evaluation training courses and programs.	Individual*	✓
4. Create "guaranteed" training sessions at the annual AEA conference to address the professional development needs of evaluators of color and cross-cultural evaluators.	Institutional	✓
5. Create nontraditional training opportunities for people doing evaluation work but who do not identify themselves as evaluators.	Institutional	(✓)
6. Organize small business development training for evaluators of color who want to start evaluation-consulting firms.	Institutional	
7. Provide financial incentives for evaluators of color and all cross-cultural evaluators to participate in training and professional development.	Individual	(✓)
8. Create a Council of Evaluation Training Programs (CETP) to serve as a forum to discuss issues of diversity and cultural competence as they relate to training and evaluation.	Structural	
9. Create and promote a "What Is Evaluation?" campaign targeting students and other potential professionals.	Individual	
10. Engage in a public education campaign to emphasize the importance of cultural context and diversity in evaluation for evaluation-seeking institutions.	Institutional	(✓)
11. Incorporate diversity issues into the review of the Program Evaluation Standards.	Institutional	✓
12. Advocate for the creation of an affirmative hiring policy (e.g., Small Business Administration Section 8(a) Business Development Program) for foundations and state and local governments.	Structural	
13. Encourage mentoring for evaluators of color and those seeking cross-cultural evaluation experience and skills.	Interpersonal	
14. Work with diverse organizations to develop a method of publicizing job opportunities to evaluators of color.	Institutional	
* While unfulfilled as of 2007, Recommendation #10 could be considered fulfilled as of 2011 by AEA membership's approval of the Statement on Cultural Competence, development of which was noted as already in progress in the Tracking Transformation report.		

Figure 29. Fulfillment of BDI recommendations by level of intervention

While the above tension was not necessarily pervasive, the critique was pointed:

> Some feel that the *expansion of interest in diversity and its definition may dilute the original intention to promote involvement and access of people of color to AEA/the evaluation field*. This relates specifically to broader issues, such as *redressing the impacts of slavery or the displacement and usurpation of the lands of indigenous people*—issues that may not seem related to the issues facing AEA, but may be *critical harms experienced by members—the foundation of injustices* upon which new injustice or even achievements rest (p. 31).

Other observations that Peak, Peters, and Fishman documented as present even if not pervasive among those interviewed included a sense that efforts to date lacked intentionality and integration into the body of AEA, that they were not as much as some other associations had done or as some interviewees had seen in their past experience. According to Peak, Peters, and Fishman, "these perceptions reflected differences in color and role"—whether champion or adopter (p. 31). They stated, "[U]ntil efforts to promote diversity are integrated into the life of AEA and as long as people see the work as a project, program or side effort, these efforts will lack needed volition to achieve long-lasting impacts" (p. 30). Moreover:

> Among certain key figures, particularly those engaged in promoting and supporting the BDI, there is a sense that progress has not been as far as could have been, commitments have not been as deep and results, not as far-reaching into the association. *Specific concerns about diluting the justice aspect of the work are prominent* (p. 30, emphasis added).

Tracking Transformation's final recommendations were for AEA to:

1. Adopt a broader commitment to multiculturalism so that all aspects of the association are included.
2. Promote justice and equity by committing to reach out to underrepresented groups and openness and acceptance for all evaluators: Maintain the legacy of reaching out to people of color.
3. Inspire excellence by expanding evaluation theories, methods, applications, and practices using a multicultural lens. Use the mission/goal/value statement, the theory of change and the multiculturalism framework to expand the dialogue on diversity and cultural competence. Continue the AEA tradition of vibrant dialogue on the

justice and excellence imperatives. Continue to bring rigor to the application of a multicultural lens in evaluation practice. Explore traditional and non-traditional venues to fuel the discussions. Use what is learned and share what is learned.

4. Reflect the unity and diversity among the family of evaluators gathered under the auspices of the AEA. Expand understanding of diversity without compromising or minimizing commitments to groups underrepresented due to harmful social processes, such as racism and xenophobia. Continue to promote international solidarity among evaluators and openness to diverse social/cognitive perspectives that can influence how evaluation is thought about and practiced.
5. Inform the membership of these efforts and encourage their full participation in the dialogue and the actions (Peak, Peters, & Fishman, 2007).

Finally, they recommended that BDI infuse multiculturalism throughout the whole AEA system by engaging in the dynamic and deeply reflective work of using lessons learned to feed actions, policies, and practices and that AEA use Phase II to both track what happens and begin pushing forward on the Guidelines and Cultural Competency Statement. However, no Phase II evaluation of BDI was found (or known of by interviewees, discussed in Chapter Seven).

The Program Evaluation Standards. From 2008 through 2018, annual frequencies for variations of the term "culture" remained higher than those of the term "race," with an increase occurring in the former between 2010 and 2012. This increase corresponded to development of the current (third) edition of The Program Evaluation Standards, which took place during the period in which AEA began formally developing and articulating its position with respect to cultural competence through the Statement discussed below. Correspondingly, the JCSEE described the Standard of Propriety as supporting "what is proper, fair, legal, right, acceptable, and just in evaluations" (Yarbrough et al., 2011, p. 106) in terms that attend to culture and inclusion.

Specifically, propriety was described as encompassing a responsive and inclusive orientation and formal agreements that take into account the needs, expectations, and cultural contexts of clients and other stakeholders (Yarbrough et al., 2011). Despite the

opportunity presented by the addition of justice among the descriptors of Propriety—a revision from the second edition, which described the standard in terms of legality, ethics, and due regard for the welfare of those involved in the evaluation (Sanders, JCSEE, & AASA, 1994)—The Program Evaluation Standards invoked culture without referring to structural dynamics,[53] namely, structural oppression (Young, 2011). As such, like the Guiding Principles for Evaluators and the Statement on Cultural Competence, the third edition differed little in this regard from the second edition of the Joint Committee's Program Evaluation Standards. Referring to the second edition (and to the 1994 edition of the Guiding Principles), Davis noted that "the evaluator's role in representing less powerful and higher-positioned voices is not explicitly articulated" (1999, p. 119).[54]

AEA Statement on Cultural Competence. Also between 2010 and 2012, AEA was engaged in communications, vetting, education, and ratification by its membership of the Association's Statement on Cultural Competence (hereafter, the Statement). The Statement was initially drafted by the Cultural Competence in Evaluation Task Force of AEA's Diversity Committee in 2005, reviewed by the AEA Board of Directors, and finally approved by a vote of the full AEA membership six years later—in 2011 (AEA, 2011). The Statement attended to the complexity of culture and noted the power dynamics among cultures: "Cultural groupings are ascribed differential status and power, with some holding privilege that they may not be aware of and some being relegated to the status of 'other'" (AEA, 2011). Importantly, it also listed racialized "others" and whiteness among its examples. It did not, however, acknowledge the asymmetrical structural arrangements that continued to shape the cultural groupings themselves or the patterns in which resources flowed among them. Nor did it name which social groups within the industries immediately surrounding evaluation and in macro-level society benefited from these arrangements at which ones' expense.

---

[53] The content of the Cultural Reading of The Program Evaluation Standards, analyzed in greater depth in Chapter Four, focuses considerable attention on power differentials and dynamics within evaluation practice and in society at large, although not on the role that evaluation as an enterprise plays therein.

[54] Davis (1999) suggested that evaluators look elsewhere, to an ethic of caring, for guidance in that regard.

CREA conference, LaRED TIG, and AEA-affiliated journal theme. Annual frequencies for variations of the term "culture" continued to increase through 2018, coinciding with several discursive events within the U.S. field of evaluation. These included the first CREA conference, which took place in 2013; Lisa Aponte-Soto's 2014 co-founding of the Latinx Responsive Evaluation Discourse (LaRED, which means "the Network" in Spanish) TIG, whose mission is to increase representation, engagement, and leadership of Latinx and other evaluators in the theory, research, and practice of culturally responsive evaluation; and a *New Directions for Evaluation* edition devoted to Indigenous Evaluation, in which culture was tied to sovereignty and decolonizing efforts. Indigeneity and colonization are discussed more specifically in the next section.

Archival Analysis: Interpretation and use of findings. Findings from the archival analysis suggest that several meso-level, institutional decisions and processes within and around AEA coincided with shifts in the relative annual frequencies (identified in the diachronic analysis) for variations of the potential signifiers of racialized difference (identified through the textual analysis). Figure 30 summarizes the findings from these three components of the CDA.

The potentially influential meso-level, institutional events identified through the archival analysis, along with the findings from the diachronic analysis and textual analysis, were used to generate a list of scholars and practitioners who were closest to the construction of racialized difference within evaluation through their role with respect to the meso-level, institutional events, as indicated by their membership on associated AEA committees and authorship of AEA documents. Prioritized among these scholars and practitioners were those who were also active in contributing to the scholarly literature containing potential signifiers of racialized difference.

Conclusion of Chapter Six

The archival analysis focused on the construction of racialized difference *within the institution* of AEA. Periods identified through the textual analysis and diachronic analysis as those during which discursive shifts occurred in the construction of racialized difference guided the search for additional material in AEA-affiliated journals as well as

documentation of AEA decisions and policies published on AEA's website. Additionally, a search for documents containing variations of the nine terms that potentially signify racialized difference identified through the textual analysis was conducted on AEA's website.

Archival analysis of professional association documents suggests that shifts in the field's construction of racialized difference coincided with many meso-level, institutional discursive events within the field's professional association, including themed editions of journals affiliated with AEA (Culture, Minority, Cultural Competence, GEDI, Indigenous); the annual conference theme and keynote speakers; its topical interest groups (MIE, IPE, LaRED); its sponsored programs (GEDI, MSI); and its standards and guidelines for practice (the Standards, Guiding Principles, Statement on Cultural Competence).

1. How has the U.S. scholarly evaluation literature constructed racialized difference?		2. How has that construction changed since the field began formalizing in the early 1970s?	3. How does that trajectory of language relate to the systems surrounding evaluation?
• Minority • Of color • White	I. Consistent use of variations of "minority" relative to "of color" and inconsistent use of variations of "white"	• "Minority" increased in late 1980s • "Minority" increased in early 1990s	• Establishment of Minority Issues in Evaluation TIG led by Madison, under Covert presidency • Publication of AEA-affiliated journal edition on Minority Issues in Evaluation by Madison
• Diversity • Inclusion	II. Rise of variations of "diversity" and "inclusion"	• Both increased in the early 2000s	• Completion of Kellogg-funded BDI study and onset of implementation under Mertens presidency • Anti-Affirmative Action sentiment
• Race • Culture	III. Replacement of variations of "race" with variations of "culture"	• "Culture" increased in late 1980s • Both fluctuated in early 1990s • "Culture" surpassed "race" in late 2000s	• International & Cross-cultural Evaluation theme of annual AEA conference under Conner presidency • Multicultural validity in annual AEA conference address by President Kirkhart • Introduction of multiculturalism by Peak, et al.'s Tracking Transformation evaluation of BDI • Cultural Reading and revised edition of the Standards • AEA membership's appproval of Statement on Cultural Competence
• Indigeneity • Colonization	IV. Rise of and decoupled relationship between variations of "indigeneity" and "colonization"	• Both increased in early 2000s • Both increased in late 2010s	• Establishment of IPE TIG • Publication of AEA-affiliated journal edition on Indigenous Evaluation by Cram, Tibbets & LaFrance

Figure 30. Summary of CDA findings by research question

CHAPTER 7: RESULTS OF INTERVIEW ANALYSIS

How Does the Trajectory of Language Regarding Racialized Difference Relate to the Systems Surrounding Evaluation?

Critical Systems Thinking (CST): Interviews

The CST strand of inquiry was intended to disconfirm, support, and explain or extend the relationship of the trajectory of the construction of racialized difference within evaluation to the systems surrounding it that was preliminarily established through the CDA strand. As such, the CST strand was an attempt to increase understanding of the systemic patterns within and around the field of evaluation in an effort to identify those that reinforce and those that counteract the efforts of racially otherized groups to create institutional and structural changes that serve their interests. Interviews of evaluation scholars and practitioners most closely connected to evaluation's construction of racialized difference—as indicated through the textual analysis, diachronic analysis, and archival analysis—elicited their mental models of the systems surrounding evaluation (Martinez-Moyano & Richardson, 2013) using expanded boundaries, structurally mediated relationships, and social group perspectives or interests as their organizing framework.

Data collection and instrumentation. The CST interviews sought impressions and interpretations of—as well as recollections and storytelling around—visualizations of the preliminary findings from the archival analysis (see Appendix D for interview outline). This was to elicit interviewees' mental models—integral to both CDA and systems thinking—of the systems within which evaluation plays a role (Lane, 2001a; Martinez-Moyano & Richardson, 2013; Thomas & Parsons, 2017).

Sampling and participant recruitment. Constituting a purposive, theoretical sample, the names of 20 scholars and practitioners most closely connected to evaluation's construction of racialized difference since the field's inception (listed in Appendix B) were narrowed from a list of approximately 80 (listed in Appendix A) who had been identified through the archival, diachronic, and textual analysis as having served on relevant

committees or published relevant literature during the years in which the diachronic analysis suggested that discourse shifts took place. They were contacted (see Appendix C) using publicly available information. Thirteen responded; however, one of the earliest to respond later became unresponsive. Interviews ended after the eleventh interview was complete, at which point the researcher determined that data saturation had been reached. Complete anonymity was difficult to achieve considering the small size of the group of evaluation scholars and practitioners serving on committees and publishing during relevant periods. While many interviewees chose to have their names associated with their interviews, discretion was used when attributing quotations to specific interviewees.

Implementation, setting, and instrumentation. All interviewees received electronic copies of preliminary findings from the archival analysis, along with an outline of the interview protocol. All interviews took place by web-interface that allowed for screen sharing or by phone, and they were recorded. They were generally between 60 and 90 minutes. The loosely structured interviews began with the researcher referring to the work of Yvonna Lincoln and Ernie House—both of whom named the power differential between program evaluators and program participants—noting that House went on to complicate that dichotomy by writing more personally about how he, as a program evaluator, also shared with many program participants the experience of childhood poverty and violence. The researcher asked interviewees to locate themselves in a similar way, by identifying the social, economic, and professional groups with whom they most identified or felt allegiance. The researcher then screen-shared with participants visualizations of the preliminary findings from the archival analysis and requested their reaction, interpretation, and storytelling around the observed trajectories. The researcher used an image of the social-ecological model and probes to clarify elements of the story that might relate to interviewees' perceptions of the system's boundaries, feedback loops and other types of exchanges, structural mechanisms underlying such exchanges, and relevant perspectives or interests.

Sample characteristics. All interviewees had doctoral-level educations. The interview sample was split nearly in half between evaluation scholars (five) and

practitioners (six, one of whom worked in a higher education setting as a practitioner as opposed to faculty). All but one interviewee were involved in AEA before or during the early 2000s, when many meso-level, institutional decisions and processes within AEA that addressed racialized difference took place. As such, nearly all the scholars and some of the practitioners had recently retired or were in the process of doing so. The only interviewee who was not involved in AEA at that time and was not nearing retirement was selected for her professional affiliation with those who had been involved in AEA at that time, the nature of her contribution to the evaluation literature, and the depth and breadth of her involvement with GEDI. Four of the 11 interviewees identified or expressed themselves as men, and seven did so as women.

When asked to name the social, economic, professional, and other groups with whom they identified or felt allegiance, nine of the 11 interviewees identified themselves with a racial group among many other social, economic, and professional groups: seven with racially otherized groups, and two as White. One of the two who identified themselves as White also identified herself as having grown up working class. One interviewee identified herself in terms of disability status alone (as not currently being disabled). Another simply identified herself as feeling closer to program evaluator than program participant.

CST Interviews: Analytical procedure and methods. Transcriptions of the recorded interviews were initially open coded vertically (within interview) and then horizontally (by question, across interview). Many codes arose from the theoretical and analytical frameworks underlying the research, which were presented transparently through the preliminary results, while others—for example, Topical Interest Groups, journal editorial boards and reviewers, specific AEA initiatives, and the importance and unimportance of language—arose *in vivo*. Analysis of the coded material was inspired by grounded theory and systems theories (Kim & Andersen, 2012).

CST Interviews: Findings. Because this dissertation rested on a theoretical framework that emphasized the constructedness of difference and identification with social group interests as well as an analytical framework that emphasized boundaries, relations,

and perspectives, interviewees' responses to initial questions about their social, economic, professional, and other group identification are quoted at considerable length to ground the remaining findings. Remaining CST interview findings are organized by preliminary findings from the archival analysis, in roughly chronological order:

- Use of variations of "minority," "white," and the phrase "of color" in relation to development of the MIE TIG
- The increase in frequencies for variations of "diversity" and "inclusion" in relation to the BDI study and implementation
- Use of variations of "culture" and "race" in relation to annual conferences and implementation and evaluation of the BDI, including the Cultural Reading and subsequent revision of The Program Evaluation Standards and the Cultural Competence Statement
- Use of variations of "indigeneity" and "colonization" in relation to establishment of the IPE TIG

Program participants and program evaluators: Intersecting personal, political, and professional commitments. All interviewees of retirement age who identified with racially otherized groups foregrounded that in their list of group identifications and affiliations. For example, Anna Madison stated:

> My professional identity, I'm involved in community development, and I've spent a lot of time with that community now….
>
> As an African American woman, I have strong ethnic identity. I'm not anti-anybody, but I'll tell everybody I'm pro-African American all day. But I'm concerned about other groups who've been marginalized, particularly the Native Americans, Latinos in general, and now the major attack on immigrants….
>
> So although I still have some identity with the [evaluation] field professionally, I've been very disappointed, and I've withdrawn. I just haven't seen the growth I think should occur within the field.

Madison was not alone in describing herself as having left the field or AEA—and subsequently lessening her professional identification with one or both—specifically out of frustration with and in response to its failure to institute changes with respect to growing

the field's understanding of racialized difference and systemic oppression. Two others—one who identified with a racially otherized group and one who identified as White—said the same explicitly; another two, both of whom identified with a racially otherized group, remained identified with evaluation and AEA, but were neither satisfied with the growth of evaluation and AEA nor necessarily optimistic about their immediate future. Rather, both spoke with muted expectations and a sense of resignation that the work is ongoing and long-term—a sentiment voiced by several interviewees.

Most who identified with racially otherized groups expressed that identity in terms that were tightly integrated with their knowledge and identity as researchers or evaluators, and many carefully parsed their personal identities as members of racially otherized groups relative to their professional identity as an evaluator. For example, Melvin Hall said,

> Well, I describe myself as being of the Black community, but not in it. And I've never been in it. So as a child growing up, I lived in a 99% African American community, but because of both social class and race segregation issues, my life was and my identity was partitioned. So in the Black community, I was viewed as a special person, and in the White community, same thing. So evaluation fits right in because, as a profession, you are dangling there between worlds. And you are in the removed observer role, and so it's a role that comes naturally to me because I've felt that role and been in that place, being an observer pretty much all my life.

For many but not all of these interviewees, personal identification with a racially otherized group was accompanied by political identification with liberation movement organizing. One said,

> Because I am a child of the Civil Rights Movement and the Black liberation struggle, the expectation for those of us going into institutions and organizations was that we were going to get prepared with tools and resources that could be of service—to our communities and some greater good, not just come out with degrees for ourselves that edify ourselves…. I came to the University of Wisconsin nearly 50 years ago from a Historically Black College to work on a Ph.D. in social psychology in 1968. And I came at a time that shortly thereafter, Black student protests, boycotts, all kinds of upheaval.

Ricardo Millett said,

> I was born in Central America, Panama. My grandparents were laborers that were brought from all over the Caribbean. In my case, Jamaica and Barbados, to Panama

> to dig that hole. And from the very beginning of that labor effort, both the railroad—by the way, the Panama Canal railroad and the Panama Canal—were basically built by imported labor, Chinese, Indian, West Indians, on the railroad and similar with the construction of the Canal. But *the whole pay system and the whole labor system were overtly and deliberately racialized…in terms of intelligence and capabilities and all of that was legitimated by research and literature* (emphasis added).

He later said:

> I go back to the days of the late '60s, when a lot of us were confronted with our—not so much our minority-ness, but it was unequivocal Black Power, kind of in your face. It wasn't social justice power [laughter]. And you were intimidated—I certainly was—*to look myself in the mirror and look at people who were just not doing well in this country and seeing a similarity*. And I could do that not only here in the United States. As I said before, I could go back to Panama, go to Costa Rica, go to Colombia, go to anywhere in South America, and I said, "God damn. There is a certain kind of thing going on here." (emphasis added)

Interviewees did not necessarily describe individual-level identification with a racially otherized group as influencing the way they approached evaluation theory and practice as much as they described their understanding of power differentials—and racialized power differentials in particular—as having done so. They sometimes but not always tied their understanding of power differentials to their individual-level identification with racially otherized groups, including political identification with larger liberation movements. For example, Hall, quoted below, is quoted earlier as seeing himself as of the Black community, but not in it—a positionality that he believed offered him insight that he drew from in his evaluation work:

> I started working, my first job out of undergraduate school and even one job I had before that were with TRiO programs. And I knew that evaluation of the outcomes of those programs was important to whether they were continued or not—the funding. So I knew from the age of 20 that people were making decisions about programs that I cared about in communities that I cared about using imperfect data and imperfect understanding of those projects…. [M]y first look at that wasn't actually along the lines of race. It was along the lines of power. And so what I understood was that people who have the power to make those observations and judgments were not the same folks who were being affected by the observations and judgments that were being made. And it happened that in the situations I was in, race and social class were two of the big differences. Later, as we became more

> sensitized to how systematic and institutionalized those issues were and I began to see and understand that it wasn't just a question of power, but that it was the power differential was sort of exacerbated by the race and social class that I became a little bit more activist in my approach, in my thinking about that.

When asked which communities he cared about, he did not refer to individual-level identification with the communities or to political identification with larger liberation movements, but rather to his professional experience with those communities. Still, the awareness that he described can be considered a political awareness in that it concerns power (Hilliard, 1989):

> In my case, it's African American communities and communities that are working-class and lower-middle class because my experiences were from those communities, and the programs that I'd been involved in were aimed at those communities, and those were the folks who were going to suffer if those programs didn't come out well in evaluations.
>
> I was very much aware that the graduation rate of students from our program was going to be used as a measure to determine if the term of the program continued or didn't continue—as one of the clear evaluation criteria, and that there would be other things that people would look at, but the whole idea of accountability and evaluation being in the background of my consciousness, if this program was going to continue, that very much affected the way I thought about evaluation when I went into it. And that program was for African American students who didn't meet the qualifications to be admitted to university through normal admissions.

Social group relations. Two interviewees did not include any racial group identification among the groups with whom they identified. They responded to questions about their group identifications by making the point that group identification is of less importance and less concern to them than relationships of respect and responsiveness. One answered as follows:

> Well, I would have rejected the dichotomy, because I really believe that in the end, we're all human beings, and we all bring resources, experiences, knowledge, insights, and when I meet with my stakeholders, my inclination is to be quite transparent about why I have to be humble in their presence, because I do not have the lived experience that many of the groups that I work with have. And it's only by us working together, forming relationships of respect and trust, that we're going to be able to share the things that we know. And that's what's going to be necessary for us to really do an evaluation that's going to make a difference.

> And I think part of that for me comes from my work in the deaf community, because I'm not deaf. And maybe someday I will be, we don't know, but I wasn't raised deaf and I entered their community as a hearing person, and through all of my interactions with them, I came to a very strong consciousness that I cannot know what it's like to be deaf only to the extent that they're able to convey to me and are willing to convey to me the nature of their experiences. And that is a physical difference. And if you put that in a broader context of: "Have I had the experience of growing up in a township in South Africa?" Absolutely not. But if I can convey that I respect what they know and what their experiences are, then we come to be more in relationship rather than us and them.

She went on to describe her introduction to diversity and inclusion in research as well as her view of positive changes in the field in ways that were unmatched by any of the other interviewees—especially by those who did identify themselves with a racial group and with a racially otherized group in particular:

> I started writing about issues of diversity and inclusion, initially from a disability perspective. That was in 1985. And then, I looked around and tried to find people who were talking about discrimination as a part of what we have to address in research, and power differences, and voices. And I found feminists, who were doing that, and of course, they were talking about gender. And then they'd always have one little line that would say, "We recognize that all women are not the same. That differences exist with religion and blah, blah, blah," but never went further than that, always one little line. And so when I wrote a book in 1985 on research in special education, I included feminist theory as a way of trying to understand the discrimination and the power differences, as a voice for people with disabilities. And some of the critics were like, "What is this woman-thing doing in this book?" And then, I think it was—I'm not 100% sure, I'm going to say 1989, perhaps, was the first edition of Research and Evaluation in Ed and Psych. And writing that book was a dig to find literature that talked about these issues. There wasn't much published. It was fugitive literature. All gray literature, knowing people who were working on things, getting in touch with them, but the progression, I mean, now I'm doing the fifth edition revision of that book, and it's fabulous because there's lots of literature [laughter]. I'm just thrilled to see that this is coming into its own in our profession.

Nearly all interviewees discussed the importance of relationships and responsiveness at length (discussed later in this segment), and most did so after acknowledging and sometimes honoring their own group identifications, which generally included racial group. One who described her identification with a racially otherized group later made the point that personal identity is insufficient and needs to be supplemented with

understanding of oneself as acting in service of those that the program is intended to benefit:

> I don't have to be in that identity group, but I am there to be of service first and foremost. I embrace a servant-leader stance, and this is not servant in the sense of being servile, but servant in the sense of…do they benefit? Not just do we do no harm, but are they left better off? And that's beyond my personal identity with them, whether I have the same attributes or not. And I have to recognize if I don't that I have extra work to do to tend to how my lens, filters, and frames are allowing me to look and actually see from their vantage point, to listen and actually hear from their vantage point, and to touch and actually feel from their vantage point. So that's why a core part of the work that I'm doing now is around cultivating empathic perspective taking, which is very different from sympathy because in empathic perspective taking, it starts with knowing self, and then you move beyond. And *this is not just in terms of touchy-feely stuff, personal and social identity, but in terms of social, structural, political location*. So power-privilege is the context within which all of this is taking place (emphasis added).

Kien Lee separately (in the context of culture, also discussed later in this segment) made the point that even representation may be unnecessary:

> To me, it's not an issue of representation. Representation is helpful, but not sufficient. And I don't think always necessary either. I think it's…knowing that you have limitations because you may not be from a particular group…. It's knowledge about what you know and don't know and who you are, your worldview, *and how it intersects with* other people's worldview. But the skill is understanding and asking questions and knowing that relationship and not that you can come in and understand a group of people (emphasis added).

Many other interviewees explicitly questioned the binary between program participants and program evaluators. More largely, based on their experience as members of racially otherized groups, they questioned "insider" and "outsider" status. Earlier, Lee had said,

> If I think about myself professionally, certainly I would say I identify with program evaluators. However, I would definitely say that there are many times when I do identify with program participants because I find myself sitting in a room sometimes listening to findings being presented, and the participants of that evaluation study are being spoken about…this third-party over here. And sometimes the language and the tone that's used is uncomfortable because it does sound like there's very much objectifying this group of people. So I'll give you an example where I've been in conferences that talk about immigrant integration, and I

> identify very strongly with the immigrant community, and so that's when it sometimes gets uncomfortable. I can't say that there's a word that they use that's negative or even anyone who's outrightly disrespectful or anything like that. It's just a very uncomfortable feeling of always feeling like who they're talking about I'm part of. So I think in that sense, I do often feel like it is that weird space of, like, you're an evaluator and you're understanding they're just representing the findings and there's probably no other better way to do it. Then there are times when you kind of go, but they're talking about, like, my peeps [laughter].

In some cases, this questioning of the binary between who program participants are and who program evaluators are was explained in relational terms that were tied to the perspectives and interests of the interviewees' group identification:

> Clearly the focus of my evaluation work is in Indian communities, and I am a member of that community. So I definitely identify with the community in which I participate as an evaluator…. [As an urban Indian], I'm distinguished somewhat from most of the evaluation work that I do because I do it in reservation communities. So I differ in that aspect from communities in which I work…. In Indian country, you don't establish yourself by, "Well, I live in a city and you live in a reservation," but by who you are tribally, what tribe you're a member of. And who you know. And the connections of who you know, through them who you might be related that you know in common. You go through that kind of a network so people can place you in terms of who you are as an Indian person. I don't know that in that placement role or establishing, I mean, I come in right away with elements of my identity that we can talk about. And so I don't know that we talk about insider, outsider, or other in the same way. We just identify us in terms of where our place is tribally and experientially as Indian people…. So I can't really say I think of myself as an outsider or they look at me as an outsider. I mean, they obviously know I don't live in their community. But then I'm very familiar with some of the shared history of Indian country and the shared language that we can talk about based on that history and that experience…. It's family and community, who you're related to, and do you know who you're related to.

Again, the tie to the perspectives and interests of the racial groups with whom interviewees identified and integrated their identities as evaluators was generally not intrinsic or automatic, but rather tied to the idea of ongoing political struggle, which several interviewees raised. For example, the interviewee quoted above continued later:

> These are big issues that have been with us, and we're in a struggle, so we continue to struggle. I'm not sure how I could explain what it is. I mean, evaluators, in many ways, I think, are somewhat at a disadvantage to be change agents, although I'm not absolving us of that role, but because we inherit a program. An evaluation is not

> driven by a research question. It's driven by this is going on, and we need to have it evaluated so we kind of come in with something that we inherit to look at, that we use that evaluation to issues that bring up broader issues. I think it's useful. I do think I have to be sensitive because, as I said, if I'm respecting sovereignty, I have to be very respectful of that, how I present that community. I mean, there are many issues around evaluators coming in and looking.

Geri Lynn Peak said,

> Well, I think it is a false dichotomy, but it's one that's very understandable because I think it comes from the false notion and the illusion of lack of bias that come with Eurocentric research training. Because White people having constructed this notion of race also think that they are exempt from the biases and pitfalls of acting as a party or a player in anything that has to do with race because they see themselves as a standard and a status quo and aspiration and all of that. So I think that—and I think it plays out in a variety of ways that disrupt that dichotomy—as a Black woman, I'm treated like I have entrée. And I don't know how many times I've shown up somewhere, and somebody thought I was someone's mama. So I'm located within that false dichotomy as the person without the power until I open my mouth. And even as I open my mouth, I might be mislocated or relocated because I choose to show up a certain way more and more… I present as an outsider and I'm kind of proud of it….

She went on:

> So I feel like we shift our relationship because we are all community members. So that's the first lie, that we're not part of community. Sometimes we're brought in from far away so we really are not part of the community. Most of the time, we are part of community because the phenomena that we're looking at has something to do with humans, and we are human. And we often live within or near or within culture of the phenomena we're trying to observe. So trying to exempt ourselves from having a role in that is problematic. So I see myself as multiple-y connected to these issues because of who I am, but I also do not see people who may exempt themselves who take on the label of White…. I feel like this work, the work of dismantling the trauma and the influence of racism is the only work there is, and the functional connections between that is the work. That's the work of this age of humanity. [All the work we're] doing is cast within that…. So I find myself, as a Black woman, multiple-y affected every time I go and work. I think, for me, the way it shows up was it just evolved out of my interest in eliminating racism, my development of an understanding that evaluation is a service activity that we work in service to…. One of the ways that I build community is through evaluation.

Peak noted that "we are human," not unlike the interviewee quoted earlier, who did not include a racial group among those with whom she identifies. However, the two

responses differed in that one contextualized the relationship within a larger system of oppression whereas the other contextualized the relationship within a larger culture. Importantly, the latter did acknowledge "systemic issues" and "discrimination," and the former did acknowledge culture. Still, the latter did not locate herself—or what the former referred to as her "role in that"—personally, politically, or professionally within or in relation to the context, whether perceived as one of systemic oppression or as cultural. She described her work as lying in conveying respect for community members' knowledge in order to "come to be" more in relationship with them, but did not acknowledge the relationship between her and them as already existing through social, economic, and political structures surrounding "the phenomena that we're looking at" in the way suggested by the former interviewee's exhortation to "…shift our relationship because we are all community members…that's the first lie, that we're not part of community."

Power differentials at interpersonal, institutional, and structural levels. Similarly, the two interviewees who identified themselves as White differed considerably from each other and in some ways from those who identified with racially otherized groups. One said:

> I identify as a White, cis-gendered female, heterosexual. I'm the grandaughter of German immigrants, and I grew up in a working-class background, and still identify very much with that, even though that's not my current tax bracket. I'm politically an independent, but left-leaning. I'm a mother and a wife, and I identify strongly as an educator. My identification as an evaluator is stronger than my identification as a social worker, but my social work values inform my evaluation. So academically, I have a dual doctorate…. And my psychology training was in community psychology, which I think comes in some of the participatory community-based evaluation. So yeah, in terms of my *cultural location*, I'd say those are my main identifiers (emphasis added).

> Social class is a big one and one of the areas of diversity that has probably impacted my life the most…. I was an Affirmative Action student, and not in terms of race, because I'm White, but in terms of social class…. [I]t was a culture shock to be around so many people with money. And to get a sense of their sensibilities and priorities, and it was definitely a learning experience. The one that epitomizes it the most for me is when I was in Introduction to Sociology, which I ended up getting a D in. I had been a straight A student all through high school, but I really couldn't grasp the frame that they were using. And for example, an assignment that I did very poorly on asked for the paper to be written in the style of a *New York Times* book review. And I had no more idea than anything. I had never read a *New York*

> *Times*. We got the *Progress Bulletin*, the little local paper and at the grocery, you could buy the *Los Angeles Times*, but I mean, I wasn't even thinking about what—I didn't even understand sort of the question—much less stylistically, what all that that implied. And so there were little moments like that where I was just sort of on a different planet. And also the fact that the school was relatively near, although, I mean, it was lightyears away in terms of culture, but geographically, it wasn't that far from my home.

When probed to connect the social, economic, and racial groups she discussed identifying with to the professional groups she mentioned and perhaps to evaluation, the interviewee went on:

> …[C]ommunity psychology was trying to look beyond just the individual. Social work will call it bio-psycho-social or ecological paradigm, and ecological crosses over to community pysch also…. [W]hen I was active in, for example, the American Psychological Association, I got a much more scholarly introduction I think to issues of race and culture than in the social work sphere where it was more about practice skills and application and less about theory.
>
> … I still do belong to APA…. I belong to the Division 45, which is the Minorities in Psychology. The title's probably changed. And also, the Women's Division….

Later, in the context of justice, this interviewee said, "[Racial justice] is not a term that I've heard used…. I mean, I've heard discussion of justice in the context of racism, but not as a phrase unique to itself."

Bob Covert, the only other interviewee who identified himself as White, described his group identifications very differently:

> I've been retired for about five years…. I was a math teacher, and I got my doctorate at Temple, basically, in it was research at the time. I mean, this was back in the early '70s, and so evaluation was not—there was not much about evaluation at that point, and so I basically, even though my degree was in ed psych, my emphasis was on research and statistics, so I started out very much as what I would probably call a researcher in the sense that a researcher working with programs, so, which gradually evolved in the literature to program evaluation. And from there, I went to University of Virginia…. [A]nd so I was interested in—I mean, I really knew a lot more about evaluation, and I was very interested in diversity issues….
>
> [W]hile at the University of Virginia—you probably don't know the history of the University of Virginia, which is a public institution, but did not admit women or African Americans into the undergraduate [program], until 1970, I think. So I'm there, in my early career, and this is now—we're in the '80s. But three of my other

> colleagues and I got together with the provost and said, "We know that there's a lot of issues here for students, based on how they're being treated by TAs and faculty, and so we would like to recommend some diversity training for our faculty." And the provost's response was, "Well, how do you know this?" Now, we already knew that. But he gave us $17,000…. Anyway, so [we] conducted a study…at the University of Virginia, and it took a year. And it was a census. We got as many students, faculty, and graduate students, and staff, and the results basically came back, and I think we ended up with maybe three or four thousand respondents. And a couple of thousand said that they would like to have training, so we went back to—so now I've shifted to being, I was both an evaluator and a kind of a, somebody who's getting the evaluation—came back, and the provost said, "Well, we don't have any money for that" [laughter].
>
> And that really was, like I said, a major [turning point]—in terms of my career, because I just figured at that point, well, we knew the answer, we spent a lot of money in finding the answer that we already knew, and then when we tell you the answer, you're not going to do anything about it. So I took it out—I just decided that I started professionally becoming more of a person involved in diversity training than in evaluation. I mean, I think—so that was the shift that I have made, and so in the latter part of my career, I basically just did diversity training and started the big class there, which was pretty successful…. [S]o I was on both sides of it…. So I mean, I guess that gets at the point about where I am relative to evaluation as an evaluator, as a trainer, a long-time trainer of evaluators, and as a consumer itself.

Only later, discussing the discourse shifts suggested by the preliminary findings from the archival analysis and a few times thereafter, did Covert refer to his personal identification with a racial group and specifically as White:

> The idea of the institutional part of it is the part that people don't get. And I think that you're talking about blaming people for being poor, or blaming people for being Black or women, or they're not doing their part, doesn't illustrate the whole society that's set up to perpetuate those ideas. I mean, that whole notion of privilege, what it means. I mean, when somebody doesn't like me or discriminates against me, I don't think, "Well, it's because I'm a White guy."

Finally, discussing the possibilities that variations of "justice" may offer, he said,

> I think that the whole idea of social justice is a wonderful way to do it. And it may even be a way to sell it. But I think probably—just think about this, though: We could talk about social justice, that's not too bad. But don't start talking about things like racial justice. Or gender justice.

Both interviewees quoted above identified themselves in terms of race—as White—and in terms of gender. One foregrounded this identification and her personal experience of class oppression and the other embedded his racial identification into his responses to later questions, having foregrounded his professional training and experience and the political context surrounding them as one of racial oppression. Their descriptions differed in the extent to which they emphasized their experience with respect to power differentials—whether racialized or classed—at individual, institutional, and structural levels, particularly in relation to their professional identities as evaluators. Both also differed from the interviews with those who identified with racially otherized groups, wherein personal identity was centered, but professional identity with respect to evaluation was almost always grounded not in that personal identity, but rather in interviewees' accounts and reflections that illustrated their understanding of power differentials more broadly, beyond their personal experience.

Overall reaction to visualizations of preliminary findings from the diachronic and archival analyses. While responses to the visualizations did vary by racial group identification, gender, and generation, they are not reported as such largely due to the small size and purposive nature of the sample. Moreover, the number of interviewees who did not explicitly identify as a member of a racially otherized group was only four, two of whom identified as White and two of whom did not include any racial group among those with whom they identified; similarly, the number of interviewees who were not of retirement age was only one.

Most interviewees understood the delimitation to language in U.S. peer-reviewed evaluation literature as an effort to provide the field with a mirror to reflect on the type of discourse it has sanctioned. Most understood the visualization and expressed no surprise. This was especially true among those who identified as members of racially otherized groups, but also for Covert, who identified as White. The omission of books, journals from related disciplines like educational research, and evaluators' on-the-ground practices—all of which were intentional attempts to focus on the construction of racialized difference by the professional association, discipline, and industry of evaluation, as opposed to by

subgroups within it—formed the basis of concerns regarding the sample to the extent that any were expressed by interviewees.

Two interviewees—one who identfied as White and one who did not include any racial group among the groups that she identified with—responded with some skepticism bordering on disbelief regarding the pool of literature, saying that the literature that they are immersed in does not seem to correspond with the visualization. One expressed concern about whether several particular pieces from the evaluation literature were included in the analysis (they all were) as well as a desire to see the raw data.

Other interviewees—most but not all of whom identified with racially otherized groups—cautioned against the focus on U.S. peer-reviewed literature, esepcially considering that it likely disproportionately reflected the work of scholars and practitioners classified as White. One suggested collecting demographic data about the authors published in U.S. peer-reviewed evaluation literature, if it were collected, to see if language use differed by racial classification. They particularly cautioned against the interpretation that such literature reflected usage and work in practice "on the ground": Not only was the typically more critical work of practitioners not necessarily, certainly immediately, reflected in the literature, but scholars' use of critical terminology also did not necessarily correspond with critical application even in their own practice. Some interviewees also raised questions about the role and importance of language—specifically, the extent to which it shaped or reflected the field's thinking—which were described in the comparison between culture and race. No interviewee had a simple answer to this question.

Interviewees' responses to each preliminary finding from the archival analysis are detailed below. Each finding starts with the factors that may have contributed to it, to the extent that interviewees speculated about contributing factors; each finding ends with its potential implications, according to interviewees, for otherized groups engaged in liberation movement organizing.

Use of variations of "minority," "white," and the phrase "of color" in relation to development of the MIE TIG. The increase in annual frequencies for variations of "minority" coincided with the establishment of the Minority Issues in Evaluation TIG and

publication of the *New Directions for Program Evaluation* edition focused on Minority Issues in Evaluation. Annual frequencies for variations of the term "minority" and "white" were consistently higher in U.S. literature than that for variations of the phrase "of color." The internalized, interpersonal, institutional, and structural factors that might have contributed to these findings and their potential implications for otherized groups engaged in liberation movements according to those most closely involved are described below.

Internalized, interpersonal, and institutional factors might have contributed to the use of variations of "minority" and "white" as opposed to the phrase "of color." As described in the archival analysis, AEA grew out of the close professional, social, economic, and racial group affiliation among academic researchers, policy makers, nonprofit organizations, and evaluation practitioners during a time when racialized neoliberalism became manifest in the dismantling of redistributive government programs like Affirmative Action. Perhaps not surprisingly, in 1986, the representation of evaluators who identified as African American in the Association was only two percent (Covert, 1987a, p. 96), as discussed in the textual analysis. The language used in efforts to address under-represented groups could be understood in this context.

While several interviewees named specific individuals who would likely be classified as White—Bob Covert, Ernie House, Karen Kirkhart, Donna Mertens, and Michael Quinn Patton—within the burgeoning Association as having supported the efforts of evaluators representing racially otherized groups, Anna Madison reflected on the sense of isolation evaluators representing racially otherized groups felt even from evaluators classified as White whose work she found intellectually compelling with respect to racialized difference.

> Probably I see the world from a different world view than they do. We have a historic—we're historically cohorts in that we were born and raised and acculturated during the same historical period in America. And probably, we share some of the same intellectual thoughts as well, but I think that's where it stops.

She continued by recounting the need to establish a Minority Task Force and eventually the Minority Issues in Evaluation TIG in this context:

> I remember we were three Black folks at the conference, and we would be looking for each other…. But Bob Covert came to us and asked us if we would create this, if—the best way we could get the organization infiltrated—we would have a TIG. And he said, "I can't do it. I need somebody who will do it." And so I picked it up. That's how we started at this.

Structural factors might have contributed to the use of variations of "minority" and "white" as opposed to the phrase "of color." Still, Madison described the TIG —while necessary and valuable—also as insufficient:

> Well, every time there were new people coming into the organization, people of color, we were trying to get them all into our TIG. So, we would have a huge group. We would have international people, and then the next year, we would have not the same people. What was happening is—the reason we wouldn't have the same people is because people didn't have funding. And so, it was really hard to get any traction going when every year we had to reeducate and start all over again. And so, at any rate, the academic people were the most stable because their universities were supporting them.

Beyond the racial stratification within academia and racialized distribution of wealth more generally that she referred to above, she alluded to structural racialization and the reproduction of racialized difference within evaluation. She discussed Ernie House's work in that regard:

> I feel that and I will—I mean until I see something different—I feel that the major people who do the evaluations or the people in positions of power when I was very much involved in it and when I was part of a national panel that looked at evaluations done by the federal government, one thing I noticed is that people…that get grants to conduct these big…multi-million dollar evaluations had no clue about the people who they were evaluating and the programs. …[T]hey look at the symptoms of our society. And the programs are designed to address the symptoms. And they are not addressing the root causes as to why we have these problems. And I think that in the field of evaluation, these people bring these attitudes into their work. That there is something very deficient about these people. And I mean it's not said in the literature. It's not said in the meetings. But it's implied by the choice of variables. By the choice of methodologies. It's implied there. So I think that there's still a big gap between first redefining the work. What is it? I think the piece that Ernie did last year really kind of brings on that whole racist issue. I think [Ernie House] gives an authentic…finally…he's giving a white-privilege point of view that has never been presented because it hasn't been acknowledged in the literature….

> [In his 2017 article in the *American Journal of Evaluation*], he's acknowledging that, hey, this group of—over here, who came over here and decided to—who from Europe—have decided to set up, make themselves gods and kings. And everybody else is deficient. But these are the very people who are leading the evaluations and who are making judgments about other people.

Madison closed this portion of the interview by crediting the TIG as having initially raised this type of awareness, however, and as having succeeded in creating change:

> [W]hen Bob Covert was president and…we created, together, we created the Minority Issues TIG for that purpose: to try to look at the race issue and the cultural issues….
>
> And I think that the field—I think that there are things that have happened in the field that have caused people to be a little more knowledgeable. And I…I think the creation of the MI—The Minority—MIE—had a lot to do within the organization. At least bringing it to the surface. It might not have changed people, but we made them acknowledge it.

Potential implications of use of variations of "minority" and "white" as opposed to the phrase "of color" for otherized groups engaged in liberation movements. Among members of this small and purposive sample, reaction to use of variations of the term "minority" tended not to vary sharply by the groups with whom they claimed identification. While sympathetic regarding the word's underlying history, many interviewees had a negative reaction to its continued use—particularly relative to variations of the phrase "people of color"—that raised power dynamics to the surface:

> [A]s I said earlier, "minority" bothers me on a number of fronts. One, because the reality is this country, in the sense of race and mixed race people…or people who are—quote…identifying as European, White ancestry are not going to be the majority. So I think in terms of certainly wealth and power, there is an element to "White" and "minority." But I think in terms of identity of this nation, what do we really mean by "minority" going forward? So that's where I have the problem with "minority" per se.

Donna Mertens captured the nuance that many interviewees shared in their interviews:

> … I don't think it's always an either/or thing. Sometimes, "culture" is used because people are uncomfortable with talking about racism or ableism or audism. They're uncomfortable with that. But you can talk about cultural differences, and then I'm okay with that. That's fine. And so that's what strikes me about talking about

> "white" and "minority." That seems to me to be the comfortable labels, but not necessarily accurate labels. And so you look at so many contexts in which African Americans are the majority, and yet we're talking about them as the minorities. Why do we do that? Because it's easier for us to talk about it that way—"minority"—because that helps us gloss over differences between different groups of people who on the basis of race and ethnicity are experiencing discrimination.

Covert noted:

> The consistency with which "minority," "white," and "race" stayed over the years, with some minor ups and downs, but I mean, basically, they stay at the same place after the beginning. And then the increase of diversity and culture. I think that one of the things that this is a great reflection on is how the intellectual community gets uncomfortable with something, and/or likes to come up with new terms. And that sort of avoids what some of the serious issues are…and it goes back to the idea of… "colonization" and "indigenous" is, I mean, not mentioned at all…. [O]ne of the general problems is that I just don't see too many people, at least practically, seeing the implications of systemic discrimination. And then I think the other thing that's really one of the other really fascinating things is the lack of the use of "of color." Because I mean, I'm a White—old, White man, but I would say that that just reflects people's lack of comfort with that term. And fortunately or unfortunately, most of the people who are writing about this are probably White, so….

Contrary to preliminary inferences drawn from the archival analysis, however, it was not necessarily the negative reaction described above to the word "minority"—to diminishment of the current global majority and to what would likely have been the majority population on North America were it not for settler colonization, including genocide and exclusionary immigration policies—that led the TIG to change its name from the Minority Issues in Evaluation to the Multi-ethnic Issues in Evaluation in 2004. Rather, Hall recollects it as an attempt to communicate a broader base for the group, beyond those who identify as African Americans. Apparently, "minority" had become synonymous with "Black" and not interpreted as applying to racially otherized groups whose sizes were even smaller than that of the U.S. population classified as Black or African American. This conflation might help explain the close proximity and nearly parallel trajectory for variations of "minority" with those of "White" on the diachronic analyses of the U.S. evaluation literature as a whole and the non-AEA-affiliated literature in particular. Corresponding with the roots of Civil Rights legislation, the conflation of "minority" with

"Black" and its contraposition with "White" illustrates the dialectical relationship between blackness and whiteness in prevailing constructions of racialized difference. Members of the MIE TIG had hoped "multi-ethnic" would communicate to evaluation scholars and practitioners representing racially otherized groups as well as representing groups racially normatized as White that the TIG represented interests that were not limited, often relegated, to African Americans. Hall described the hope that the new name would increase the TIG's size and strength within the Association.

> And it was essentially when African Americans were needing to think about, strategically, how to incorporate the other groups that were beginning to come in, people of color, coming into the association. And we used the multi-ethnic label as a—it turned out not to work out this way—but as a way of saying to Chicano/Chicana, Latino/Latina, indigenous, that everybody could be under this tent. Because at that time we could all fit into a Volkswagen van anyway [laughter], okay, and so the idea that they would not be welcomed was really nonsensical. So we changed the label, but as that was happening, the social consciousness in the country—and I don't know if—I'm sure it wasn't reflected in the evaluation literature, but the social consciousness of the country was such that those groups came in, and after there were four or five, they wanted their own TIG also. So of course, we had the Indigenous TIG and then, followed by, LaRED. So I was a part of that and witnessed all of that unfolding.
>
> … I know for me, it was not a reaction to "minority," although that was certainly part of the conversation and part of our consciousness. But we had several options in terms of moving from minority. So it was the decision about which of the options to take after moving from minority. [Attracting other groups was] considered to be the real selling point.

The attempt at building coalitions and solidarity across groups through use of the word "multi-ethnic," while possibly miscalculated, differed from use of variations of "diversity" in that it was an attempt by one racially otherized group to communicate to other racially otherized groups that there was a place for them within AEA. Potential factors contributing to, and implications of, the rise in annual frequencies for variations of "diversity" are discussed below. The establishment of the IPE TIG that the interviewee mentioned is discussed at the end of this segment.

The rise of variations of "diversity" and "inclusion" in relation to the BDI study and implementation. The increase in annual frequencies for variations of "diversity" and

"inclusion" as potential signifiers of racialized difference in U.S. evaluation literature between approximately 2001 and 2016 coincided with the BDI, establishment of the Diversity Committee on AEA's board, and other related events. The factors contributing to, interpretations of, and implications of these increases and coinciding meso-level, institutional decisions and processes according to those most closely involved at the time are described below.

Internalized, interpersonal, and institutional factors may have contributed to the increase in annual frequencies for variations of the term "diversity." In the 1990s, Ricardo Millett "found [him]self" serving as a program officer at the Kellogg Foundation. As such, he had access to a sum of discretionary funds that he could use and combine with other program officers' discretionary funds to support projects that fit within the confines of the Foundation's interests. Driven by a "desire to see changes in research and evaluation methods…that would inform more effective policy and programs…and a desire to have AEA be a champion of these kinds of methodological changes and outcomes"—and parallelling his own path from Panama to Brandeis and eventually Abt Associates—he started using these funds to develop and support programs to recruit young evaluation practitioners and provide them internship opportunities.

In the late 1990s, Millett gave a keynote address at the Canadian Evaluation Society meeting. In the audience was Donna Mertens, cited earlier as having invited John Stanfield to keynote AEA's annual conference under her presidency. Having heard her work cited repeatedly in Millett's talk, she decided to introduce herself to him after the address. Their mutual appreciation for each other's work and shared goals allowed Millet to approach Mertens about the possibility of AEA instituting a Kellogg-funded program to increase the pipeline of evaluators representing racially otherized groups. Then president of AEA, Mertens had run on a platform that involved addressing "issues of diversity and international relations." She was thus receptive to the idea, which grew into the BDI described in the archival analysis. Mertens described the financial support from the Foundation to hire staff as having facilitated the effort's success. She described the tension and lack of support among AEA board members as having hindered it.

In 2002, Melvin Hall served as a replacement for Denice Ward Hood, cited earlier as co-author of the first—still among very few—pieces of evaluation literature to name white supremacy and intersectionality, as a discussant on a panel by the MIE TIG. In the audience were Karen Kirkhart, cited earlier as having introduced the concept of multicultural validity into the field and having initiated the Cultural Reading of The Program Evaluation Standards that were described in the archival analysis, and Elmima Johnson, who held a position at the National Science Foundation, which would later underwrite portions of the BDI implementation. Kirkhart shared Hall's comments and her appreciation of them with her husband Nick Smith, then president-elect of AEA, who appointed Hall to chair the Diversity Committee that the AEA board had established as a standing committee a year prior. Hall described the Committee as follows:

> [T]he Diversity Committee, if you think about that label, the Diversity Committee was strategically composed of someone from special populations, someone from the Feminist TIG, and so all of the TIGS that represented what you might call special communities, had a representative on the Diversity Committee. And in a sense, it was our Committee of Others.

Hall's comments regarding otherized status of these groups was noteworthy when one considers that they represented the interests of the majority of program participants. In addition to describing the composition of the Committee as strategic, Hall described the approach taken more largely as both strategic and effective.

> The Building Diversity Initiative, their report was considered powerful as they came to the board. It wasn't a hot potato because it was very reasoned in the way it was written and the people who were behind it, Donna Mertens and all those folks, Hazel Symonette. So it wasn't a hot potato, but it was a report that created some challenges for AEA organizationally, when the board tried to figure out, "so what do we do with these 14 recommendations?" They basically cut across the whole Association. And so they put the Diversity Committee in charge of figuring out how to implement them, but recognizing the implementation could not all be done by the Diversity Committee because it cut across the turf of these other standing committees. So Sharon Rallis…Sharon was the incoming president—so Sharon and I had a conversation and talked about this dynamic, and I suggested that as a standing committee, the Diversity Committee could not have other committees needing to report to it. But we also didn't want to just turn this over to the board because it wouldn't receive the kind of focused attention that it needed. So we held a retreat at the end of an AEA conference. Sharon came and we had the incoming

> chairs of all the other committees that would be asked to take on a BDI recommendation. And so in small groups, each recommendation was discussed, what it meant, what we might do about it, all those sorts of things we worked out in small groups that then presented to the whole group. And we came out of that retreat with an implementation plan of who was going to do what.
>
> And so it was Sharon's clout as the incoming president behind me as the figurehead pulling the retreat together and then inviting all these other committees. So we finessed the question of how does diversity work with each of these committees by having a liaison from the Diversity Committee join each of these other committees in their follow-up on the BDI initiatives. So I think the Association avoided another kind of injunction that could have disrupted the implementation. So it went as smoothly as it did because we found a gracious way to weave it back into the organizational structure.
>
> It was one of the major events going on at the same time, and that is that the Joint Committee was revising The Standards. And Karen Kirkhart as part of the Diversity Committee, one of our other subgroups was led by Karen, and it did the Cultural Reading of The Standards, which impacted the revision of The Standards. It very much impacted the language. Don Yarbrough was very appreciative of the 100-plus-page document that we provided, and I think it was—they said it was influential in how they revised The Standards and the little scenarios that are used to illustrate The Standards.

Structural factors may have contributed to the increase in annual frequencies for variations of the term "diversity." In addition to describing the process as strategic, Hall felt that use of less threatening terms, like "diversity," was necessary initially to build broad support. He considered use of variations of "diversity" as strategic as the composition of the Diversity Committee and approach to implementation:

> [S]ince we were trying to get this on the radar and get people to join and respond so forth, we sort of eagerly made a switch [from 'race' to 'diversity'] because it mattered for launching the conversation and faster movement. We lost something in the process, but we gained a lot as well. So I think in my own thinking about it and my own words, when I reflect on it now, there were two important factors. One was Karen Kirkhart because—well, I mentioned Karen was there in the beginning. She was President of AEA, but she also was then the person that asked me to come on to the Diversity Committee. In fact, I said, "Okay. You got me into this thing, I want you… [laughter]." And so she became quite a force. She was also part of the Cultural Competence [crew—the] other task force for the six years, and she even stayed on after I bailed. And she was developing [the concept of] multicultural validity at this time. So her voice was very influential in the phrasing and in the

> philosophical orientation we were taking to things. Jennifer Greene was the liaison on the AEA board to the Cultural Competence Statement, and so as we developed drafts, Jennifer was the person on the board who would react to our words. And so she was also influential in helping to firm up and support some ways of thinking and challenge other ways of thinking. So I think those two were particularly influential and supportive, and most definitely, they both became close friends of mine at that time, through that process. And they have definitely influenced my thinking and writing ever since.

This was particularly so at a time when AEA was averse to alienating members—having recently done so by taking a stance on high stakes testing and federal privileging of experimental design.

> So [the departure of high-profile members in response to AEA taking a public position on high stakes testing and methodological hierarchy] super sensitized the board and other people to the politics of language, and so at that time, our coming in with the cultural competence and other things, "diversity" was a much more palatable term. And so I think that there was a little bit of the mixing of streams of politics that contributed to those things flowing the way they did.

Importantly and perhaps not suprisingly, however, the history recounted above was complicated. At the time of the BDI and Diversity Committee, the MIE TIG was also addressing recruitment and professional development—and had been doing so for two decades as described in the archival analysis and in regard to the preliminary finding about use of the term "minority." Mertens interpreted the 1999-2001 Building Diversity Initiative study that grew out of the discussion between herself and Millett as having raised the level of attention beyond the TIG level to the association level and as having broadened the focus beyond racially otherized groups to all otherized groups. Other interviewees interpreted both the process and the language (discussed under implications for liberation movements) differently. For example, one interviewee interpreted raising the level of attention as follows.

> [T]he pillars for the Multi-ethnic Issues, namely to increase the pipeline and numbers of professional evaluators of color—that was one pillar. The other pillar was to increase the competencies and the skill and knowledge of all evaluators to work in different communities. Those were the two pillars. Those were the twin pillars of BDI. And so, the non-involvement of MIE directly in that was what created some turbulence.

Another interpreted the broadening of focus beyond racialized difference as follows:

> [T]his is my personal observation over time being in academia for 42 years. Because I perceive your chart—I started in '76. [A]round 1999 to 2000, there was a national academic discourse and backlash on equal opportunity. The backlash on Affirmative Action created a new language that got away from race and started talking about inclusion and diversity. And inclusion and diversity was a way of getting rid of that strong racial dichotomy between Black folks and White folks and Black folks and racial minorities as we were called, and still. So an academic univers—this is interesting. Once an idea gets out in the main discourse, it's taken on by the whole industry. So there was a conscious shift in America trying to get away from the realities of race disparities, between race and ethnic disparities, between non-white ethnic groups and racial groups. And so if you do inclusion, then you could include White women, LBG—and I'm not against these groups, but I'm saying…. Because I served on one of those Diversity—two of those Diversity Committees.
>
> There was a little period in AEA where there were a bunch of radical White women that ran it. And they were the ones I think that—I'm not sure, but I think they were the ones who first…pushed the diversity agenda because naturally, that opened it up for White women…. [I]t was after Will [Shadish] that this little group of White women came in, and they had this Diversity Task Force and they were supposed to have been—they asked the MIE TIG to be involved with it, but of course they were going to control it. But the whole thing was that they came up with—oh, they were getting ready to also write the Principles and they wanted to make sure that diversity somehow—some element of diversity was included in the Principles…. And they wanted to make sure that diversity was addressed there and inclusion and all those issues. But yeah, that's why, putting that diversity was really promoted. And I mean, it was really pushed hard in our organization. So the Diversity Committee has been around for a while in various iterations because a lot of different perspectives and views have come into it, but it has survived in terms of actually a set of diversity principles being actually adopted and supported by the… Association.

The concerns raised by this interviewee regarding social group interests are discussed in the section devoted to implications of the increase in frequencies for variations of the term "diversity" below.

Potential implications of the increase in annual frequencies for variations of the term "diversity" for otherized groups and liberation movements. Ricardo Millett crystallized implications of the construction of racialized difference for evaluators as follows:

> The question is, "Does the application of these kinds of designations lead to differences in the way research questions are framed? Or how data is analyzed? And if not, what is it an artifact of? Politeness? And if only politeness, does it affect the validity and utility of the research product in addressing the systemic issues?" Another thought that comes to mind is, "Has the fuller recruitment, inclusion, participation, and research activity of evaluators of color over time, had an influence on the way these terms are used? And does it get more authentic kinds of analyses and conclusions?"

Some interviewees raised concerns about the term "diversity" that may help explain the proximity of its trajectory on the diachronic analysis to that for variations of "inclusion":

> Diversity is an outcome, and we use it like it's an intervention. That means, we're going to get more Blacks, we're going to get more women, and we're going to get more Latinos. We're going to get more people who have different backgrounds and shades, we're going to get more people with different abilities. And we're going to get more people who aren't evaluation-trained—which made me laugh because there's a whole group of us that weren't, but we're professional evaluators. And so we're going to get all these things. But really it's inclusion that leads to diversity. Because you can bring people in. And it's more apparent I think in schools than it is—like independent schools than it is, say, in at AEA. You can bring in teachers of color into a Waldorf school environment [but] if you don't make it comfortable for them to practice without setting aside all of their soul…. [Y]ou can bring children and families into a Waldorf school, but if they keep that Eurocentrism at the forefront, many of the children will not stay because people haven't learned to be inclusive. They just want to have diversity. They want to look a certain way. They want to put it on like a dressing. They want to put it on. But if we become inclusive, if we think about how should we be such that different kinds of people want to be with us? We're back to Hazel's two-way mirror.[55] We're back to transforming selves in order to transform the system. We have to become different beings in order for people who want to be with us, right?

Others found value in recognizing and using variations of the term "diversity" in the context of the liberation of racially otherized groups:

> I'm more comfortable with the idea of understanding diversity [as opposed to 'minority']. And that's just not diversity that cuts across race [but] class and economic realities. And I think maybe we've created a Trump because we ignore too much some of the effects of that economic diversity that was happening among

[55] Peak was referring to the Integral Evaluator model (2014), in which Hazel Symonette advocated for internally facing reflection and self-work as an ongoing and necessary part of outwardly facing structural work.

> White people and the impact of that in terms of, in essence, politicizing them. When somebody came along and politicized that in a very negative way. And I think a lot of things go into that, not just economic reality, but just that change in dynamics that's going on in terms…we've talked about with "minority." That's a difference, which was made very, very visible with a Black president. So it stirred the pot and very deeply, as we saw.

More, but not all, interviewees expressed concerns about the implications of using variations of "diversity" as opposed to terms with greater analytical specificity. Many of them simultaneously acknowledged this lack of specificity as underlying its attractiveness. For example, Kien Lee, co-author of the BDI, said:

> It suggests that the word "diversity" and "culture" are certainly code words that are safer to use than race, White, or of color. And so there's more comfort in using those terms. It also says that, it could be that, it's used to describe more broadly, the groups beyond, differences beyond race. So if you're talking about sexual orientation, gender, urban culture versus rural culture, it may be used more broadly….

Indeed, Hall himself expressed some ambivalence even as he continued to characterize the language choice as intentional.

> I do believe that if you calculate how to get forward movement and make some concession in order to expedite forward movement, once that beachhead is established, you can then appropriately and usefully come back and pick up and fill in some things that you had to give up along the way. And that's where I think we are with race…. [S]witching to "diversity" allowed a much stronger foray, and now that—in fact, it was so successful, that it is getting the backlash it's getting. In some ways, I put diversity and Affirmative Action in the similar situation of having made so much sense, and having been so well sold, that it made too much progress. And some people looked at that and said, "Oh, no. This is…. We've gone too far." And so the backlash then tries to chip away at it. And it was because it was so successful. So what's a backlash to Affirmative Action, and somewhat to diversity, it is now both easier and most appropriate to try to now bring forward the race discussion because we did lose some of the edge—switching to "diversity."

With respect to the need to describe "groups" and "differences" more broadly and the potential in evaluation for the concept of intersectionality, Covert and Hall raised the following tensions that they had observed. Covert said:

> [T]he language is only one dimension of the whole problem. It goes back to what can we do about it? You can even know about it, intersectionality, and you can pass

> the test. But you don't know—that, how, that whole notion of—I can't tell you how many times, for example, I've been in situations where we're talking about race, and a woman will stand up and say, "Well, you think you've been discriminated against? Let me tell you." And I think that, until people start to think about this notion of social justice in a way that it's equitable and takes into consideration a whole variety of different things, then it's just playing the minority groups against each other, competing with each other, and letting the status quo continue.

Hall said:

> Yeah. Well, I'm a pessimist, unfortunately because I question whether we have the bandwidth to do what I think is needed…. [*T*]*he intersection lives within a single body.* Yet in behavior, we act as if, depending upon the situation, we're not all part of the same. *And the psychological principle that I think explains that is when we let one of our identities, one component of our identity be the front face and put the others in retreat mode, we take on a role that has with it all kinds of other perspectives*. So White women who believe themselves to be feminists in California voted against Affirmative Action. Because they voted not as feminists, but as White women with sons and husbands. So that's how we allow one component to cause us to forget the other components (emphasis added).

Hall's description of the centrality of White men as a kind of pivot corresponds with critical constructions of difference, wherein otherized groups are dialectically differentiated from those that are normatized. He elaborated upon it—again, in psychological terms as opposed to social group interests—in considering the possibilities for working across group identification:

> So how do we fix that? The answer is something I don't know that we know how to do, which is—we are so used to thinking of each of those components as defined against White maleness that we don't have a way of thinking about those components defined against each other. So Black people don't know how to talk to brown people, but they do know how to talk to White people. And so Blacks and [brown people] only communicate effectively through White people, not directly [to] each other. And I don't mean through White people, but when the White person's the enemy. So somehow intersectionality has to lead us to cut out the White devil as the counterpoint in order for us to reunite *the different components of our individual selves* (emphasis added).

The above comments convey the difficulty of multiple identities working together when identity is conceptualized as psychologically (or culturally, genetically, phenotypically) intrinsic to individuals. Conceptualizing identity as atomized attributes that differ within

and between individuals—as suggested by the term "diverse"—focuses attention on representativeness. In contrast, conceptualizing identity critically—as socially constituted around shared interests arising from shared experiences of oppression—focuses attention on the structural arrangements underlying the oppressive dynamics. For otherized groups, increased understanding of their shared interest in liberation from oppression can fuel coordinated action to change the structural mechanisms, which sometimes differ among groups, reproducing their otherized status.

Use of variations of "culture" and "race" in relation to annual conferences and implementation of the BDI, including the Cultural Reading of The Program Evaluation Standards and the Cultural Competence Statement. The annual frequencies for variations of "culture" as a signifier of racialized difference started to fluctuate relative to "race" in U.S. evaluation literature in 1988 and in AEA-affiliated literature in 1994. In both cases, the shift coincided with AEA conference themes and keynotes. "Culture" finally surpassed variations of "race" in 2008, coinciding with implementation and evaluation of the BDI, including the Cultural Reading and subsequent revision of the Program Evaluation Standards and development of the Cultural Competence Statement. The factors contributing to, interpretations of, and implications of these trajectories and coinciding meso-level, institutional decisions and processes according to those most closely involved at the time are described below.

Internalized, interpersonal, and institutional factors may have contributed to the trajectory for variations of the term "culture" relative to that for those of "race." The importance of the distinction, or relationship, between culture and race varied depending on whether interviewees identified with a particular racial group. One who did not identify with a racial group said,

> What ["culture"] does is allows us to have an umbrella under which we can identify those aspects, those characteristics that are relevant within a particular context. And sometimes, race is the most salient characteristic. Sometimes, it's not. But I want to be able to, as an evaluator, to have that broader umbrella. What I'm looking for is what's relevant within this context that tells me, "These are the systemic issues that have to be addressed. These are the basics of discrimination within this particular context."

All interviewees who did identify with a racial group saw the distinction somewhat differently. One said,

> … I found [it] interesting that this term, "culture," had actually eclipsed the term "race." I'm not sure that they're—I'm not sure without the orientation to how they're used in the literature. [That's important]. But obviously, that growing—and I don't think that they're synonymous by any means, although they're interrelated. And I would find it worrisome a bit if we're downplaying or softening the issue of racism by finding a different terminology. On the other hand, since I work within a culture and the focus is to be culturally-responsive, I'm pleased to see that this attention to culture, or the use of the word, has grown so much. So that would be my concern. It couldn't replace "race."

Before elaborating with more nuance, Covert somewhat succinctly said, "My interpretation of it is that people don't want to talk about race." He then talked about how language, however carefully chosen, can be co-opted:

> [I]t's so complicated. I think that the thing about cultural competence and that stuff—to a certain extent, that's been co-opted, right? It's not that the language is bad language…. It's okay to talk about culture and diversity and inclusion, but what do they mean? That's what I'm saying about being co-opted.

Peak echoed this concern, and went into more detail:

> For me, it's not so much the word that's used, but it's the ability for people to take a word and redefine it. Because the cultural competence always was about White people and power. But there are other terms that can be turned around and misutilized so when we look at—oh, what was this? It's one of the terms. "Culture," "diversity," "inclusion," "multiculturalism"—those can be used to obscure specific experiences that communities have had…. And what we're talking about, in terms of a very specific justice orientation, is helping people understand that we're not talking about immigrants who chose to work in exchange for transport [laughter]. We're talking about real terrorism. We're not talking about supplanting a group of people who decided that they were just not up for that work, but people who had autonomy and the desire to maintain their freedom, who were then very specifically annihilated, or very intentionally marginalized. And there's a number of people who don't understand that there were polio blankets, which is fairly well known…. [It's] phenomenal. And you tell somebody that the penalty for becoming pregnant without the master's express say-so—and I know this was true in Haiti—the penalty for that was to have the baby cut out, and you either live or die. That's mind-numbing.

Many interviewees described how they used variations of "culture" as opposed to "race"—with varying degrees of consciousness, intentionality, and resignation—to keep the conversation broad, avoid alienating potential supporters, and create space for the work they thought was necessary to happen on the ground. For example, Ricardo Millett, who had initiated the BDI two decades prior, reflected without a hint of defensiveness or regret on not necessarily having the luxury to choose words carefully—with some anticipation of how they could be interpreted and later used—during periods of tension between immediate needs and strategic goals:

> To be honest with you, back at that time, I personally was not as focused as maybe I should have been on the issue of whether I should coin this "culture" or "race." I was more interested in making a dent, creating an opportunity, a space. Since many of the foundations, including Kellogg, were saying to me and other directors of evaluation, "Where are the candidates? How can we build a pipeline?" So I was more interested in getting that started and getting other foundations to kick in, to begin to think of building their own evaluation pipeline to address many of the programs and policy programs and the interventions that they thought required more program evaluators of color. So let's build a pipeline. So I did not want to frighten folks away by getting deeply into whether it was "culture" or "race" that should be used. My own writings, my own analysis, I was very conscious of it, but I did not want to be—if I was going to ask another foundation to kick in and further ask them to consider supporting a candidate from their own ranks: "I prefer that you call it whatever you want!" [laughter]
>
> But now I guess it's pulling—it's become critical—every conversation and every fight requires its own timing. I can't say that I was that sophisticated back then in thinking and articulating it this way. But I was much more interested in them supporting this pipeline, developing programs themselves, or supporting my program.

Kien Lee, who co-authored the BDI report, shared similar recollections and concerns.

> … I think the discussion around that was still how can White evaluators work in communities that are different from theirs, particularly, if you define them by race and ethnicity. And so a lot of White evaluators felt like they could do this work if you talk about cultural competency.
>
> The AEA definitely was not comfortable with the word "race." So that was one of the issues that came out of the BDI, that the board that took over at the end of the initiative—not the beginning, but at the end of the initiative—was not as

> comfortable and wanted to talk about culture and diversity and that was the only way that we—those of us who represent communities of color—could really, almost very—in a more subtle way—start to push the field forward. So it was that point of entry to be able to hopefully get the field to talk more and more about this. So I think if you see the term "race," I mean, over time it holds stable but it is slowly increasing, sort of the influx. I also think that at around that time as this took off, then the concept of cultural—not only cultural competency but cultural responsiveness, stakeholder responsiveness, all of those concepts became more prominent and if you talk about responsiveness to your stakeholders and the communities, then the word "culture" captured more of it than just the word "race." So within AEA, there's never been any pushback [from racially otherized groups] around questions of race and I think it's only in the last year or so where that word became a lot more prominent and people started saying, "Hey, it's really about this. Okay? We can't dance around it." And I think it's only the last year or two that that really started to happen.
>
> But by default, "culture" and "diversity" just became, A, the safer words: I don't know if this is true but my speculation is that as people understood also the concept of—as the Building Diversity Initiative brought attention to more of these issues around can evaluators of color and the lack of, I think it also brought forefront questions about like, "Well, what about gender?" Or, "What about sexual orientation?" So I think "culture" and "diversity" sort of became the catch all. And it was also safe. B: Because of the comfort with it, it was a way for those who work in racial justice to, like, "All right. If we can't hit them over their head with it, we can at least begin to interject into the field very gradually and subtly." And then I think the consciousness that there are other kinds of groups besides racial groups that are being discriminated against, I think that just became a broader term.

Importantly, interviewees' strategic use of language was not limited to words that signify racialized difference. More than one parsed out the meaning of and reaction to other words, not specific to racialized difference, that challenged the field's understanding of who it was and what it was expected to do. For example, Hall shared:

> When I chaired the group that ultimately came up with the Cultural Competence Statement for AEA, I actually tried to change the title from "Cultural Competence" to "Cultural Responsiveness," and I got pushback from the AEA Public Affairs Committee and [others]. But we had a grant through Kellogg to write those statements, and after—it took a while, but after we got the proposal approved to do a statement—when I wanted to go from "cultural competence" to "cultural responsiveness," I was forewarned that it might upend the whole approval and we'd have to go through a whole 'nother process. So I decided just to leave it as "cultural competence"…. "[C]ultural competence" is what other organizations were writing at the time. And so when I was trying to sell the idea, I used what they were using,

> but in my own mind, what we wanted was to go beyond competence to responsiveness. But when I got that pushback, I said, "You know, trying to go a bridge too far means the whole thing might not happen." So I settled for what others were doing in terms of cultural competence as language

Another interviewee placed this strategizing within the tradition of resistance struggles more largely.

> Well, I think, I think it kind of depends. One of the things was making a decision about how to keep people engaged in the conversation without them shutting down. And we've always known that, I think, probably. Probably the best example of that is probably Dr. [Martin Luther] King [Jr.]'s work. And so it's—we've always made use of code. Because—out of necessity in some ways. And so you try to find—you have to keep the engagement going in some kind of way. And you want people to be able to hear you. And if they shut down—and I shut down, too. I mean we all do in terms of certain things. And so you have to be able to in some way do that, you know, in terms of…make some decisions about how can…. It's almost, one of the things that I talk about even when I talk about assessment, rather than call it "culturally responsive," sometimes, I mean, what we—we're really talking about is coming up with culturally specific assessments. But people couldn't, I mean, it was just too hard for them to hear that—that language was a little bit too—I don't know if the language was harder, but when you said, "culturally specific" versus "culturally responsive," in some ways, was a way that allowed people to still hear you…. My use of the term "responsive" in this context is very specific because being responsive requires you to DO something. Well, you can talk about being competent—you could be competent and stay at home and do nothing.

Structural factors may have contributed to the trajectories for variations of "culture" and "race." Structural factors raised by interviewees ranged from the national level to the industry level, encompassing AEA and the evaluation field, philanthropy, and training institutions. For example, when asked what might have contributed to the recent increase in variations of "race" that she noticed and brought up in her interview, Lee responded:

> Well, I think the elections had a lot to do with it…. I think the elections had some people going, "We gotta deal with this head-on." I think that it's something that not only the evaluation sector, but other sectors have elevated. And so for those of us working in other areas, whether it's been economic issues, education issues, "health equity" is a word that's catching on so much. I think that term "equity" is just catching on everywhere. I don't think we understand it a whole lot deeper [laughter], but it's catching on, and I think for those of us who really understand the

> issue, it comes down to skin color. So I think that's why, in the last year, in last year's AEA [conference], it just became more prominent.

In addition to pressure coming from national-level discourse, however, interviewees perceived structural factors within AEA and within the nonprofit industry as contributing to the trajectories of variations of "culture" and "race." Lee made the following observations:

> I think also, something not to be underestimated is that over time, those of us who've been involved in the Building Diversity Initiative, after that Initiative and realizing the power that the board has, has been very intentional about making sure that people of color are represented on the board. And so over time, I think the selection of board members that are more prone to talking about race, I think that voice is becoming slowly louder and louder.

She alluded to the shared social group and mutually reinforcing exchange of capital among leaders in the professional association and in philanthropy:

> I think in the foundation sector, I would say only in maybe the last two years, the number of initiatives they're funding around racial equity has also increased. And so I think because those initiatives have increased, the requirement for evaluators who can evaluate that work increased. And I think because of that, the evaluators who are doing this work are also bringing more attention to it. And these folks are folks that are part of the professional association's leadership, so it's kind of this bi-directional relationship, right? So the people who are being asked to evaluate this kind of racial equity work are pretty much the same old people. I mean, there's a core of us that do this work. Then there's the overlap between these individuals and the board membership.

Lee also considered structural factors related to education and pipeline development, echoing the ecosystem analysis she wrote about with Gilbert (2014):

> … If you're not somebody who naturally thinks that way because of your personal life experience, then it will take a lot more to do that. And it's about reading, but it's also about engaging and talking and all of that. And doing. So it's beyond just reading coursework and stuff.
>
> For me, training—evaluators and especially the GEDI folks, to understand culturally competent evaluation and to push it, while a step forward in the field, was a little shortsighted in two ways and continues to be shortsighted in two ways: One is that the training did not get them to think about when you go outside of this individual level competency and this community, this family, as they call—as we all call it, the GEDI family. What is the change you want to see? …And you alone cannot make that change. What does it take to push that change beyond you at an

> organizational level, right? And then, if you think about these organizations who then can—some of these organizations will be on board and some won't. That means at the systems level at a bigger level…. And what is the role of AEA? Influence?

Potential implications of using variations of the terms "culture" and "race" for otherized groups and liberation movements. Interviewees carefully considered the limitations and possibilities, from the perspective of justice, that are offered by specific language and roles for AEA in relation to that of funders and training institutions. For example, Millett shared:

> The way I'm understanding the current use of "culture" versus "race" or "race" versus "culture," "culture" seems—it allows the user of the term "culture" to include in that category people of different races, people who share a certain set of standards or values or educational attainment, appreciation for the rules of the game, and they have mastered it. "Culture" allows the users of this term to differentiate between, for example, Blacks who have attained a PhD or a law degree or doctoral degree or living a cultural lifestyle that approximates, aligns very much with Whites or any other category. So it erases the driver of race as a determinant of success, reduces the presumed effects of race on opportunity….
>
> So these persistent people at the bottom of the ladder here don't share certain kind of cultural attributes that those of us who are more successful share. So it is a polite way to say that opportunity and status differences in this country are not attributable to race but rather to this kind of a…cultural exposure, attributes, which I think is nonsense because it fails to recognize that if the opportunity exposure covers all citizens, then the accumulation of poverty indexes would not be focused at all on—interesting enough, in this culture, even the culture thing fails to attribute, under that umbrella, White folks on the worse side of the socioeconomic indicator to those—so we're just not being honest with these uses.
>
> Do I think ultimately it will get us there? Quite honestly, no.

Like Millett and others, Lee saw important limitations to variations of "culture." As with Hall, Covert, and others, her considerations related to intersectionality and power, with ramifications for the field's professional identity and training:

> I don't think the competency is about culture. The competency is about understanding structural inequities, understanding systems, understanding how they have shaped where we are today, and knowing that the outcomes you see have been shaped by all these economic, political, and social factors that have brought people

> to where they are today. And being able to sift through all of that to recognize what is an outcome, whether it's an outcome or a lack of outcome, but it's due to structural issues versus poor program implementation, or really good program implementation.

Lee went on to explicate the perceived risks of focusing on white supremacy—a primary driver of socio-eonomic inequality that works with and through other systems of oppression—for intersectional analyses. This perception arises from confusion about the meaning of intersectionality—understanding it as *multiple identities* as opposed to a legal response to *interlocking* forces of *oppression*.

> It's a weird tension because I think there's definitely more possibility in becoming more and more explicit. You almost have to become more and more explicit if we're going to get to justice. It's [three] things that we have to become very, very clear about, you know—evaluators…. One is that, the importance of recognizing intersectionality. Right? That none of us are made up of a single social identity. And so if you recognize that, then being more and more explicit about racial justice risks diminishing the impact of multiple social identities that we all comprise, and therefore multiple communities that we all belong to.
>
> At the same time, the second thing is that we have to recognize and be able to articulate, at the same time, that despite the multiple social identities, despite multi-sectionality, race is probably very, very salient, and all evidence point towards that, right? So you can have a Black family and a White family of equal income, and you still have disparities, right? We know race is really salient, and this is a conversation that happens in any field, including in health when we talk about health equity. So then it's recognizing that. So it's trying to figure out how do you actually articulate not only the concept conceptually, but also how it impacts us and our work. I don't think anyone's been able to articulate it well.

Lee began to question the role of evaluators and evaluation as a field in the context of power.

> The third part of it is that all of this means—what does all of it mean? And what it comes down to is power. And so we don't really talk about power because, at the end of the day, evaluators have a lot of power on one hand, because they have the power to be the ones that hear what people are saying, to interpret the findings, and to share the findings on one hand. On the other hand, really, as evaluators, if you identify with that profession, you don't have the power to make social change. The power lies in those who are making decisions about the programming. Sure, you can influence that. But there's only so much you can influence…. And we think we do, which is part of the issue that I think—the tension that is going into these

> concepts around evaluation and how we can promote social change and all of that…. So much attention's going into that that it's taken away the depth and thinking about how evaluators are trained and what they're really supposed to do.
>
> And if evaluators really want to change the world, we've got to step outside of that evaluation profession and think about ourselves as social change agents. And think of ourselves as community organizers. And if we have to do that, that means we have to expand our skill set. That means evaluation no longer becomes about technique, about methodology. It becomes bigger than that. And that's not something many of us think about a lot or can articulate really well. It's certainly not something the profession has even discussed.

With that broader role, Lee directed attention pointedly at the field's professional association and its relations with philanthropy, government, and higher education institutions—specifically, the implications for funding and training institutions:

> And so as a professional association that actually has a lot of influence, right, where are they in the change situation? Or does it become this sort of—this nice little GEDI program here on the side…. This nice little program here on the side for people of color to deal with these issues on the side.
>
> And that's what I was trying to really convey in that [2014] article, was are we even looking at this at a higher level beyond this program? It's much bigger. It's a structural issue and institutionally, AEA is not doing anything.
>
> … [I]f AEA were to be able to step up, then there's all those people who fund evaluation. They are not going to do any different unless the professional association pushes that—right?—and the professional association is not going to do anything different unless the funders who fund evaluation demand something different….
>
> I think the next step is trying to figure out how do you train evaluators to develop that critical thinking beyond just—it's not just about techniques and methods. That's really where we have to go. That's a conversation we have to push…. And if it's not just about technique and methods…then what is it about? What is the identity of this profession?
>
> … So in many ways, in order to advance, we have to shake up the professional identity and struggle with it. Really struggle with it. Because we should be—and it may be figuring out how evaluators and the profession would partner more effectively with other fields—right?—and have that bi-directional influence and support because folks in education and evaluators of education initiatives have to figure out how to push education equity. Right? So the evaluators have to understand education from that perspective and work with those that are

> implementing educational initiatives. The same thing with health and the same thing with—and I think what—I don't know how it will work but I can't see evaluation—if we're really going to be a participant in social change, how we can disconnect ourselves from these fields. We almost have to be integrated even more into them. Yeah. I haven't thought more about that other than I know that we have to shake up what we really mean by evaluation and how to practice it. We really want this profession to be able to be an active participant in social change.

Not all interviewees were sure that evaluation training had a role to play, however. One said:

> Now, if you want to say, "Okay, …if you come into evaluation…much of what we do deals with populations that are in need of resource and development, then we probably need a better understanding of why that is…." But I don't think we're under the obligation to teach American history. So how do you parse out what piece of that evaluation should own and represent? It may be an open question to the organization, to the TIG on teaching evaluation. How much should we acknowledge that as a responsibility and then deal with it?

Use of variations of "indigeneity" and "colonization" in relation to establishment of the IPE TIG. The annual frequencies for variations of both "indigeneity" and "colonization" in U.S. evaluation literature, and especially within literature affiliated with AEA, increased especially in 2003 and 2005 as well as 2012 and 2018. The factors potentially contributing to, interpretations of, and implications of these trajectories and coinciding meso-level, institutional decisions and processes according to those most closely involved at the time are described below.

Internalized, interpersonal, and institutional factors may have contributed to increases in the frequencies for variations of the terms "indigeneity" and "colonization." Few interviewees beyond those who identified as members of indigenous groups spoke directly to the trajectories regarding indigeneity and colonization. One interviewee who identified as indigenous said she saw the establishment of the IPE TIG as pivotal to the development of literature addressing both indigeneity and colonization.

> … I do believe that the development of the Indigenous Peoples TIG introduced that language into AEA because I don't think they had used that term that much. And it kind of built that understanding about seeing people from a perspective of being indigenous, and not necessarily American Indian, but indigenous peoples. So I

> think that had some influence maybe on evaluators and probably from their thinking about how they were going to write about it and use those terms.

Structural factors may have contributed to the trajectories for variations of "indigeneity" and "colonization." The interviewee pointed to structural factors tied to larger dynamics of systemic oppression—namely the material distinctions in the ways that racially otherized groups are oppressed under white supremacy—that played out within AEA. The label of "multi-ethnic" might have given racially otherized groups more presence and power relative to Whites, as discussed by Hall with respect to both "minority" and "diversity" and "inclusion." It also served to mask material distinctions among survivors of settler-colonization, enslaved labor, and imperialism, however, leading some members to establish a distinct IPE TIG that subsequently seems to have influenced the language used in U.S. evaluation literature.

> I will say that, oddly enough, we established the Indigenous TIG—we had a lot of people within the Minority or what we would call now—Multi-ethnic TIG. There was some leadership in that group that was very discouraged by the idea that we wanted to form an Indigenous People's TIG. And that was all very politically motivated. There was a fear that they had made this progress through the Building Diversity Initiative. And here we were, branching off and would maybe—which they perceived as branching off, which really wasn't what it was meant to be. But the perception was that it would allow the dominant powers to play wedge issues. And that was by far not the majority I think. But it was a vocal group.
>
> And it was just—we did it, and one member of that group did call me, pretty unhappy about it. But I think that at that time, that even minority people were having to understand the difference between thinking racially and thinking culturally. So I think it stirred all that pot up, but, I think, overall, for the health of the evaluation organization. And then, of course, I think the development of CREA certainly intensifies the kind of attention to the idea of culturally responsive, using that term. So I see all of that kind of growing in the [2000s] up to now. I mean I think it's important to understand how you want to view the competency of people to be good evaluators. And the idea of understanding your project, your program, and your participants, in terms of a cultural lens rather than just an ethnic lens or a racial lens, I think kind of is a healthy growth that we see in evaluation thinking.

Potential implications of using variations of the terms "indigenous" and "colonization" for otherized groups and liberation movements. One interviewee who did not identify as indigenous, but did identify with a racially otherized group, expressed some

indignation at the decoupling of indigeneity and colonization that they tied to larger structures, potentially beyond AEA.

> I think that this notion of indigeneity without referring to colonization is, again, trying to be culturally competent and admit that there are people who are indigenous to any place…. [T]o be writing about [the notion of colonization] is, basically, [to] have an activist standpoint and require both White people and their multiple gatekeepers to admit that colonialism is driving what we're seeing. It's not part of the calculus so it's not surprising not to see it.

The only interviewee to identify as a member of indigenous groups spoke about the distinction between indigeneity and colonization, and the increases in both, somewhat differently, however. Echoing the Editors' Notes to the 2018 *New Directions for Evaluation* edition on Indigenous Evaluation (Cram, Tibbetts, & LaFrance, 2018), in which indigeneity was both considered globally and affirmed as existing outside the confines of colonization, she said,

> "[C]olonization" is not as big a term that was used, I think, in Indian Country because here we're dealing with issues of sovereignty, and defining ourselves from a colonialized reference was not a comfortable way to do it. It's like, "This is who we are and this is our sovereignty." However, I have my own lessons …doing some writing and realizing, so if I were to deal with people from a native Hawaiian experience, "colonization" was a very important term for them. So I think "colonization" from the perspective of evaluators that I worked with say from a New Zealand perspective, African and other parts of the world who are indigenous, that row of colonization is very important relatively speaking compared to what I think the term and terminology has been in Indian Country.
>
> Although there is a similar issue in the idea of de-colonializing our thinking. And you will see that and a growing understanding in Indian Country about freeing yourself from really the—not so much from colonialization but from that mindset of seeing things in terms of that Western dominance as being preferred or not even understanding how you're giving it privilege without questioning it.
>
> … I like the use of "indigenous." I mean, I have no problem with that. We are not defined by colonization. We're defined by our own sovereignty within a situation of oppression that occurred, for sure, and in which there are still some real issues to deal with. But I think from an indigenous woman's point of view, I want to be seen as a sovereign person within a sovereign identity related to my Turtle Mountain Chippewa tribal membership. So I don't have a problem with "indigenous" at all.

> And I don't want to be defined in terms of just that colonial reality. So I don't have a problem with it.

Centering colonization as opposed to indigeneity also seemed to present a practical problem that would affect indigneous peoples and liberation movements:

> You don't really be responsive to colonialism. You understand its impact and its effects, but you want to be responsive to culture. So I have no problem seeing that change.

Connecting the comparison between indigeneity and colonization with that between culture and race, and considering larger questions of justice, the interviewee continued:

> [M]y sense is that racism is kind of, to me, similar to colonialism as a force that explains a lot of what you see that's inequitable today, and so I don't think we can ignore it. When it comes to decolonialization or looking at sources such as the funder and that sort of stuff, in some ways I would say when you get right down at the community base, they'd be luxuries because you're just working to improve and you go after what resources are available. To the degree, I think, evaluators can show that there's a tension between what is desired at a community base level and what is imposed by a funding level. When there's tension, I think, we have a responsibility to understand and to give voice to that tension.
>
> But going forward, I think a whole—I think going forward, I have no problem with the understanding of indigeneity.

CST interviews: Interpretation and use of findings. Overall, findings from the CST interviews suggest that choices and decisions made with respect to language within AEA—by those who identified as members of racially otherized groups and by those who did not—were generally conscious, sometimes even calculated and negotiated explicitly, in relation to meso- and macro-level dynamics within and surrounding AEA and evaluation more generally. Interviewees who identified as members of racially otherized groups ranged from having intentionally quelled the focus on racialized difference, on one hand, to prioritizing commitments to action and resources at the expense of language, on the other. This effort to create the space necessary to do the work they thought needed to be done on the ground seemed true of interviewees regardless of their role—for example, in their classrooms, in their board rooms, or in their evaluation practices. While some felt the right decisions about language had been made at the time, most interviewees who identified as

members of racially otherized groups felt the struggle was not yet complete, but ongoing. Aside from the interviewee who identified as indigenous and pleased with the trajectories for variations of "indigenous" and "colonization," none of the interviewees expressed satisfaction with the three other trajectories that were depicted in the visualizations of preliminary findings of the diachronic analyses. Only two interviewees—both of whom did not identify with any racial group—reacted to them positively.

Meta-interpretation and action stimulus. The CST-strand interviewees generally confirmed the relationship between specific meso-level, institutional decisions and processes within AEA and the discourse shifts that appeared in the U.S. peer-reviewed literature. This was especially so for the rise of diversity and inclusion, the replacement of race with culture, and the rise of indigeneity and its decoupled relationship with colonization. Interviewees described many of these institutional-level decisions and processes as directly tied to interpersonal relationships and, simultaneously, directly tied to structural-level dynamics. Institutional mechanisms—for example, AEA TIGs, conference sessions, and published literature—facilitated many pivotal interpersonal relationships, such as the relationship between Melvin Hall and Karen Kirkhart as well as between Ricardo Millett and Donna Mertens.

At the same time, interpersonal relationships facilitated many of those pivotal institutional appointments and subsequent discussions, negotiations, and decisions. For example, the interpersonal relationship between Melvin Hall and Karen Kirkhart, who had an interpersonal relationship with AEA's president-elect, resulted in Hall's appointment as Chair of the Diversity Committee. Indeed, more than one interviewee lamented changes in the pace, management, and structure of the annual conference, which they felt no longer facilitated discussion and interpersonal relationship building. As much as interviewees tied together interpersonal and institutional dynamics, they also tied both to macro-level, structural dynamics. For example, some explicitly questioned the solidarity that those classified as White women may or may not have with members of racially otherized groups as well as the solidarity that racially otherized groups may or may not have with each other;

two tied those questions to the backlash against Affirmative Action, which was waged largely through those classified as White women.

These discourse shifts described in Chapter Five were tied conceptually and chronologically to discursive events within AEA that were identified through archival documents. Interviewees described these discursive events as related to interpersonal dynamics; meso-level, institutional decisions and processes; as well as macro-level, structural discourses and policies (summarized in Figure 31). The trajectory of the field's construction of racialized difference generated from CDA of the evaluation literature, historicized in relation to archival documentation of discursive events within AEA, and contextualized in relation to larger policy debates through CST interviews with scholars and practitioners closely involved in those events, comprised four discursive patterns:

- Minoritization and ambivalence toward whiteness
- The invocation of diversity and inclusion
- The replacement of race with culture
- The rise of and decoupled relationship between indigeneity and colonization

These four discursive patterns characterize the over-arching language trajectories' relationship with the systems surrounding evaluation. The patterns correspond with a liberal as opposed to a critical construction of racialized difference.

Patterns in Evaluation's Trajectory of Racialized Difference	Discourse Shifts	Related Dynamics		
		Interpersonal	Institutional	Structural
I. Minoritization and ambivalence toward whiteness	• late 1980s	• Madison and President Covert	• Establishment of Minority Issues in Evaluation TIG	• Lack of continuity among members unless university-affiliated and funded
	• early 1990s	• Madison	• Publication of AEA-affiliated journal edition on Minority Issues in Evaluation	
II. The invocation of diversity and inclusion	• late 1990s/ early 2000s	• Millett and President Mertens • President Smith, Kirkhart, and Hall	• Kellogg-funded BDI study • BDI implementation	• Anti-Affirmative Action lawsuits filed against higher education institutions by applicants classified as White women
III. The replacement of race with culture	• late 1980s	• President Conner	• International & Cross-cultural Evaluation theme of annual AEA conference	• Racialized neoliberalism manifested in devolution of redistributive programs
	• early 1990s	• President Kirkhart	• Multicultural validity in annual AEA conference address	
	• 2000s	• President Smith, Kirkhart, and Hall • Peak	• Cultural Reading and revised edition of the Standards • Introduction of multiculturalism by Tracking Transformation evaluation of BDI	
	• early 2010s	• Hall	• Membership's appproval of AEA Statement on Cultural Competence	
IV. The rise of and decoupled relationship between indigeneity and colonization	• early 2000s • late 2010s	• Cram, Tibbets & LaFrance • Cram, Tibbets & LaFrance	• Establishment of IPE TIG • Publication of AEA-affiliated journal edition on Indigenous Evaluation	

Figure 31. Summary of interpersonal, institutional, and structural dynamics by pattern of trajectory

Conclusion of Chapter Seven

The CST interviews addressed the construction of racialized difference at multiple levels, but provided the only opportunity within the research design, aside from the search for outside literature to interpret results, to focus on the construction of racialized difference *among institutions* surrounding AEA—including philanthropy, government agencies, and higher education. Interviewees were identified through the results of the textual analysis, diachronic analysis, and archival analysis. From an original list of 80 such scholars and practitioners identified through the literature and documents as being most closely involved in evaluation's construction of racialized difference, 20 were invited to participate, and 11 interviews were completed.

Interviewees historicized and contextualized the shifts in language in relation to interpersonal, meso-level/institutional, and macro-level/structural relations, decisions, and processes. Figure 31 summarizes these findings, and Chapter Eight will discuss each finding from the perspective of critical theories of systemic oppression and system dynamics.

Review of Results

Chapters Four through Seven detailed the findings from the CDA and analysis of the CST interviews, which are summarized below. First, the textual analysis identified peer-reviewed literature that directly addressed racialized difference from the inception of the U.S. field of evaluation in 1977 through 2017. Within this selection of literature, also since the field's inception, one thread constructed racialized difference critically, whether by acknowledging it as produced rather than natural, treating it as relational rather than individual; or historicizing or contextualizing it within asymmetrical structures rather than as fixed. This critical thread was not solid, however, in that the authors had generally not built upon their own or each other's work to produce a coherent body of empirical or theoretical scholarship within evaluation that addressed difference, racialized difference specifically, or systemic oppression. Indeed, aside from Hood, most of the authors had published regarding racialized difference in the peer-reviewed evaluation literature only once between its inception in 1977 and 2017. An important exception to this finding was

the literature that addressed indigeneity and colonization, which not only grew immediately following meso-level, institutional events and decisions, but also coalesced into a coherent body and community of scholars and practitioners.

Textual analysis of the critical thread of literature directly addressing racialized difference also yielded nine terms or phrases whose variations have been used to signify racialized difference within the evaluation literature: colonization, culture, diversity, inclusion, indigenous, minority, of color, race, and white. A diachronic analysis of U.S. peer-reviewed literature containing one or more of those nine terms revealed four over-arching trajectories: (1) consistent use of variations of "minority" as opposed to "of color" and inconsistent use of variations of "white"; (2) a rise in the use of variations of "diversity" and "inclusion"; (3) the replacement of variations of "race" with "culture"; and (4) a rise in the use of variations of "indigeneity" and "colonization," which were decoupled for most of the evaluation literature's history—albeit in different ways during different periods. While the first finding regarding minoritization and whiteness remained fairly consistent throughout the 40-year period that followed the initial increase observed soon after the establishment of AEA, the latter three showed more dramatic change over time.

All four discursive shifts coincided with institutional level decisions and processes within evaluation. These decisions and processes involved AEA conference themes and keynote speakers; journal editions; task forces, committees, and topical interest groups; standards, principles, competencies, and other policies; and programmatic initiatives. Interviews with those closely involved in these decisions and processes tied them materially or conceptually to specific interpersonal, institutional, and structural dynamics in the systems surrounding the professional association. These dynamics ranged from interpersonal connections and relationships instigated and cultivated largely at the meso/institutional level through AEA to macro-level/structural forces like high stakes testing, the privileging of experimental design, and the neoliberal backlash against redistributive policies like Affirmative Action and Equal Opportunity.

Historicized through the archival analysis and contextualized through the analysis of CST interviews, the trajectories of the field's construction of racialized difference were organized into four discursive patterns:(1) minoritization and ambivalence toward whiteness; (2) the invocation of diversity and inclusion; (3) the replacement of race with culture; and (4) the rise of and decoupled relationship between indigeneity and colonization. These discursive patterns are each discussed in Chapter Eight.

CHAPTER 8: DISCUSSION

Chapter Eight starts with a review of the theoretical and analytical frameworks through which each discursive pattern identified in Chapter Seven is discussed. Using concepts from critical theories of systemic oppression and system dynamics, each of the four discursive patterns is discussed in relation to the areas of contrast between critical and liberal constructions of difference.

Organization of Chapter Eight

Following a review of salient concepts, discussion of each discursive pattern includes a brief summary of the findings from the textual analysis—the specific terms that have been used to signify racialized difference. Second, it includes a brief summary of the diachronic analysis findings—the intertextual changes over time. Third, it includes a summary of the archival analysis and CST interview findings—the interpersonal, meso-level/institutional, and macro-level/structural level dynamics that coincided with the field's discourse shifts. These dynamics include discursive events and mechanisms—decisions and processes—within and around AEA. They constitute structurally mediated flows, exchanges, or feedback loops of various types of capital, facilitated through various media of exchange—or currency—including whiteness and discourse (Ahmed, 2007a, 2007b).

Review of Theoretical and Analytical Frameworks

Chapter One contrasted liberal constructions of difference from critical constructions of difference. In the former, particularly in a climate of racialized neoliberalism, difference is constructed as natural and individual while structures are obscured. In the latter, difference is constructed as produced and reproduced over time through structurally mediated feedback loops. These feedback loops involve exchanges of capital through socio-economic relations and decision-making processes that are

characterized by super-ordination and sub-ordination because the structural mechanisms mediating them are asymmetrical.

These areas of contrast between liberal and critical constructions of difference include whether difference is:

- Natural or produced
- Individual or relational
- Lateral or asymmetrical
- Static or cumulative

Enabling and distributive conceptualizations of justice. Chapter One also introduced an enabling conceptualization of justice in contrast to distributive conceptualizations of justice. Enabling conceptualizations of justice consider the extent to which conditions enable otherized social groups to *exercise* collective agency to create more symmetrical structural *dynamics*. In distributive conceptualizations of justice, financial and other types of capital are *static* and *held*. In enabling conceptualizations of justice, justice is the obverse of oppression.

Review of Theoretical and Analytical Concepts

Both critical theories of systemic oppression and systems thinking are thus concerned with dynamics and relations—the exchange or flow of multiple types of capital—across actors within a system through structural mechanisms. This dissertation has argued that racialized difference is relational and that relations are structurally mediated. Difference is discursively and materially (re)produced through the compounding of capital exchanged through asymmetrical relations, cumulatively over time and interactively across actors. Discussion of results in Chapter Eight thus rests on an understanding of the concepts of (1) difference, (2) boundaries, and (3) relations.

Difference. Critical theories of systemic oppression offer a framework for making meaning of perceived difference. Under such theories, difference is (re)produced dialectically. Racialized difference, in particular, is (re)produced through dialectical relations centered on whiteness. Under Smith's (2016) three pillars of white supremacy (illustrated in Figure 4), Blackness is (re)produced as property in need of the ownership

associated with whiteness; indigeneity is (re)produced as savagery in need of the civilization associated with whiteness; and Asianness is (re)produced as alien in need of the belonging associated with whiteness.

Boundaries. Models like Reskin's (2003) Mechanisms-based Model of Ascriptive Inequality offer multi-level frameworks for analyzing the process of racialization. This model (illustrated for the nonprofit industry surrounding evaluation in Figure 3) includes the four levels of racialization listed below:

1. Internalized: dynamics *within* individuals (implicit biases regarding racially otherized groups among individual leaders, staff, and evaluators—regardless of racial group identification—affiliated with the institutions listed below, resulting from the repeated association of whiteness with authority, ownership, and belonging; explicit biases to advance social group interests—as in cases of Whites wanting to leave desegregated schools or neighborhoods)
2. Interpersonal: dynamics *among* individuals (individual leaders, staff, evaluators affiliated with the institutions listed below); interpersonal dynamics can amplify or counteract individual interests in maintaining or ending racial inequality
3. Institutional: dynamics *within* institutions (professional associations, nonprofit organizations, evaluation firms); institutional dynamics can amplify or counteract individual interests in maintaining or ending racial inequality
4. Structural: dynamics *among* institutions (multi-lateral agencies, governments and government agencies, educational systems, philanthropic foundations); structural dynamics can amplify or counteract institutional interests in maintaining or ending racial inequality

Relations. Structural mechanisms shape the nature of decision-making processes and socio-economic relations. Chapter Seven included quotations from two interviewees who elaborated on their social group identification by referring to a common humanity. One said, "In the end, we're all human beings, and we all bring resources, experiences, knowledge, insights." The other said, "The phenomena that we're looking at has something to do with humans, and we are human." Both challenged the dichotomy

between program participants and program evaluators. Both spoke about culture. Both spoke about systemic issues such as oppression and discrimination. Their responses differed, however, in that one described her work as conveying respect by "*coming to be more in relationship* rather than us and them" whereas the other described her work as "*shift[ing] our relationship*" because we all "live within," or have a "role" in the "phenomena we're trying to observe" (emphases added).

The nuance pivots on interpretations of the word "relational." One interpretation of it is involving interpersonal relationships. Another is acknowledging existing relations. Program participants and program evaluators are already structurally related through exchanges of various types of capital in the industries surrounding the evaluation and in the larger economy and society; the structures mediating these exchanges are asymmetrical; and that asymmetry is racialized (as well as gendered, classed, ableized, etc.). In much the way that a third interviewee described how the binary between insider- and outsider-status made little sense in Indian country, relationality could be understood as explicitly placing oneself in relation to others, by naming the nature of the relation:

> [I]n that placement role or establishing…I come in right away with elements of my identity that we can talk about…. It's family and community, who you're related to, and *do you know who you're related to*" (emphasis added).

Indeed, feminist psychoanalysis posits relationality as the original state of personal identity—because all human beings begin life in-relation, in utero and then as infants still attached to their mothers, as opposed to coming to be in relationship with them (Hollway, 2010).

Such an interpretation of relationality requires that not just individual evaluators, but—more importantly—evaluation as a field, meaning institutions such as AEA, place themselves in relation to other actors in the surrounding system by naming the nature of the relation: What is the dynamic? To what extent is the pattern of exchange symmetrical? To what extent is it asymmetrical?

Young's (2011) Five Faces of Oppression offers language for naming how dynamics are asymmetrical. They are summarized below.

- Exploitation describes socio-economic relations whose asymmetry lies in the systematic transfer of capital produced by one group's land or labor to another.
- Marginalization describes socio-economic relations whose asymmetry lies in the systematic exclusion of groups from the exchange of capital.
- Powerlessness describes decision-making processes whose asymmetry lies in the infantilization of some groups by others.
- Violence describes decision-making processes whose asymmetry lies in the discipline of some groups by others.
- Cultural imperialism describes cultures whose asymmetry lies in the otherizing of some groups and normatizing of others.

Flows, stocks, and feedback loops. Flows are exchanges, relations, processes, dynamics, or activities—verbs that take place over time. Flows update the magnitude of stocks. Stocks are quantities of resources, assets, or capital that are measured at specific times, although they have necessarily accumulated and been drawn from (through in-flows and out-flows) over time. In-flows fill or add to stocks. Out-flows drain or subtract from them. The distinction between flows and stocks is pivotal to the distinction between an understanding of power and privilege in the form of access to and control over multiple types of capital as dynamically *exercised*—corresponding with an enabling conceptualization of justice—rather than static and *held*—corresponding with a liberal conceptualization of justice. Such an understanding is necessary to analyze how difference is (re)produced through asymmetrically structured flows and to re-arrange structures that channel flows of capital more symmetrically and in ways that enable collective action among otherized groups.

Feedback loops. Flows can change the magnitude of a stock in ways that influence the rate or nature of the original flow or exchange, making subsequent change in magnitude either more or less likely. These are feedback loops. In positive feedback loops (see Figure 32), the direction of growth is positive; initial changes become amplified or magnified over time. Achievement of the product encourages the process that created it. Positive feedback loops reinforce or amplify some phenomenon that a

cause helped effect. By contrast, in negative feedback loops (see Figure 33), the direction of growth is negative; initial changes are counteracted or balanced out so that conditions remain relatively stable. Achievement of the product inhibits the process that created it. Negative feedback loops balance or counteract some phenomenon that a cause helped effect.

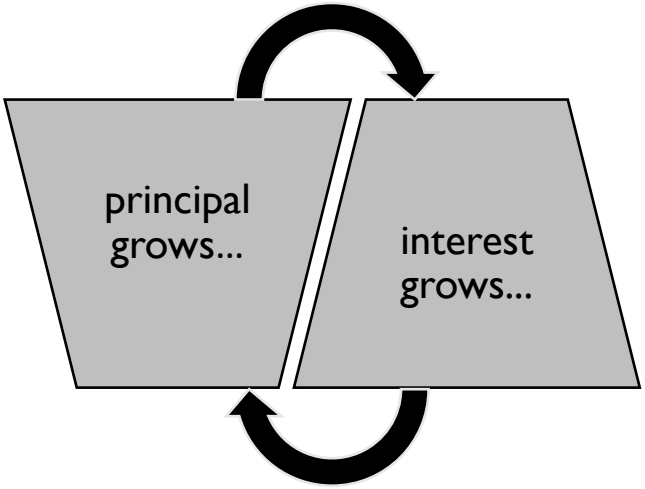

Figure 32. Positive Feedback Loop

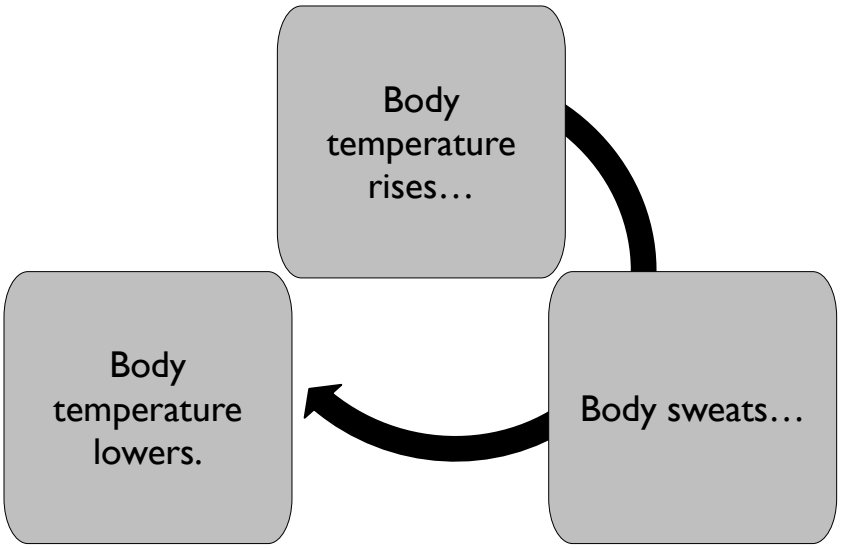

Figure 33. Negative feedback loop

Flows and feedback loops offer ways to think about the structural mechanisms underlying a dynamic's asymmetry. What is the history of that dynamic's underlying

structure? How could that structure be shifted to counteract rather than amplify the asymmetrical pattern? To what extent do the structural mechanisms amplify otherized social groups' efforts to exercise collective agency to create more symmetrical dynamics? To what extent do they counteract such efforts? The former would be considered enabling and just, while the latter would be considered oppressive and unjust under Young's (2011) conceptualization of justice.

Capital. Capital is the total sum, store, or stock of accumulated contributions—past flows invested in (appreciation), minus flows out (depreciation)—of some valuable, positive, or advantageous quality that allows efforts to operate. Ecological and cooperative economics proffer multiple types of capital (Šlaus & Jacobs, 2011; see Figure 34 for some examples), beyond financial capital, to encourage thinking about capital in terms of relations and processes as opposed to the more static, dehistoricized, and decontextualized idea of "money." Thinking in terms of multiple types of capital not only makes the commodification—and theft—of nature, human beings, knowledge, and culture more visible, but it also illustrates the sources underlying the agency that otherized groups exercise.

1. Natural capital is the breathable air, stable climate, flood protection, and water purification that planetary systems and local ecosystems provide. • Natural non-living capital consists of raw and processed elements of nature, like stone, metal, timber, and fossil fuels.
2. Relational capital arises from the number, strength, and nature of social, economic, political, and professional relationships. • Human capital consists of individual human beings' capacity to apply themselves to a purpose beyond their individual survival.
3. Learned capital consists of the systems and processes for sense- or meaning-making. • Created capital consists of the repertoire of work that people engage other capital to cultivate and create.
4. Structural capital includes facilities, equipment, technology, and supplies. • Technological capital is the artistic and scientific processes that transform capital.
5. Cultural capital is generated by a community to support the restoration and regeneration of other capital. • Spiritual capital connects individual lives/lifetimes with a larger universe of lives/lifetimes.
6. Economic capital accounts for the economic effects an organization has on other entities.

- Financial capital represents other types of capital, providing purchasing power and facilitating social, economic, professional, and political activity.

Figure 34. Types of primary and secondary capital derived from ecological economics

Chapter Two discussed whiteness as a form of currency—a medium through which capital is exchanged. Discourse is another. As a form of currency in the exchange of capital (whether relational, human, learned, or cultural), discourse compounds cumulatively and interactively. Ahmed described this connection between the discursive and the material as follows (2006):

> … [A] new approach to the relation between texts and social action…still considers texts as actions, which "do things," but it also suggests that "texts" are not "finished" as forms of action, as what they "do" depends on how they are "taken up." To track what texts do, we need to follow them around. If texts circulate as documents or objects within public culture, then our task is to follow them, to see how they move as well as how they get stuck. So rather than just looking at university documentation on diversity for what it says, although I do this, as close readings are important and necessary, I also ask what they do, in part by talking to practitioners who use these documents to support their actions.… (p. 105)

Typically in a capitalist economy, capital is used to create more value as opposed to being consumed. An understanding of capital rooted in critical theories of systemic oppression or systems-oriented fields like ecological economics offers a framework for asking about any socio-economic flow or exchange of capital: What kinds of capital have historically been—and are now being—exchanged between the social groups that each actor involved represents? For whom does the exchange create value? At whose expense?

Patterns in Evaluation's Construction of Racialized Difference

Through its material and discursive role related directly to the flow of knowledge, financial, and other types of capital in racially stratified industries, the field of evaluation inevitably engages with the construction of racialized difference (House, 2017). Chapter Seven reported four discursive patterns identified through CDA and CST interviews. The first three of the four discursive patterns identified, all of which are listed below, suggest that the constructions of racialized difference reflected in and reproduced through evaluation are largely liberal as opposed to critical:

- Minoritization and ambivalence toward whiteness
- The invocation of diversity and inclusion
- The replacement of race with culture
- The rise of and decoupled relationship between indigeneity and colonization

Each of these is discussed below in relation to the above concepts from critical theories of systemic oppression and system dynamics.

Minoritization and ambivalence toward whiteness. It will become necessary for any future analyses that may consider the frequency for variations of "minority" in the evaluation literature to code items containing "minorit*," because Fiona Cram introduced the critical term "minoritize"—which is common in other disciplines—into the evaluation lexicon in 2018.[56] Minoritization refers to the minority-making process through which otherized groups—regardless of relative size—are diminished.

Difference as relational rather than individual. Covert's use of the word "minority" and "minorities" in his paragraph about the extent to which evaluators who identified as "Black," "Asian," "Hispanic," and "Native American" were represented in AEA, but not in the paragraphs about the extent to which international members and graduate students were represented, reflected usage common at the time, in which variations of the word "minority" implicitly referred to racial minorities rather than simply referring to any numerical minority—for example—minorities from the perspective of gender and sexuality. Indeed, "minority" is still often used as a synonym for members of racially otherized groups. Its use suggests that the generally lower status of racially otherized groups relative to groups racialized as White is simply a natural result of their smaller size.

This suggestion is problematic in two respects. First, it ignores the fact that globally, racially otherized populations are the majority. Whether globally or in countries like South Africa, however, the smaller size of the population racialized as White does

[56] While items containing variations of "minority" that were included in this dissertation's diachronic analysis were coded neither for relevance nor for substantiveness, the analysis did confirm that no items before Cram's in 2018 include variations of the term "minoritize."

not correspond with lower status. In other words, it is not size that determines the relative status of groups. Second, it is problematic in that it ignores the fact that the USA's "White majority" was artificially produced and is continually reproduced through settler colonialism, including the genocide and sexual violation of indigenous populations, anti-miscegenation laws, and exclusionary immigration policies that explicitly limited entry and naturalization of nationals from countries outside Europe to maintain the proportion of settlers of European ancestry relative to those of Asian, African, and Central or South American ancestry as well as to indigenous peoples. Ignoring these facts constitutes the process of minoritization.

Structurally mediated relations. Overall, U.S. peer-reviewed evaluation literature tended to construct racialized difference in terms of "minorities," as opposed to people "of color." Small increases in the annual frequencies for variations of "minority" within AEA-affiliated literature coincided with meso-level, institutional discursive events and mechanisms within AEA—specifically, the development of the MIE Topical Interest Group (TIG) and publication of an AEA-affiliated journal edition devoted to Minority Issues in Evaluation. However, they remained fairly consistent relative to the annual frequencies for variations of "people of color" and "white" and did not have a visible effect on the U.S. evaluation literature at large. Additionally, annual frequencies for variations of "minority" have varied minimally over time or by AEA affiliation. In fact, they held steady (from 20% to 21% between 2004 and 2018 in AEA-affiliated publications) even after the Minority Issues in Evaluation TIG changed its name to the Multi-ethnic Issues in Evaluation TIG in 2004, soon after implementation of the BDI began. As reported in Chapter Seven, interviewees described this name change as an attempt to broaden the TIG's membership beyond evaluators who identified as African American or Black, as "minority" was believed to suggest.

The relationship between variations of "minority" and "of color" provides necessary insight into the three remaining patterns identified in evaluation's construction of racialized difference. While references to "colored people" are centuries old, adoption of the phrase "of color"—although not without critique—represents a relatively recent

attempt among members of different racially otherized groups engaged in liberation movements to work in solidarity together toward gender justice. Veteran organizer Loretta Ross, from Sister Song Women of Color Reproductive Justice Collective, explained the origins of the phrase "women of color" as follows:

> In 1977, a group of Black women from Washington, DC, went to the National Women's Conference…as part of the World Decade for Women…. This group of Black women carried into that conference something called "The Black Women's Agenda" because the organizers of the conference—Bella Abzug, Ellie Smeal, and what have you—had put together a three-page "Minority Women's Plank" in a 200-page document that these Black women thought was somewhat inadequate. So they actually formed a group called Black Women's Agenda to come to Houston with a Black women's plan of action that they wanted the delegates to vote to substitute for the Minority Women's Plank that was in the proposed plan of action. Well, a funny thing happened in Houston: When they took the Black Women's Agenda to Houston, then all the rest of the "minority" women of color wanted to be included…. [The Black women] agreed…but you could no longer call it the "Black Women's Agenda." And it was in those negotiations in Houston [that] the term "women of color" was created. And they didn't see it as a biological designation—you're born Asian, you're born Black, you're born African American, whatever—but it is a solidarity definition, a commitment to work in collaboration with other oppressed women of color who have been "minoritized." Now, what's happened in the 30 years since then is that people see it as biology now…. And people are saying they don't want to be defined as a woman of color: "I am Black," "I am Asian American"…and that's fine. But why are you reducing a political designation to a biological destiny? That's what white supremacy wants you to do. And I think it's a setback when we disintegrate as people of color around primitive ethnic claiming. Yes, we are Asian American, Native American, whatever, but the point is, when you choose to work with other people who are minoritized by oppression, you've lifted yourself out of that basic identity into another political being and another political space. And, unfortunately, so many times, people of color hear the term "people of color" from other White people that [they] think White people created it instead of understanding that we self-named ourselves. This is a term that has a lot of power for us. But we've done a poor-ass job of communicating that history so that people understand that power. (Western States Center transcript of online videorecording, 2011)

Difference as asymmetrical rather than lateral. While annual frequencies for variations of "minority" were consistently higher in the U.S. evaluation literature than they were for variations of the phrase "of color," what did vary considerably by AEA

affiliation were the overall and annual frequencies for variations of "white." In AEA-affiliated publications, overall frequencies for variations of the phrase "of color" were as low as they were for "white" (5% and 6% overall, respectively), and annual frequencies were consistently parallel from the field's inception through 2018. In U.S. evaluation literature unaffiliated with AEA, however, overall frequencies for variations of "minority" and "white" were the same (17%), and annual frequencies were nearly parallel from the field's inception through 2018. Corresponding with the finding that "minority" might be considered a synonym for "Black" or "African American," the parallel treatment of "minority" and "white" illustrates the dialectical relationship between "Black" and "White." The parallel treatment also corresponds with the finding that literature unaffiliated with AEA containing variations of "race" tended to treat it as a static attribute affixed to individuals, by which reported studies were often disaggregated, as opposed to a topic of discussion.

Difference as cumulative rather than static. Further interpretation of this pattern of minoritization and ambivalence toward whiteness in evaluation discourse requires examining

> … how institutions become white [*sic*] through the positing of some bodies rather than others as the subjects of the institution (querying, for example, who the institution is shaped for and who it is shaped by). Racism would not be evident in what we fail to do, but what we have already done, whereby the "we" is an effect of the doing. [This type of] recognition of institutional racism…*reproduces the whiteness of institutions by seeing racism simply as the failure to provide for nonwhite others because of a difference that is somehow theirs*. (Ahmed, 2006, pp. 107, emphasis added)

Amplifying relations of violence. As reported in Chapter Four, Stanfield's 1998 keynote address at AEA's annual conference (printed in the AEA-affiliated *American Journal of Evaluation* in 1999) was the first documented instance in the evaluation literature in which "white" referred not to an attribute of individuals, but rather to a system—specifically, white supremacy. From 1977 through 2016, Ward Hood and Cassaro's 2002 article entitled "Feminist Evaluation and the Inclusion of Difference" was the only other piece to have used the phrase "white supremacy." Importantly, however,

five items containing the phrase were published between 2017 and 2018. This sudden increase in 2017 and 2018 might have been in response to violent national discourse, as interviewee Kien Lee is reported in Chapter Seven as having suspected might underlie the increased frequency of variations of "race" in U.S. evaluation literature during approximately the same period.

The invocation of diversity and inclusion. As reported in Chapter Five, despite having entered the field's lexicon only in the late 1980s and early 1990s, respectively, the percentage of U.S. peer-reviewed evaluation literature containing variations of "diversity" (32%) and "inclusion" (22%) exceeded the percentage containing variations of any of the terms or phrases that potentially signify racialized difference more specifically, perhaps not surprisingly. The latter included "race" (16% in U.S. literature as a whole, with wide variation between AEA-affiliated and -unaffiliated literature); "minority" (14% in U.S. literature as a whole); "white" (11% in U.S. literature as a whole, with wide variation between AEA-affiliated and -unaffiliated literature); and "of color" (8% in U.S. literature as a whole). While the overall frequencies for variations of "diversity" and "inclusion" were higher in the U.S. literature as a whole and in U.S. literature unaffiliated with AEA relative to that affiliated with AEA, the sharpest sustained increase in annual frequencies was in AEA-affiliated literature. That increase coincided with meso-level, institutional discursive events and mechanisms within AEA—namely, the BDI study, which started in 1999, and its implementation, which started in 2001. The potential relationship between these events and the increase in annual frequencies for variations of "diversity" are discussed below, after brief consideration of the term "inclusion."

Results for "inclusion" reported in Chapter Five are not considered in depth in Chapter Eight for the following reasons. First, unlike "colonization," "culture," "race," and "white," items that contained inclusion (and those that contain diversity) were not coded for relevance or substantiveness. Second, the increase in annual frequencies for variations of the term "inclusion" was very gradual as opposed to representing a sudden discourse shift. Third and relatedly, no seminal items in the literature or in AEA's

documents could be identified through the archival analysis that corresponded with fluctuations in the term's trajectory. Finally, no interviewee picked up the thread of inclusion as a point of discussion in the way that they picked up diversity. It is possible that the increase in annual frequencies for variations of the term "inclusion" was simply a function of the now-routine phrase "diversity and inclusion" as opposed to any specific meso-level discursive event or mechanism within evaluation or AEA. While this pattern regarding diversity and inclusion was thus somewhat specific to "diversity," "invocation" is intended to convey how the phrase "diversity and inclusion"—currently often heard as "diversity, equity, and inclusion"—reflects and taps into national discourses and policies of racialized neoliberalization.

Difference as relational rather than individual. As reported in Chapter Four, considerations of diversity, like culture and inclusion of under-represented groups, were introduced into the evaluation discourse and AEA, in particular, in the late 1980s (Conner, 1989a, 1989b, 1989c; Ellmer & Olbrisch, 1983; Mark & Shotland, 1985; Patton, 1985; Roberts & Attkisson, 1983). For example, Conner's presidential address was entitled "In Praise of Differences" (1990). During this period, the evaluation field began openly seeking greater and earlier involvement in programs' funding and design (Hendricks, 1985). Evaluation's relationship with philanthropy also began tightening, explicitly in response to Reagan-era reductions in government spending:

> …[A]s resources usually available for evaluation have begun to dwindle and the federal government has begun encouraging increased private sector support for social programs[, e]valuation researchers are looking more widely for financial support and are asking whether foundations might be a source of evaluation funds. (N. L. Smith, 1985)

At the same time, philanthropic foundations sought greater legitimacy in response to pressures from results-oriented donors and trustees by demonstrating their knowledge of such "best practices" as evaluation (Behrens & Kelly, 2008; Hall, 2004).

Difference as asymmetrical rather than lateral. What distinguishes the earlier efforts to increase the representation of evaluators from racially otherized groups, consider internalized biases, and include stakeholders in the evaluation process (e.g.,

Covert, 1987b; Madison, 1992) from later streams of discourse is that the latter were heavily predicated on the incongruence between program participants and program evaluators and the value that evaluators representing racially otherized groups could bring to the quest for valid data. In contrast, the former were predicated on fairness and the intrinsic value of heterogeneity. This trajectory corresponds with the larger trend identified in the USA. The origins of representativeness as an issue of fairness lie in compliance with Equal Employment Opportunity and antidiscrimination laws set forth by the Civil Rights Act of 1964. Affirmative Action represented a response to legal requirements as well as "a moral imperative for organizations to more proactively recruit individuals from traditionally underrepresented groups to remedy past discrimination" (Weisinger, Borges-Méndez, & Milofsky, 2015, p. 9S). It is widely acknowledged to have benefited those classified as White women more than those racialized as "other" (Weisinger, Borges-Méndez, & Milofsky, 2015).

In 1987, the Hudson Institute projected that "by the year 2000, more than 80% of the net new entrants into the U.S. workforce would be women, non-Whites [*sic*], and immigrants" (Weisinger, Borges-Méndez, & Milofsky, 2015, p. 10S, citing Johnson & Packer, 1987), leading employers to look beyond the legal compliance requirements of EEO and Affirmative Action toward more proactive ways to "manage" this workforce diversity. As opposed to the fair employment emphasis of EEO, the early emphasis in diversity was on recruiting individuals to increase heterogeneity. Once organizations began to diversify without changing the status of under-represented groups, organizations' emphasis shifted to celebrating differences. Justifying the role that new, "diverse" organizational members could play in reaching previously untapped markets, or in fostering creativity and innovation, led to the so-called "business case" for diversity that permeates many organizations today across sectors and around the world (Weisinger, Borges-Méndez, & Milofsky, 2015, p. 10S). Much like Stanfield's (1999) comments reported at length in Chapter Four with respect to minoritization and whiteness, however, "the business case does not adequately take into account intergroup inequality, often confuses cultural differences with inequality differences, and frequently relegates

historically excluded groups to positions where they work only with those from their own community" (Weisinger, Borges-Méndez, & Milofsky, 2015, p. 13S).

Amplifying relations of cultural imperialism. Moreover, for difference, identity, and even cultural competence to generate value, whether measured financially or in terms of efficiency (Hood, 2001), they must be treated as quantifiable and commodified characteristics affixed to or residing in individual evaluators as opposed to representing political identification with social group interests. Such a location advances liberal as opposed to critical constructions of difference. For example, as reported in Chapter Four, in showing how students with ancestries outside Europe or who speak languages other than English (among other characteristics) were disproportionately labeled "at risk," as well as how such labeling serves to maintain the status quo, Richardson (1990) implied that those students were the ones in whom ethnic and linguistic difference and diversity resided—as a "personal characteristic" (p. 65). A classroom of multiple students whose ancestries differed from one *another* or who spoke languages that differed *from one another* could be characterized as diverse and might constitute a state of ethnic and linguistic "diversity." In such a classroom, students with ancestries outside Europe or who spoke languages other than English would only be different *in relation to* students whose ancestries were within Europe and who spoke English.

To characterize students with ancestries outside Europe or who spoke languages other than English as inherently "different" or "diverse" without specifying *from whom they differ or diverge* locates "difference" and "diversity" *in* them rather than *between or among them and the groups from which they differ* ethnically and linguistically. The groups from which "at risk" students differ remained unnamed and unmarked, as was the stratified system by which the former are sub-ordinated and the latter are super-ordinated. As such, Richardson centered students with European ancestry and who spoke English as the standard of comparison. Perhaps unwittingly, Richardson thus reinforced the normativity of whiteness.

Amplifying relations of exploitation. The invocation of diversity and inclusion serves to assert the value of identity and difference as lying in their potential to benefit

not those who have been and continue to be excluded—not those who are racially otherized—but rather in their potential to benefit decisions, projects, or enterprises that continue to be racially stratified. Indeed, these decisions, projects, or enterprises remain disproportionately led or owned by those racialized as White. As such, the invocation illustrates the white supremacist logic of "blackness as property" (Smith, 2016; described in Chapter Two) as well as critical race theory's precept of interest convergence (Bell, 1980; Guinier, 2004; Stec, 2007). According to critical race theory, Whites will make concessions to demands by Blacks only to the extent that the concessions ultimately advance their own interests. Crusto (2005) described this very particular valuing of "diverse" bodies—specifically, those classified as Black—for the benefit of those classified as Whites in his analysis of the U.S. Supreme Court's decision on the *Grutter v. Bollinger* case, which was brought by an applicant classified as a White woman against the University of Michigan Law School. While the decision upheld Affirmative Action, it did not do so because the presence of students considered "diverse" at the University of Michigan would benefit them or their communities through their practice of law. The court upheld Affirmative Action because, it stated explicitly, the presence of students considered "diverse" would benefit students classified as White, who would presumably need to engage in an increasingly desegregated society in which the experience and understanding of heterogeneity would serve their interests. Ahmed (2007) describes how the term "diverse" allows for this application in ways that not all terms signifying difference do.

> [W]hat is meant by 'diversity' is kept undefined for strategic reasons. What is interesting to note here is how quite contradictory logics are used simultaneously: in other words, the business model and the social justice model are used together, or there is a 'switching' between them, which depends on a judgement [*sic*] about which works when, and for whom. (Ahmed, p. 242)

Amplifying relations of powerlessness. Within evaluation, Covert's concerns represented the fair employment roots of diversity as representativeness, and Conner's column and presidential address, both reported in Chapter Four, represented its movement through the intrinsic-value-of-heterogeneity and celebration-of-differences

phase. The field's prevailing discourses of diversity, inclusion, and culture between 2002 and 2016, reported on in Chapter Five, represented the business case for diversity. In the U.S. field of program evaluation, the business case lay in validity—the need to gain entrée into and solicit accurate information from social groups that were observed as over-represented among program participants and under-represented among program evaluators (Lee & Gilbert, 2014). Lee and Gilbert tied AEA's business case to evaluators' role with respect to institutional interests of the industries in and around evaluation when they described how Ricardo Millett, evaluation manager for the Kellogg Foundation at the time, had been receiving feedback from private funders that they were struggling to find evaluators of color.

> These funders, who were paying increasing attention to the racial, ethnic, and cultural diversity of communities and the impact of this diversity on neighborhoods, institutions, and systems, assumed that evaluators who shared the racial, ethnic, and cultural characteristics of program participants would yield more culturally relevant and valid evaluations.… (Lee & Gilbert, 2014, p. 98, citing the Association for the Study and Development of Community [ASDC, currently known as Community Science], 2002)

In other words, it was philanthropy's interests in validity and perhaps justice for program *participants* from racially otherized groups—as opposed to justice for program *evaluators* from racially otherized groups—that led them to seek evaluators from racially otherized groups by funding BDI, and that ultimately impinged upon AEA.

Structural rather than individual change. Those most closely involved in AEA's decisions and processes at the time of the BDI, as reported in interview findings in Chapter Seven, described the discourse shift away from terms more specific to racialized difference toward variations of "diversity" and "inclusion" as strategically tied to more macro-level, structural dynamics in multiple ways. First, interviewees described it as corresponding with a national policy shift away from redistributive programs and toward privatized efforts at diversity and inclusion. This policy shift, associated with racialized neoliberalization, often manifested materially in anti-Affirmative Action suits filed by students classified as White women. Second, and related, interviewees described it as an attempt to attract support from those harmed by systems of oppression other than white

supremacy or interested in dimensions of difference other than racialized difference. These were related to macro-level, structural discourses through the misinterpretation and misuse of intersectionality as a wedge in the dismantling of redistributive programs that were achieved through constituent-led Civil Rights movement organizing efforts. "[D]iversity enables action because it does not get associated with the histories of struggle evoked by more 'marked' terms such as equality and justice" (Ahmed, 2007, p. 238).

Amplifying relations of marginalization. What is notable not only from the perspective of an enabling conceptualization of justice abstractly, but also materially from the perspective of the Affirmative Action roots underlying establishment of AEA's MIE TIG is that both previous literature and interview findings reported in Chapter Seven suggest that BDI was not initially developed in close conjunction with the MIE TIG, despite claiming to share its goals (Symonette, Mertens, & Hopson, 2014). The decision to take these goals up to the board level of the association—which was white-led—and broaden the focus beyond race effectively removed them from the hands of the MIE TIG—which was led by members of racially otherized groups, particularly African American. In addition to fundamentally shifting the language and corresponding purpose, this elevation and broadening potentially weakened the MIE TIG's *raison d'être*. As reported in Chapter Seven, results of the interviews conducted with those most closely involved in these decisions suggest that they might have made them with an intuitive, if not calculated, understanding of interest convergence.

The MIE TIG seems to have responded by attempting, unsuccessfully, to fortify itself by communicating its relevance to racially otherized groups beyond those classified as Black or African American. Soon thereafter, the IPE TIG was established, and a decade later, the Latinx Responsive Evaluation Discourse and GEDI TIGs were established. As a result, it is possible that today, an AEA "committee of others" (as Hall is reported in Chapter Seven as having described it) would include members from multiple TIGs representing racially otherized groups, as opposed to just one (the MIE TIG). Together, these members representing multiple racialized groups may constitute a

greater proportion of such a committee. As such, they could potentially achieve the critical mass necessary to identify and address the ways that white supremacy manifests in evaluation. Especially in conjunction with members of TIGs representing groups otherized along other dimensions of identity, these members could potentially identify and address the ways that white supremacy intersects with other systems of oppression with enough analytical specificity to drive changes in the structural arrangements that mediate decision-making processes and socio-economic relations among groups in the industries surrounding evaluation.

Social group rather than individual interests. Results reported in Chapter Six from the archival analysis and interviews suggest that evaluation scholars and practitioners who identified as women representing racially otherized groups were at the forefront of efforts to recruit and mentor evaluators from racially otherized groups and deepen understanding of systemic oppression and racialized poverty among all evaluators. However, interviewees tied the shift toward discourses of diversity and culture as a stated attempt to reflect multiple dimensions of identity to the tenure of scholars and practitioners classified as White women inhabiting positions of power within AEA. Regardless of their racial identification, interviewees described the positions of power that these White women inhabited within the Association as both formal, gained through their elected leadership roles, and informal, gained through their kinship relationships with White men in elected leadership roles.

Difference as produced rather than natural. Interviewees who identified as African American women saw the shift toward discourses of diversity and inclusion as an attempt by those classified as White women to advance their own social group interests. Noteworthy in this regard is the finding, reported in Chapter Four, that the percentage of U.S. peer-reviewed evaluation literature containing variations of the liberal terms "diversity" and "inclusion" far exceeded the percentage of literature containing variations of the term "intersectional" (1%). This was despite the latter having emerged from scholarship in Critical Race Theory and Black Feminist Thought more than a decade

prior—at approximately the same time that "diversity" entered higher education and business discourse.

Intersectionality. Hall's comments regarding Californians who identify as White women voting against Affirmative Action, which were reported in Chapter Seven, illustrate how liberal interpretations of multiple identities—wherein dimensions of difference are perceived, first, as intrinsic and, second, as competing with rather than working through each other—serve racialized neoliberalism. California voters classified as White women simultaneously experienced the intersecting oppressions of cis-hetero-patriarchy, under which their status as women is sub-ordinated, and White supremacy, under which their status as White is super-ordinated. The oppressiveness of cis-hetero-patriarchy as a system lies precisely in the status of those classified as women—regardless of racial classification—being one of dependence; the status of those classified as women rises or further falls depending, overwhelmingly, on their economic relationship with those classified as men. Maximizing their economic interests involves attachment, typically through the cis-hetero-patriarchal institutions of marriage and other kinship structures, to those whose social groups status are super-ordinated by one or more systems of oppression—in other words, those who are classified as White, men, able-bodied, or owning-class.

A critical interpretation reveals how voters classified as White women share many more socio-economic interests with those classified as White men than they do with women representing racially otherized groups, from whom they are residentially, socially, economically, and occupationally segregated. This is because a critical interpretation constructs difference as borne from social group interests that interlock with and inflect each other. It further recognizes identity as socially constituted through institutionally and structurally mediated relationships—including marriage and kinship—as opposed to individual characteristics. Thus, Californians classified as White women did not vote for their husbands' and sons' interests as men at the expense of their own interests as women; on the contrary, they voted *for* their self-interest as women whose status they knew remains sub-ordinated under cis-hetero-patriarchy. In other words, they

did so *because* they identified as women, not *in spite of* identifying as women. The calculus underlying their voting pattern corresponds with the finding that marriage fails to yield significant wealth gains for Black women (Zaw, et al., 2017).

The material focus on social group interests, which is fundamental to the inherently structural metaphor of intersectionality as originally conceptualized in critical legal studies (Crenshaw, 1989), offers otherized groups possibilities for liberation movement building in ways that "diversity" does not. In the California example, the power of intersectionality would lie in all voters with an interest in unfettering economic security from attachment to those classified as White men building solidarity around that shared interest. Voters from groups whose land, labor, and other capital are exploited in or excluded from the formal economy; whose decision making is infantilized or disciplined; and whose social group norms are otherized would organize liberation movements. Through collective action, they would change the structural and institutional arrangements that mediate those oppressive relations and subsequently maintain the contingent determination of their status.

The replacement of race with culture. Like the frequencies for variations of "diversity," Chapter Five reported that the annual frequencies for variations of "culture" in the U.S. peer-reviewed literature underwent some shifts that coincided with meso-level, institutional decision-making processes and socio-economic relations within and surrounding the field's professional association. While overall frequencies were comparable for variations of "culture" and "race," annual frequencies for variations of "culture" surpassed those for variations of "race" both in the literature in AEA-affiliated publications and in the U.S. evaluation literature at large in the 2000s after having been lower through most of the 1980s. The greatest increases in annual frequencies for variations of "culture" coincided with Karen Kirkhart's 1994 keynote address on multicultural validity in the annual conference and its 1995 publication in the AEA-affiliated *American Journal of Evaluation*; completion of AEA's Cultural Reading of the Program Evaluation Standards in 2004; and approval of AEA's Statement on Cultural Competence in 2011. The latter two were both part of the implementation of BDI.

This finding regarding the increase in annual frequencies for variations of "culture" relative to the annual frequencies for variations of "race" suggests that discursive events and mechanisms within the professional association may influence language use and subsequent understanding not just among AEA members, but perhaps more broadly within the evaluation field. This is particularly so considering that items containing variations of "culture" were coded for relevance and substantiveness. In other words, the analysis directly compared items signifying racialized difference using variations of "culture" with those signifying racialized difference using variations of "race." AEA's formal emphasis on culture as opposed to race was followed by a decline in direct references to race not only within AEA-affiliated publications, but also in those representing the evaluation field as a whole.

Difference as asymmetrical rather than lateral. Findings regarding the increase in annual frequencies for variations of "diversity" and "culture" relative to terms more specific to racialized difference reported in Chapter Five also suggest a relationship among discourse; meso-level, institutional decision-making processes within AEA; and macro-level, structurally mediated socio-economic relations among social groups. Evaluation scholars and practitioners closest to the field's language and meso-level, institutional decision-making processes generally described the shifts in language as conscious attempts to subdue the focus on race so as not to alienate AEA's board and membership. This strategy was tied by interview results reported in Chapter Seven to the professional association's shift away from anything that might be perceived as political. The Association was avoiding the perception of political activism after having made formal statements, with federal implications, on high-stakes testing and on the privileging of experimental design, both of which elicited the departure of longstanding and high-profile members.

Amplifying relations of cultural imperialism. As stated in discussion of the finding regarding diversity and inclusion, interview results reported in Chapter Seven suggest that those most closely involved in evaluation's construction of racialized difference may have made their case for increased attention to diversity, inclusion, and culture with full

awareness of the need to advance White interests. It is unclear, however, whether this tactic has served the interest of racially otherized groups in cultivating a critical understanding of racialized difference and white supremacy in relation to other dimensions of difference and systems of oppression. For example, Hood stated:

> In the case of *some* racial and cultural groups, it is not always what they say, but how they say it, as well as the observed nonverbal behaviors. *An evaluator of color* could play an important role as an "interpreter" in the design stages, during the dialogue process, during the implementation of the evaluation, and in the interpretation of evaluative findings. (Hood, 2000, p. 82, emphasis added)

The implication is that in the case of *other* racial and cultural groups—presumably those *not* racially otherized, but rather racially normatized as White—what is said is somehow disconnected from how it is said. Much like Richardson's (1990) use of the words "diverse" and "different," in the absence of historical context or an analysis of power, Hood's specification of only "*some* racial and cultural groups" as relying largely on nonverbal cues serves to essentialize culture within racially otherized individuals and groups and reinforce the normativity of whiteness. It implies that the low-context communication styles typically associated with those racialized as White are standard, against which the high-context communication styles typically associated with racially otherized individuals and groups are "different." Rather than exploring both communication styles as manifestations of culture—specifically as practices developed in response to experiences shaped by racially differentiated access to authoritative power—Hood characterizes the latter alone as "cultural."

Social group rather than individual interests. Similarly, instead of recognizing the skill developed by many racially otherized individuals and groups to "switch codes" actively (even if subconsciously) to survive within white hegemony, Hood's language—particularly his conflation of race with culture—naturalizes the ability of "an evaluator of color" to serve as an interpreter. Finally, it obviates consideration of the nonverbal communication that does take place among those racialized as White, including the codes that are exchanged in exclusively White spaces, which allow participants to perform and associate themselves with whiteness. The acquisition and assertion of relational and

human capital—which is inherently racialized, gendered, classed, and ableized—become salient even to members of super-ordinated groups in more homogenous settings. Indeed, the lack of fluency with such cues and codes among racially otherized individuals and groups—as well as among those perhaps racially normatized as White, but otherized in terms of class, gender, sexuality, or ability status—can hinder the ability to wield influence in such spaces.

Importantly, Hood's piece excerpted above was published during the same years as the BDI study and early implementation. References to racism inherent in U.S. institutions and structures—including U.S. democracy—that Hood made in that piece declined precipitously between 2002 and 2016. However, the attribution of "culture" to racially otherized individuals and groups increased, as did literature expressing interest in increasing the extent to which program evaluators reflect the racially otherized groups disproportionately represented among program participants—although that disproportionality was never historicized nor contextualized, but rather treated as a cultural incongruence between program evaluators and program participants. Again, the interest expressed was largely to increase what Kirkhart (1995) referred to as "multicultural validity" in evaluation results. Since the racial composition of program participants has become the rationale motivating calls for increased diversity and cultural competence among program evaluators, the evaluation literature has increasingly treated it as fixed and depoliticized.

Difference as cumulative rather than static. Hood attributed the shared lived experience and understanding of issues framing a program and its evaluation that he believed existed between program participants and program evaluators practicing during the era of legally sanctioned segregation to shared culture. He further conflated culture with racial classification, rather than to a shared history and experience of oppression and subsequently shared interest in and commitment to justice and emancipation for African Americans. He noted that he focused his study on evaluators racialized as African American and practicing prior to the *Brown v. Board of Education* decision (1954) and Civil Rights Act (1964). Intersectional nuances within the community based on class,

color, and other dimensions notwithstanding, evaluators racialized as African American who were practicing during the era of legally sanctioned racial segregation in education, housing, employment, and social interaction would have likely spent their entire lives in fairly close geographic, social, and generational proximity to other African Americans—including those participating in the programs they evaluated—whose levels of education and other forms of capital may have differed from their own (Collins, 2002). Such conditions, wherein racial classification and geographic space map onto each other so perfectly (by design), can give rise to essentialist notions of culture that are uncomplicated by different experiences rooted in socio-economic class, generation, sexual orientation, and gender identity.

Social group rather than individual interests. Between-group segregation based on racial classification flattens within-group differences to those outside the group; the cultures that develop in such settings routinely become synonymous with race. In extrapolating that this notion of shared culture could persist in uniting program participants from racially otherized groups with program evaluators from racially otherized groups beyond the era of legally enforced residential, educational, economic, and social segregation, Hood lent credence to "[t]he fiction of cultures as discrete, object-like phenomena" (Gupta & Ferguson, 1992, p. 7). Consider in contrast the current concentrations of wealth and poverty in the United States. Today's program participants who are racialized as African American may live in closer proximity to—and share more lived experience with—other *program participants* who are *racialized differently from them* than they do with *program evaluators* who are *racialized similarly*, as African American.

Very similarly, AEA's Statement on Cultural Competence, which was approved a decade after Hood's piece, treats privilege, like culture, as a static possession that one can "hold" (AEA, 2011). In doing so, the Statement evades consideration of the mechanisms through which different social groups within the structures surrounding evaluation continue to depend on and interact with each other, which mediate how—and how much—each can *exercise* its power and privilege. These mechanisms include socio-

economic relations and decision-making processes within the nonprofit/nongovernmental industries and larger society, in both of which evaluation plays a pivotal material and discursive role (House, 2017). Leaving the mechanisms underlying the power dynamics, as well as whom they super-ordinate, unstated normalizes and reinforces the association of some social groups with higher status and others with lower status. Through its failure to contextualize the associations and problematize the underlying mechanisms, the Statement thus reinforces the process of systemic oppression (including structural racialization) rather than counteracting it.[57]

The rise in indigeneity and its decoupled relationship with colonization. As described in Chapter Four, most of the authors who have written critically about racialized difference within the U.S. peer-reviewed evaluation literature have done so only once, and much of that literature was written no later than 2002. Hood alone has consistently written about racialized difference critically within the evaluation literature—typically from the perspective of educational evaluation—before as well as after 2002. Additionally, though, especially after 2003, literature that critically addressed racialized difference from the perspective of indigeneity and colonization—often written in the first person—began to appear and grow.

Counteracting relations of powerlessness. In contrast to that of the MIE TIG, whose work was arguably diluted by its elevation to the board level and the board's broadening of the focus, the work specific to indigenous peoples in evaluation within AEA largely remained at the TIG level and thus driven by those most familiar with, affected by, and active on issues to which indigeneity and ongoing settler colonization were central. The work of this community of scholars and practitioners coincided with great shifts in language within the field—especially in AEA-affiliated journals—wherein use of variations of both "indigenous" and "colonization" increased dramatically and

[57] In the era of neoliberalism, which focuses attention on individuals rather than structures, research suggests that disparity data are as likely to reinforce negative stereotypes as they are to stimulate support for structural interventions (powell, 2013). Amid national narratives of the American dream and land of opportunity, even the well-intentioned often assume that differences in power and privilege are a natural result of individual factors such as educational level and access to social networks rather than the effect of structural forces.

their decoupling by authors who identified themselves as members of indigenous groups represented an intentional act of self-determination. Their decoupling contrasts with the disavowal of colonization that the decoupling by authors who either identified as White or did not identify themselves in terms of racial group membership represents.

The increase in variations of "indigenous" and its decoupling from variations of "colonization" by scholars and practitioners who identify with indigenous groups counteracts relations of powerlessness and amplifies indigenous-led collective action. The contrast between this pattern, including its relationship with the IPE TIG, and that of minoritization and ambivalence toward whiteness, including its relationship with the MIE TIG, bears implications for otherized groups and liberation movements that are discussed in Chapter Nine.

Conclusion of Chapter Eight

Evaluation's construction of racialized difference falls into the four discursive patterns reported in Chapter Seven. Interviewees who identified with racially otherized groups tied all four patterns to dynamics at interpersonal, institutional, and structural levels. Chapter Eight articulated each patterns' correspondence with liberal and critical constructions of difference (i.e., as natural, individual, lateral, and static or as produced, relational, asymmetrical, and cumulative). It also articulated the nature of the relation that each discursive pattern amplifies or counteracts (summarized in Figure 35).

Chapter Seven reported that interviewees described the relations underlying the four patterns as both interpersonal and tied to larger discourses and political action associated with racialized neoliberalism at the macro/societal level. Both levels were structurally mediated by discursive events and mechanisms at the meso/institutional level. These discursive events and mechanisms included AEA's themed journal editions; annual conference themes, sessions, and speeches; establishment of topical interest groups within the membership; and board-level committees, task forces, policies, and initiatives. These meso/institutional level discursive events and mechanisms varied in the extent to which interviewees who identified with racially otherized groups said they represented and served their social group interests, as Chapter Eight discussed.

AEA's role in mediating relations integral to the four discursive patterns identified suggests that AEA could play a role in mediating relations that discursively expand and deepen evaluators' understanding of racialized difference and systemic oppression. The variation among the extent to which interviewees who identified with racialized groups felt these institutional mechanisms served their social group interests bears implications for otherized groups and liberation movements more generally. Chapter Nine discusses these implications. It explores institutional mechanisms in terms of their potential to enable collective action among otherized groups working to counteract asymmetrical—or amplify symmetrical—social group relations. To the extent relevant this includes the nature of the asymmetry, types of capital exchanged, and media of exchange (currency).

	Discursive pattern	Relevant Structural Dynamics	Relevant Meso-level/ Institutional Mechanisms	Nature of social group relations and relevant theoretical and analytical concepts
I.	Minoritization and ambivalence toward whiteness	• Higher education	• Professional association topical interest group • Themed edition of professional association-affiliated journal	1. Reflects/reinforces violence
II.	The invocation of diversity and inclusion	• Philanthropy	• Professional association research study • Professional association program implementation	2. Reflects/reinforces cultural imperialism 3. Reflects/reinforces exploitation 4. Reflects/reinforces marginalization • Individuals and social groups • Intersectionality
III.	The replacement of race with culture	• Political discourse • Government policies	• Professional association conference theme • Professional association conference address • Professional standards • Evaluation of professional association program • Professional association statement	5. Reflects/reinforces powerlessness • Individuals and structures
IV.	The rise of and decoupled relationship between indigeneity and colonization		• Professional association topical interest group • Themed edition of professional association-affiliated journal	5. Counteracts powerlessness

Figure 35. Summary of identified discursive patterns by relevant structural dynamics, institutional mechanisms, and face of oppression

CHAPTER 9: IMPLICATIONS

As discussed in Chapters One and Two, the field of evaluation has noted an incongruence between program participants and program evaluators since its inception. Throughout the field's history, concerns regarding the incongruence have been framed, to varying degrees, in terms of validity and justice. In addition to introducing critical constructions of difference and enabling conceptualizations of justice, both of which are rooted in critical theories of systemic oppression, Chapter One problematized the absence of a body of evaluation literature that considers difference and justice in relation to evaluation's role within the racially stratified systems that surround it and utilize its services and results. The field's framing of difference as cultural incongruence among individuals has left the asymmetrical dynamics within and around it largely unexamined and unchallenged, naturalizing racial stratification and reinforcing the normativity of whiteness. This naturalization and normativity have implications for the current context of racialized neoliberalism that are even more urgent in the present moment of racial panic described in the introduction (Desmond-Harris, 2016; Gitlin, 2018).

Chapter Two outlined the racial stratification of U.S. society and the nonprofit industry and described the tension between individuals and structures as central to the nonprofit industry from its inception through the present day. It contextualized the industry's racially stratified structure in relation to its historical origins and ongoing reluctance to fund structural change by enabling the exercise of collective action toward that end by racially otherized groups. Additionally, Chapter Two reviewed U.S. evaluation literature that addressed racialized difference. The bodies included utilized a context-sensitive or multi-level analysis; questioned prevailing, liberal constructions of difference by recognizing racialized difference as produced; and recognized social group interests in relation to racialized difference. The latter fell broadly into literature that reinforced the normativity of whiteness and literature that emphasized self-determination

among racially otherized groups, especially from indigenous and decolonizing frameworks. In the absence of critical theories of systemic oppression and system dynamics, even evaluation literature utilizing contextual or multi-level frameworks in discussions related to difference failed to consider structurally-mediated relationships in the historical production and ongoing reproduction of difference.

Chapter Three detailed the methodology that this dissertation employed to link critical theories of systemic oppression with systems thinking to "… move beyond the individual and consider the interconnected parts of the ecosystem (i.e., organizational policies and practices, institutional norms) that made…the programs [necessary] in the first place…." (Lee & Gilbert, 2014, pp. 106-107). This link was made with the explicit intention to contribute to a body of research that builds empirical evidence that can inform remedies for the structural asymmetry underlying racialized dynamics within and among institutions. Chapter Three thus named the dissertation's philosophical and practical positions as borne from an intersectional paradigm rooted in critical theories of systemic oppression and systems thinking.

Chapter Eight discussed the results reported in detail in Chapters Four, Five, Six, and Seven—specifically, the four discursive patterns—in terms of the constructions of difference that each advances (critical or liberal). Using Young's (2011) Five Faces of Oppression, it also articulated the nature of the relations that each discursive pattern amplifies and counteracts. Specifically, the nature of these relations is as follows: (1) minoritization and ambivalence toward whiteness reflect and reinforce relations of violence arising from the racial panic associated with their extreme manifestation, which is white nationalism; (2) the invocation of diversity and inclusion as well as (3) the replacement of race with culture reflect and reinforce relations of exploitation, cultural imperialism, powerlessness, and marginalization; and (4) the rise of and decoupled relationship between indigeneity and colonization increasingly counteract powerlessness.

Those closest to evaluation's construction of racialized difference and related discursive events who were interviewed were reported in Chapter Seven as having tied each pattern to interpersonal, institutional, and structural dynamics that were mediated by

meso-level mechanisms in and around AEA. In other words, the interpersonal dynamics simultaneously initiated and supported—and were initiated and supported by—institutional and structural dynamics. As such, the implications of the four discursive patterns for research and practice exist at every level of Reskin's (2003) Mechanisms-based Model. Chapter Nine rests on the conclusion drawn from this dissertation's research results that AEA could more intentionally approach the role that it has already been playing in mediating the interpersonal, institutional, and structural dynamics that discursively construct racialized difference. Furthermore, AEA could approach that role with the explicit intention to challenge racialized neoliberalization and to amplify collective knowledge production and action among racially otherized groups. Chapter Nine uses concepts from critical theories of systemic oppression and system dynamics to consider the implications of the four identified patterns' development for evaluation practice and for research on evaluation.

Organization of Chapter Nine

Chapter Nine considers the ways that mechanisms, events, and actors at each level of the ecosystem within and around evaluation can restructure the flow of capital to counteract liberal constructions of difference and racialized neoliberalization. It also considers the ways that they can amplify otherized groups' ability to engage in collective knowledge production and action to effect structural change. It is organized by level: internalized, interpersonal, meso/institutional, and macro/structural. Exploration of the implications for the meso- and macro-levels concentrates on specific mechanisms and events or actors and entities described by interviewees as having led to and arisen from interpersonal, institutional, and structural relations that they described as having influenced—and having been influenced by—the field's construction of racialized difference.

Specific actors and mechanisms that interviewees mentioned within AEA include:

- Board of directors, committees, TIGs, and membership
- Decisions, policies, or statements
- Professional development opportunities, especially the annual conference

- Programs such as those implemented as part of the BDI
- AEA-affiliated journals

Specific meso- and macro-level actors and entities around the professional association that interviewees mentioned include:

- Educational institutions
- Philanthropic foundations
- Government agencies
- Large evaluation firms
- Peer organizations (such as the Joint Committee)
- Individual representatives thereof

Within each level, implications are discussed in terms of their value for research on evaluation as well as for evaluation practice.

Implications for internalized racialization

At the internalized level, evaluation scholars and practitioners—regardless of phenotype, identity, or upbringing—can reflect critically on themselves and their work as inevitably representing the interests of the social groups with whom they identify and with whom others identify them. Such reflection and identification challenge the notion of individualism that is foundational to liberal constructions of difference and racialized neoliberalization. They can further historicize and contextualize the types of capital to which they have access and in which they trade, both personally and professionally. They can counteract racialized neoliberalization and amplify collective knowledge production and action by engaging in the critical reflexivity necessary to recognize, unlearn, and sit with the anxiety that they may experience when they intentionally share or relinquish the power to control and predict processes, outcomes, and narratives.

Critical reflexivity and critical praxis. As reported in the interview results in Chapter Seven, Ricardo Millett raised several questions in response to preliminary results that suggested four discursive shifts had taken place in the signification of racialized difference in the evaluation literature. Paraphrased, he asked:

- Does the application of the language patterns identified through the textual and diachronic analyses lead to differences in the way that evaluation questions are framed? Or how data are analyzed?
- Does it affect the validity and utility of the [evaluation] product in addressing systemic issues?
- Has the fuller recruitment, inclusion, participation, and research activity of evaluators representing racially otherized groups over time influenced the way these terms identified through the textual and diachronic analyses are used?
- Does it get more authentic kinds of analyses and conclusions?

Evaluation scholars and practitioners can prepare themselves emotionally, cognitively, socially, and politically to collect demographic and other data without automatically reproducing liberal narratives—and corresponding categories—of difference. Rather, they can work in solidarity with otherized groups to collect data with enough analytical specificity and relevance that otherized groups can use the process and results to effect changes in structural mechanisms that mediate socio-economic relations. Collectively integrating theory/cognition, reflection, and action is often referred to as critical praxis.

For example, programs frequently collect and report disaggregated data on "Whites," "Asians," and "Blacks" as racial categories. The results of such efforts cannot inform action in the way that collecting data on nativity and generation and circumstances of migration can. The outcomes of refugees from Europe, Asia, and Africa may be more similar with each other than they are with those of immigrants or U.S.-born members of their respective racial classifications. These similarities may arise from the trauma of war, precarious documentation status, or language differences—all of which, unlike "race," can potentially be addressed through a combination of service provision, advocacy, and constituent-led organizing. Likewise, the outcomes of foreign-born individuals from the same continent may differ depending on whether their countries of origin had been colonized by European power and, if so, which ones. For example, refugees from Sri Lanka, who were likely educated in English, may be at a relative advantage in U.S. workplaces compared to immigrants from Korea, who likely were not.

In the same way, it may be that those classified as "Asian" who arrived as refugees from the American war in Southeast Asia have more outcomes in common with those classified as "African American"—arising from a shared experience of disinvested public housing—than they do with those also classified as "Asian" who arrived as math and science professionals prioritized in the 1965 Immigration and Nationality Act. It may be that Americans of South and West Asian ancestry—regardless of nativity, generation, or circumstances of migration—who are classified as "Arab" or "Muslim" have outcomes more in common with those classified as "Black" than they do with Americans of East and Southeast Asian ancestry, who are classified as "Asian." These may arise from the shared experience of racial profiling, wherein the former are perceived as Islamic terrorists and the latter are perceived as criminals. The flattening of differences within racial classifications heightens the differences between racial classifications, reifying "race" as a static attribute affixed to individuals, which corresponds with liberal constructions of difference and ultimately serves racialized neoliberalization.

At the same time, it may be that within a decade, immigrants from Africa—regardless of educational level, income level, and insurance status—face the same health outcomes that those classified as "Black" who survived the Middle Passage, centuries of enslaved labor, and generations of segregated housing and education do. It may similarly be that recent immigrants and refugees from Africa experience the same rates of police violence that those classified as "Black"—regardless of educational and income level—have been experiencing since Emancipation. In these cases—when racial classification *does* bind individuals that differ along other presumably more influential dimensions such as nativity, generation and circumstances of migration, educational level, income level, or insurance status—it becomes necessary to historicize and contextualize the systems involved in relation to the (re)production of racialized difference. Without longstanding social, economic, and political relationships with those already addressing these issues, evaluators—of any phenotype and social group identification—will be ill-equipped cognitively to achieve the racial literacy and levels of analytical specificity and relevance that are necessary for political action and that are feasible with the populations

and situational constraints involved. Evaluators who have been trained to control or predict the process or product must therefore prepare themselves emotionally to approach this work interdependently and intersubjectively with those already engaged in knowledge production and collective action on the relevant issues.

Collective rather than individual action. Because those racially normatized as white are an artificially-produced majority that is super-ordinated by white supremacy, and those racially otherized are sub-ordinated by white supremacy, for relations to be symmetrical and just rather than asymmetrical and oppressive, the latter must typically be organized to contribute and benefit as a collective body—such as a TIG—to effect structural changes that represent their interests. **Error! Reference source not found.** illustrates the feedback loop through which EvalTalk, a forum created for the exchange of ideas and opportunities among evaluators disproportionately racially normatized as White, preserves homeostasis: Individuals attempt to post content representing the interests of racially otherized and sub-ordinated groups to EvalTalk. The White majority—which was artificially-produced through relations of marginalization—that created the list-serv to advance its socio-economic interests had near-exclusive access to it until relatively recently. It had established group norms and policies stating that content posted to it must relate to evaluation. Such norms and policies themselves exemplify cultural imperialism, in that they reflect and reinforce the low-context, disciplinary independence that is characteristic of colonial knowledge systems and antithetical to many indigenous knowledge systems. Unless the content representing the interests of racially otherized groups also serves its interests (illustrating both exploitative relations as well as the CRT concept of interest convergence), the White majority perceives such posts as group-specific and irrelevant to evaluation—which it perceives as both universal as well as socially and politically neutral. This perception is rooted in liberal constructions of racialized difference, wherein the White majority is considered natural, ahistorical, and pre-ordinate. Moreover, to enforce the rules that it created to serve its interests, the White majority relies on infantilizing and disciplining tactics (illustrating relations of powerlessness and violence) that are characteristic of oppressive systems.

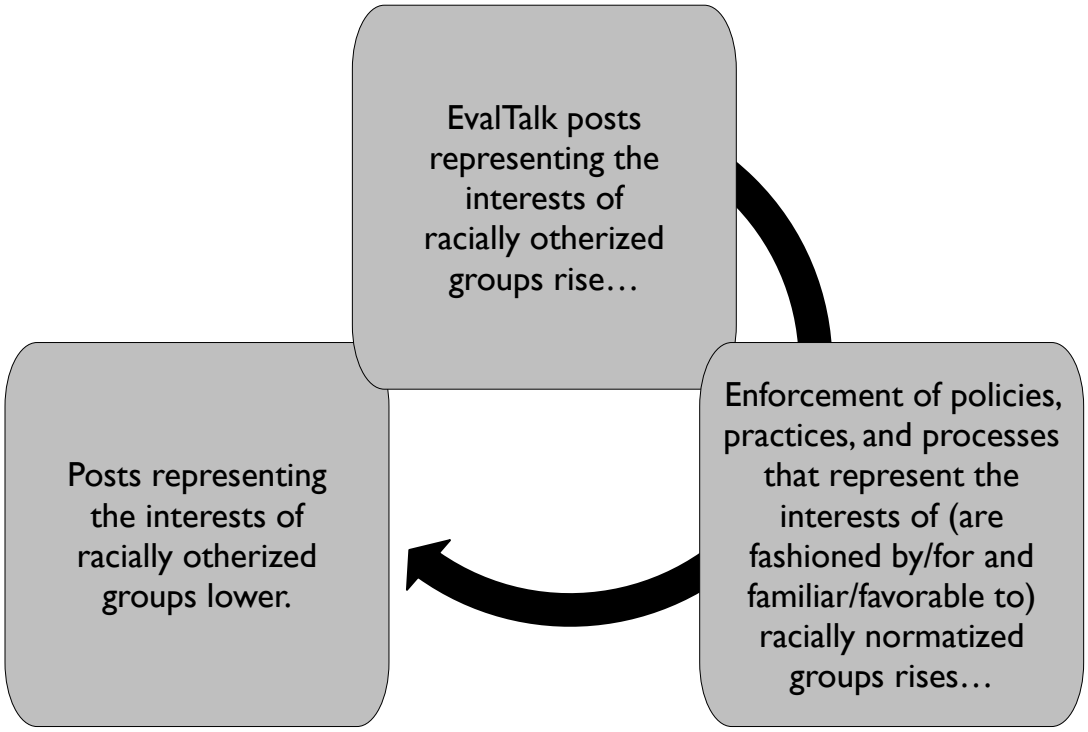

Figure 36. Negative feedback loop counteracting efforts by individuals representing otherized groups against marginalization and cultural imperialism

Even among those racially normatized as white, however, the individual efforts listed above—while necessary—are insufficient to effect structural change. To maximize the impact of individual evaluation scholars' and practitioners' efforts in the above areas, actors within and around evaluation would need to create conditions that enable evaluators representing the interests of otherized groups to provide feedback that challenges the discursive reproduction of liberal constructions of difference and the concentration of capital among evaluators who benefit from such constructions. The next three sections discuss some of the necessary interpersonal, institutional, and structural conditions for evaluators representing the interests of otherized and sub-ordinated groups to participate in the exchange of ideas and opportunities.

Implications for interpersonal racialization

At the interpersonal level, evaluation scholars and practitioners of all phenotypes, social group identifications, and life experiences can acknowledge themselves as already related—engaged in the exchange of capital—to others through the racially stratified

socio-economic structures surrounding their evaluation work, at a minimum. They can acknowledge many of these relations as characterized by super-ordination and sub-ordination and reflect critically on repeated association, rooted in racial stratification, as the source of their implicit biases. They can cultivate critical literacy in themselves and in their clients, partners, peers, and students by engaging with and citing scholarship and activism led by otherized groups.

Evaluation scholars and practitioners can cultivate relations through which they engage in mutual learning—as opposed to capacity building, which is inherently asymmetrical—intentionally with members of their own groups and intentionally with members of other groups, whether "group" is defined by identity, role, occupation, discipline, etc. The exchange of relational, learned, and created capital, among others, could be driven by collective action on specific issue areas such as housing education, health as opposed to being driven by philanthropic or government interests. What can evaluation scholars and practitioners learn from activists and organizers in these areas? How would it affect their conceptualization of social problems, how change happens, and what data are relevant?

Opportunities for mutual learning can allow evaluation scholars and practitioners to cultivate and practice relations of interdependence and intersubjectivity with members of otherized groups as co-producers of knowledge and agents of change. In relations of interdependence and intersubjectivity, evaluation scholars and practitioners would not necessarily control or predict the outcome of situations as they may have grown accustomed to, depending on their role, level of authority, and social group identification. Evaluation institutions, including professional associations like AEA, can create conditions that enable interpersonal relations rooted in interdependence and intersubjectivity—as opposed to sub-ordination and super-ordination—to grow.

Implications for institutional racialization

As with individual evaluation scholars and practitioners, evaluation firms, internal evaluators, evaluation training programs, and professional associations like AEA can acknowledge themselves as already related—engaged in the exchange of capital—to

others at the institutional level through the racially stratified socio-economic structures surrounding and integral to evaluation. These relations, characterized by super-ordination and sub-ordination, include the exchange of capital not only with clients/contracting agencies, students, members; funders; partners; subcontractors, organizational leadership, program staff, and program participants, but also with their own management and staff—at multiple levels—as well as with vendors and neighboring businesses and residents.

Institutional mediation of interpersonal relations. Evaluation institutions can systematically ask which internal and external structural mechanisms trigger different kinds of feedback loops, as opposed to one-way flows—such as charity and enslavement—that are unsustainable. A matrix such as Figure 37 can help institutions identify feedback loops that amplify the interdependent exchange of capital exclusively within groups who are racially normatized as White. Which ones counteract the circulation of capital either between groups who are racially normatized as White and groups who are racially otherized or within groups who are racially otherized? These mechanisms reinforce asymmetrical, unjust relations between social groups. In contrast, which structural mechanisms trigger feedback loops that counteract the circulation of capital exclusively within groups who are racially normatized as White? Which ones amplify the interdependent exchange of capital either between groups who are racially normatized as White and groups who are racially otherized or within groups who are racially otherized? These mechanisms cultivate more symmetrical, just relations between social groups.

	List structural mechanisms that:	
	1. amplify the circulation of capital…	**2. counteract the circulation of capital…**
a. …within groups who are racially normatized as White?	Mechanisms listed here reinforce oppression	Mechanisms listed here disrupt oppression
b. …between groups who are racially normatized as White and groups who are racially otherized?	Mechanisms listed here build justice and disrupt oppression	Mechanisms listed here reinforce oppression and weaken justice
c. …within groups who are racially otherized?	Mechanisms listed here build justice	Mechanisms listed here reinforce oppression

Figure 37. Matrix to identify the feedback loops that structural mechanisms trigger

For example, evaluation institutions can organize issue-oriented communities of practice focused on mutual learning that embrace norms for discussion and protocols for decision making rooted in otherized and sub-ordinated worldviews, values, and ways of knowing instead of relying on routine processes and codes of "professionalism" rooted in the comfort of normatized and super-ordinated groups. As structural mechanisms, norms and protocols rooted in otherized groups can counteract relations of cultural imperialism if they foster the interdependent and intersubjective exchange of cultural and spiritual capital between otherized and normatized groups. Norms and protocols rooted in otherized groups can also amplify exploitative relations, however, if they foster exchanges wherein members of normatized and super-ordinated groups determine what, when, where, how much, and by whom they are used.

An example of the distinction between a norm or protocol counteracting relations of cultural imperialism as opposed to amplifying relations of exploitation lies in use of the circle, which is a structure and process for discussion and decision making that is common but not universal among indigenous societies. Use of the circle can disrupt structural asymmetries, hierarchies, and stratification if the decisions made through it are binding. For such decisions to be responsible, however, everyone around the circle would need access to the same information and everyone would need to bear some level of accountability for the outcome. When super-ordinated groups use the circle to solicit input and ideas from sub-ordinated groups who are outside—marginalized from—the central decision-making body, creating the illusion of participation in decision-making (often referred to as "buy-in") without any commitment to provide circle-members with the decision-making parameters transparently in advance or to adhere to their decision, however, it becomes a culturally-encoded fetish. Its value is symbolic and its use is disingenuous. In such situations, those using the circle reinforce their own position of power—which reinforces exploitative relations as opposed to challenging the cultural imperialism of business as usual.

Similarly, evaluation scholars and practitioners can resist acting on the urge to serve individually as the sole point of access to and entrée for otherized groups—even

under the noble auspices of diversity, cultural competence, and inclusion—which reinforces their own position of relative super-ordination. This characterizes the pattern described by interviewees, reported on in Chapter Seven, wherein AEA leadership recruited evaluators associated with racially otherized groups to serve as one of many committee members—but marginalized the MIE TIG as the body that had been organized to represent the interests of racially otherized evaluators—in shaping the professional association's approach to the BDI. Instead of individually recruiting otherized individuals, evaluation scholars and practitioners can work in solidarity with otherized groups toward institutional and structural changes that amplify rather than counteract their ability to exchange relational, human, and learned capital. This dynamic is discussed in the implications for professional development, especially the section on TIGs.

Evaluation institutions can increase both the demand for and supply of critical scholarship and practice by supporting education and publication in a variety of media. Such work would create opportunities for otherized groups to exchange types of capital among themselves—strengthening that feedback loop—such that they represent their own social group interests, as the IPE TIG—unlike the MIE TIG—was able to. The IPE TIG's relations with journals and higher education enabled it to do so (see Figure 38).

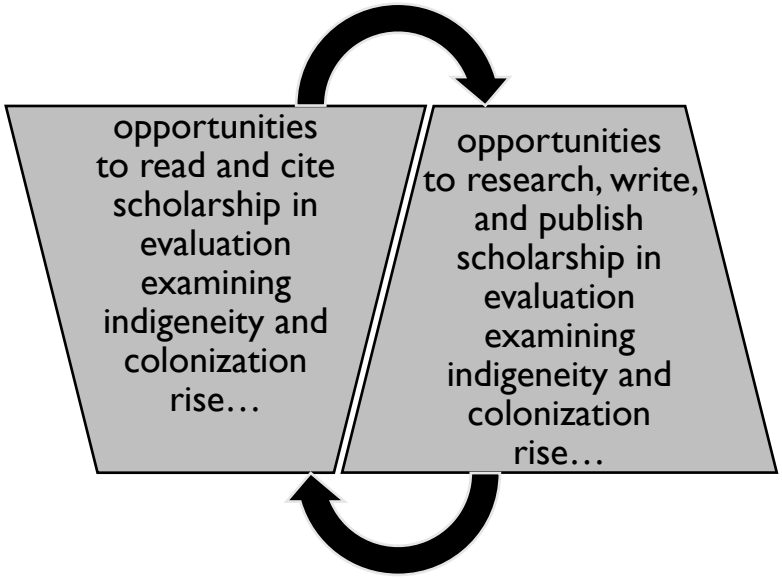

Figure 38. Potential positive feedback loop amplifying self-determination, maximized by coordination among philanthropy, AEA, TIGs, journals, and higher education

Conferences and other professional development opportunities. As reported in Chapter Seven, interviewees described conferences and other professional development opportunities as having influenced evaluation's construction of racialized difference by feeding keynote speeches and sessions, committees and other decision-making bodies, and subsequently the literature. Several interviewees referred to the annual conferences of AEA and the Canadian Evaluation Society in noting how pivotal keynote speeches and sessions—beyond raising awareness about diversity and culture among attending practitioners—facilitated interpersonal relationships among leaders. These relationships were described as having influenced evaluation's use of variations of diversity, inclusion, and culture largely by influencing key processes and the composition of key decision-making bodies tied directly to both the BDI study and its implementation. While committee membership in AEA is increasingly determined by a formal application process rather than appointment—potentially increasing access among otherized groups—more than one interviewee remarked that the current conference size and structure are less conducive to relationship building. They said that today's conference-goers have little informally structured time to engage with each other across experience level and interest area; indeed, they are often segregated by TIG and experience level.

In addition to potentially feeding practice as well as committee composition and subsequent policies and initiatives, however, conferences potentially feed the field's literature. This is especially so for keynote speeches that have historically been printed in association-affiliated journals, but also for ideas that grow out of sessions. Finally, conference participation and sessions simultaneously influence and are influenced by conference theme and TIG membership—each of which also bears a reciprocal relationship with the board's composition. Discussion of relations with TIGs and educational institutions that challenge the helping dynamic associated with the consulting relationship follows.

Individual Topical Interest Groups. Archival analysis of AEA documents and interviews with those most closely involved in the field's construction of racialized difference reported in Chapter Six suggest that efforts to recruit evaluators representing

racially otherized groups, and to train evaluators more generally in working with racially otherized groups, had been underway at the TIG level nearly since AEA's inception in 1986. These organized efforts were led by members of the MIE TIG who largely, but not exclusively identified as African American and in fact as African American women. In the late 1990s, however, when philanthropy expressed an interest in, and provided funding for, recruitment and training efforts that they hoped would help them understand racialized patterns and evaluate the programs that they were funding, the MIE TIG was only one among several TIGs that comprised the Diversity Committee. The remainder, while referred to as "others" by Hall, were represented by those classified as White. The Diversity Committee was charged with thinking through implementation of the resulting BDI study.

Institutional mediation of structural relations. Amid macro-level, structural shifts away from redistributive programs conceived as redress—crystallized in the anti-Affirmative Action suits filed against higher education institutions by applicants classified as White women, however, interviewees perceived intentional efforts to elevate and broaden the focus of recruitment and training efforts beyond race as having resulted in adoption of the discourses of diversity and inclusion. Important from the perspective of enabling conceptualizations of justice is the dissatisfaction with the broadened focus and sense of displacement at its elevation to the association level that some involved with the MIE TIG felt (Collins & Hopson, 2014).

When contraposed, the pattern regarding indigeneity and colonization differs considerably from that of diversity and inclusion as well as that of culture. Interviewees perceived the increase in annual frequencies for variations of both indigeneity and colonization as having resulted at least partially from the establishment of the IPE TIG; additionally, because much of the increase in literature using variations of these terms were contributed by authors who identified as indigenous, the decoupling of indigeneity and colonization was perceived not as a disavowal of colonization, but rather a refusal to define First Nations peoples in terms of colonization. Unlike the work of the MIE TIG, the work of the IPE TIG was neither "broadened" nor "elevated." Indeed, the discourse

shift following establishment of the IPE TIG could be considered an act of self-determination.

Topical Interest Groups collectively. The sequence of events raises questions about the extent to which the MIE TIG's 2004 decision to change its name—in a deliberate attempt to represent members who identify with racially otherized groups beyond African American—was influenced by AEA's "elevation" and "broadening" of the recruitment and training work it had been engaged with and by establishment of the IPE TIG. AEA has started to cultivate informal relationships among members of differently otherized TIGs. A unified body of evaluators representing multiple otherized groups cannot as easily be marginalized. This work can continue through intentionally intersectional conference themes, keynote speeches, and sessions and themed journal editions that broach the analytically specific questions that Millett raised. The risk is that these conference sessions and journal editions will continue to be "ghettoized" rather than fully integrated into the conference and field's canon of literature. This is where educational institutions and philanthropy can play a role, which is described after discussion of ways to circulate capital with and within otherized groups.

Business development and affirmative hiring policy for racially otherized evaluators. Beyond pipeline development and training in cultural competence, both the BDI study completed in 2001 and the Tracking Transformation evaluation completed in 2007 included recommendations for business development and increased access to philanthropic and government contracts among evaluators from racially otherized groups. These included marketing evaluation as a career path to students and professionals, outreach to non-traditional evaluators, small business development training, financial incentives for evaluators of color, mentorship for evaluators of color, a job bank for diverse organizations, and affirmative hiring policies. While implementation of the BDI led to development of the GEDI, which has trained 16 cohorts of nearly 100 emerging evaluators from racially otherized groups, and the Minority Serving Institution Fellowship, which has engaged faculty from institutions that serve students from racially otherized groups, recommendations that could have potentially shifted asymmetrical

socio-economic relations and decision-making processes have largely gone unfulfilled, as reported in Chapter Six.

As reported in the interview results of Chapter Seven, however, Kien Lee saw firm-owners and -leaders (including that of her own) often serving on or having close relationships with individuals who served on AEA's decision-making bodies as well as with higher education, philanthropy, and government agencies. In other words, this group formed a tight, fairly homogenous social group that not only fed the circulation of capital among those classified as White, but also shaped the field's training and professional ethics. Additionally, small businesses classified as disadvantaged (formerly referred to as "minority"-owned, including Lee's own) are widely regarded as providing employment to members of racially otherized groups in ways that White-owned businesses do not. As such, building businesses among racially otherized evaluators and allocation of a proportion of philanthropic and government contract and funding dollars to racially otherized evaluators might have had cumulative and interactive effects on the interests of racially otherized groups. Tracking Transformation recommended that AEA consider approaching such training as a partnership between racial and ethnic identity-based TIGs and the Independent Consulting TIG.

Institutions, particularly AEA as the field's professional association, could approach business development among evaluators representing groups who are otherized racially, and perhaps along additional dimensions of identity, through the community of practice model described earlier in this section on institutional-level implications. They could also do so through a searchable database and directory of evaluators representing otherized groups. Finally, they could advocate and coordinate ways for philanthropy and government agencies to allocate a percentage of all funding or proportion of all research/evaluation contracts to efforts that are led by members of otherized groups and that address structural asymmetries as opposed to individual-level interventions.

Implications for structural racialization

For the above efforts within AEA and related institutions to be viable, however, the institutions must see themselves as part of larger industries and work in concert—in

the same way that they have been working together in concert around an unarticulated, but shared liberal construction of difference. AEA, philanthropy, government agencies, higher education, and large firms can work together intentionally and explicitly around a shared purpose—to counteract the concentration of capital among those classified as White through not just the practice, but also the enterprise of evaluation. AEA's intention and purpose can be articulated in and through the professional association's conference themes and selection of keynote speakers, its professional ethics and related statements, and its special programs such as BDI.

Data access. AEA can show its commitment to critical examination of racial stratification by advancing evaluation as a clear employment category. Publicly available data on the occupation are largely unavailable through Department of Labor and similar bodies because evaluation workers' roles, functions, positions, and titles range from database technician to data scientist to management analyst to research associate. Additionally, AEA's administrative data regarding the demographics of its membership fail to distinguish, for example, between Asian American evaluators and evaluators who are Asian nationals, perhaps even residing in Asia. This approach to data is indicative of the field's failure to develop theoretical and analytical specificity regarding dimensions of difference in ways that are functional to policy and program planners, including within evaluation as an industry.

Educational institutions and other training programs, philanthropy, and AEA (through its TIGS) can advance the intersectional collection of administrative data that could be meaningfully disaggregated within AEA and in surrounding fields to document and track demographic trends in organizational leadership. This would advance research *on* evaluation, evaluators, and surrounding systems from a critical perspective rather than just *for* them. Achieving recognition as an employment category—it is worth noting—would have ramifications for how the field approaches its professional ethics, credentialing, and training.

Data justice. Also at the structural level, the professional association, philanthropy, government agencies, educational institutions, and large evaluation firms

can coordinate efforts to support the development of evaluation theory and practice around the knowledge and collective action of otherized groups by underwriting or developing case studies, educational and professional development materials and opportunities, and data visualization strategies that historicize and contextualize difference and inequality beyond individual lives and individual lifetimes. They can underwrite and engage in the development of dynamic and multi-dimensional approaches to data disaggregation and visualization that demonstrate intersectionality and fluidity with respect to systems of oppression, difference, and identity.

Evaluation education and other training. Interviewees raised the role of educational institutions, as reported in Chapter Seven, as did the BDI and Tracking Transformation reports discussed in Chapter Six. BDI included among its recommendations a Council of Evaluation Training Programs to serve as a forum to discuss issues of diversity and cultural competence as they relate to training. Tracking Transformation found that this recommendation had not been fulfilled. Additionally, more than one interviewee mentioned the positive role that a faculty appointment plays in the ability of TIG and committee members representing racially otherized groups to sustain their leadership and be considered credible within a TIG and the professional association as a whole. This can make such participation inaccessible to practitioners and junior faculty. Educational institutions, AEA, and philanthropy can consider ways for AEA TIG and Board/committee participation to serve the academic, professional, and financial interests of faculty representing racially otherized groups in the same way that teaching, publication, and internal service are currently intended to. They can also consider ways for such participation to be accessible and immediately beneficial to practitioners.

Critical intersectionality. Despite the field's origins in educational research—including the eugenics movement (Besag, 1981)—on one hand, and in Great Society programming—including the War on Poverty and the Moynihan Report (1965) on which the "war" was based—on the other, the field of evaluation has failed to develop an analysis of how systems of cis-hetero-patriarchy, capitalism, and white supremacy work

together to reinforce each other in its literature, ethical standards and guidelines, graduate curricula, professional development offerings, or competencies. Indeed, *The Case for National Action* inherently implicates not just race and socio-economic class, but also gender, heteronormativity, and ability status in its characterization of "the Negro family" as pathological because it is supposedly matriarchal (Moynihan, 1965). As noted in Chapter One, a search of U.S. peer-reviewed evaluation literature as of December 2018 identified exactly six articles that contained the phrase "economic justice" and another six that contained the phrase "racial justice." In recent efforts to develop competencies as part of the field's efforts to professionalize, professional association members rejected the expectation that evaluators demonstrate competence with respect to "social justice" (AEA, 2018a).

Educational institutions and other training programs can increase their role in the production and consumption of—supply of and demand for—knowledge about difference and inequality that is historicized, contextualized, and critically intersectional, which they can build in the next generation of evaluators. By working with philanthropy and AEA, through its TIGs, educational institutions can fund and conduct research on existing training and educational materials as well as develop new materials—teaching cases, journal articles, and textbooks—that delve into the types of questions that Millett raised. They can create space within evaluation for a critical body of literature.

Funding and contracting. Philanthropy and government agencies can continue engaging with leaders in evaluation and in educational and training programs as both a consumer and producer of knowledge related to the role of evaluation in enabling collective action among otherized groups. They can create more opportunities for the exchange of different types of capital among constituent-led groups, evaluators, and funders that are designed for mutual, reciprocal capacity-building and critical reflection. They can ensure that a percentage of all their funding and contracts goes to efforts that are led by members of otherized groups and that address structural asymmetries as opposed to individual level interventions. Such a priority would, necessarily, require accompanying methodologies and systems for organizations to collect and track data

about staff's demographic characteristics, professional development, compensation packages, job titles, and decision-making authority. The D5 Coalition, Philanthropic Initiative for Racial Equity, and Candid (formerly Guidestar.com)—a web-based repository of information on all nonprofit organizations based in the USA—as well as various charity watchdogs have started this effort within the last decade.

The above effort applies only to nonprofit organizations, however—not to for-profit firms or government agencies. Moreover, even among nonprofits, reporting on such data is voluntary. Furthermore, collecting and reporting such data is inaccessible to many grassroots organizations that are led by constituents who are racially otherized. Large, white-led organizations that can allocate part of their budgets and staff to development and public relations are again at a relative advantage. Philanthropy and government agencies can ensure that organizations led by otherized groups have both the technological and human infrastructure necessary to collect and report on relevant data in ways that are meaningful from the perspective of understanding racial stratification within evaluation and surrounding systems by funding organizations' efforts to engage in critical inquiry about their own work, on their constituents' own terms.

Organizational structures. Additionally, philanthropy can work with AEA and educational institutions to advance research into and experimentation with organic, nonhierarchical organizational structures. This is as important at large, white-led firms as it is in grassroots organizations that struggle to be perceived as legitimate. In both cases, it would allow entry-level staff and emerging evaluators with close ties to families and communities representing otherized groups to learn about, participate in, and influence organizational processes and decisions rather than simply using their phenotype, relationships, and cultural knowledge to collect and transport data whose purpose, analysis, and interpretation have been pre-fabricated. Alternative organizational structures would fundamentally shift asymmetrical dynamics wherein the skills and identities—relational, human, and cultural capital—of otherized evaluators at all experience levels often remain at lower levels of organizations where they are commodified and exploited rather than engaged to effect structural change.

Workforce development. Finally, educational and training programs can work with AEA TIGs and philanthropy to honor the analysis of those most intimately familiar with systems of oppression by working with local governments' workforce development agencies, which offer training for temporarily displaced workers as well as members of groups who are chronically unemployed and systematically excluded from the formal economy, to create evaluation training opportunities appropriately tailored to public programs' eligibility and compliance requirements. These could be offered through nonprofit- and government-administered employment services that are part of many government assistance and re-entry programs such as Unemployment Insurance, Temporary Assistance for Needy Families (TANF), and Refugee Resettlement. They would provide those who have completed their high school diploma or GED and had direct personal experience with the justice system or public assistance with stackable evaluation-related credentials and income-generating opportunities. The income-generating opportunities in evaluation would allow them to apply their understanding of systemic oppression to public and nonprofit agencies rather than working in low wage jobs that reinforce their status as surplus labor, which is easily exploited and marginalized.

Conclusion of Chapter Nine

Chapter Nine discussed the implications for each of four levels of racialization: internalized, interpersonal, institutional, and structural. At the internalized level, it offered ways to think about relations as already existing and simply needing to be named, generally as oppressive or enabling, before they can be shifted by thinking materially about the exchange of various types of capital—including through the currencies of whiteness or discourse. At the interpersonal level, it offered issue-oriented mutual learning and reflection as an alternative to one-directional capacity building.

At the institutional and structural levels, it offered a matrix for institutions to use in thinking systematically through the impact that their business-as-usual is having on racially otherized groups, regardless of their intent. These included interpersonal solutions, which reinforce gatekeeping roles, to institutional and structural problems

regarding access. It also offered ways that AEA could work internally—through the annual conference and other professional development opportunities, with its TIGs, and with its journals—as well as externally with educational institutions, training programs, philanthropy, and government to deepen and expand understanding of difference, identity, and systemic oppression within the field and AEA membership. This is particularly necessary for disaggregating and visualizing identity intersectionally and critically. Overall, these involved creating space for evaluators representing otherized groups to develop necessary relationships, scholarship, and practice. Mechanisms proffered included business development for racially otherized evaluators, affirmative hiring policies, access to relevant data on the field and related industries, alternative organizational structures, and workforce development efforts.

Conclusion of Dissertation

Regardless of what they may look like and how they may have grown up—the social groups that they may come from or may currently belong to—all evaluators are in a position to reflect critically on how the exchange of capital takes place through their specific role in the larger system of evaluation and its surrounding industries. What internalized, interpersonal, institutional, or structural changes can each evaluator make to ensure that the exchange is symmetrical and that feedback loops no longer circulate capital exclusively among those who have accumulated it at the expense of others?

Using critical theories of systemic oppression and system dynamics, this dissertation has examined the construction of racialized difference in and through evaluation. Textual analysis, diachronic analysis, archival analysis, and analysis of interviews with those most closely involved in the field's construction of racialized difference yielded four discursive patterns: (1) minoritization and ambivalence toward whiteness; (2) the invocation of diversity and inclusion; (3) the replacement of race with culture; and (4) the rise of and decoupled relationship between indigeneity and colonization. Interviews with those most closely involved in the field during the periods of relevant decisions and processes suggested that the professional association did and can continue to play a role in the field's construction of difference, through vehicles like

the annual conference, themed journal editions, topical interest groups, and decision-making bodies. Interviewees tied each of these to interpersonal relationships as well as macro-level, structural dynamics.

Results suggest that mechanisms and actors already existing within evaluation's professional association supported the development of the four discursive patterns named above as characterizing evaluation's construction of racialized difference. They could potentially continue to do so. The extent to which they do so henceforth in ways that amplify otherized groups' collective action toward structural change or counteract such efforts by reinforcing individualized, dehistoricized, and decontextualized constructions of difference and inequality may depend on the extent to which AEA chooses to work with other institutions representing super-ordinated groups (educational institutions, philanthropy, government, evaluation firms) to cultivate enabling conditions with TIGs and other bodies representing otherized and sub-ordinated groups. This choice rests on the field's professional identity: Will evaluation acknowledge its close relationship—indeed, overlapping membership—with philanthropy, higher education, and government? Will it continue using phrases like "social betterment," "equity," and "justice" in its literature and ethical guidelines? Or will it claim to be at the mercy of its clients and subject to the whims of market forces? Whose socio-economic interests does each identity advance?

This dissertation concluded by offering possibilities for the professional association—particularly at the TIG level—educational institutions and training programs, large firms, government, and philanthropy to work more intentionally with each other in ways that challenge oppressive dynamics and the reproduction of racialized difference. Each of these actors depends on one or more types of capital from the others—and each contributes one or more types of capital to the others—in the process of fulfilling its function within the larger system. While philanthropy and government agencies can easily marginalize individual evaluators or evaluation firms for their ethical positions, they cannot as easily marginalize the entire field of evaluation as a collective represented by AEA. This is true even if some increasingly engage related but less

expensive occupational roles (e.g., data visualization specialists, management analysts, database technicians) instead of engaging evaluators—and even if some do so in a deliberate attempt to circumvent ethical discussions. In her paper on leverage points, systems theorist Meadows advised not bothering with reactionaries (Meadows, 1999).

As noted in the section addressing implications for the structural level, pivotal to AEA's coordination among related institutions of education, philanthropy, and government is its demonstration of leadership through an explicit articulation of purpose, which Meadows considered among the most sustainable and most difficult points of intervention.

> The shared idea in the minds of society, the great big unstated assumptions—unstated because unnecessary to state; everyone already knows them—constitute that society's paradigm, or deepest set of beliefs about how the world works… Paradigms are the sources of systems. From them, from shared social agreements about the nature of reality, come system goals and information flows, feedbacks, stocks, flows and everything else about systems (1999, pp. 17-18).

Referring to Thomas Kuhn, Meadows went on to ask,

> [H]ow do you change paradigms? … In a nutshell, you keep pointing at the anomalies and failures in the old paradigm, you keep speaking louder and with assurance from the new one, you insert people with the new paradigm in places of public visibility and power….
>
> Systems folks would say you change paradigms by modeling a system on a computer, which takes you outside the system and forces you to see it whole. We say that because our own paradigms have been changed that way. (1999, p. 18)

This dissertation was written in an attempt to offer a view of evaluation's whole system, from the outside, in the explicit hope that such a view may change its paradigm.

EPILOGUE

Late in my doctoral program in Evaluation Studies, in the melancholia that has always accompanied the period of transition from late summer into early fall for me, my immediate and extended family gathered in one of the few Malayalee restaurants in the USA to celebrate even more changes than usual at that time of year. The adolescent version of my evaluator self—anticipating my upcoming birthday and the impending schoolyear—would have been calculating grade point averages within and across subject areas and establishing goals for the next grade. Fortunately, our own children were enjoying every last bit of summer before school started in full force. We were all rejoicing in my successful defense and the upcoming completion of my dissertation. I shared the recent news that both sessions that I had proposed to the American Evaluation Association's annual conference, which was to take place in the city of my longtime residence—Minneapolis, Minnesota—had been accepted, to my surprise considering their focus on white supremacy.

On hearing that my family was visiting from Washington, D.C., the restaurant owner began discussing U.S. politics with us. Three years into the violent separation of refugee children from their parents at the border of the country of my birth and upbringing, we opined about each of the 2020 presidential candidates. In the process of discussing Tulsi Gabbard, we learned that the restaurant owner, while opposed to the party in power in the USA, supported the very similar party in power in India—the country of my ancestry, at the cusp of Hindu nationalists' passage of anti-Muslim citizenship laws.

Before dinner was over, I received an email from one of the most illustrious evaluation scholars inviting me to share the stage with him during one of the conference plenaries. Specifically, he asked me to join him in song—about the power of evaluation. I was welcomed to address the issue of white nationalism in my verse—one out of a total of three—for which I could write the lyrics and select the images. I had shared other professional associations' explicit condemnation of white nationalism in relation to their

work on EvalTalk—AEA's evaluation list-serv—and raised the need for evaluators to do so as well. While flattered by the possibility of a platform visible to 4,000 evaluators, I was apprehensive. My concern was not that I would foreclose professional opportunities. My concern, which I shared with the scholar, was that we would be making light of it. White nationalism does not lend itself to the audience sing-along that is customary at such plenaries. My concern was that we would not be doing the topic justice. He said it would serve as consciousness raising for 95 percent of the audience, who had likely never contemplated it at all. The thought of seeing White people laughing at brown people's pain and fear sent me back to the sick feeling I used to get when I experienced it in grade school and then again in college, until the extent of my focus on racial justice effectively segregated me from those inclined to laugh. After considerable discussion at home and at work, I decided to let the scholar decide based on the lyrics that I wrote and images that I chose. When he did not object to them, I cautiously agreed.

The song was scheduled for Friday's plenary. I had six other conference-related commitments, five of which would take place before it would.

Infantilization

On Tuesday, I co-facilitated an all-day pre-conference professional development workshop that I entitled "The Revolution Will Not Be Culturally Competent (or Diverse…or Inclusive)," based on the results of this dissertation. In it, I reprised "The Revolution Will Not Be Culturally Competent," which I had originally presented as a conference session seven years earlier—the last time the conference was in Minneapolis. In 2012, fortunate to have the annual conference in my adopted home, I assembled a cast of local artists and activists to perform the story of my experience in Bangladesh that I shared in the Prologue as stop-action theater, wherein audience members are encouraged to interrupt scenes and play characters' roles to change the outcome of the plot. This approach to theater as a forum for social change is indigenous, but not unique, to India.

I had put out an open call for partners to unpack the performance on the MIE TIG's list-serv. One person, a woman of color, expressed interest. In AEA proposal terms, she would serve as the discussant.

I had titled the performance and corresponding session intentionally: In homage to "The Revolution Will Not Be Televised," a song by Gil Scott-Heron that exposed the illusion that consumerism was power, and "The Revolution Will Not Be Funded," a book by INCITE! that exposed the illusion that nonprofit service provision was social change.

The discussant did not understand the references, though. She said this openly: "The revolution *will* be culturally competent?" She later expressed concern about how the piece would be received based on an email she shared from a member of her firm's advisory board—a prominent scholar in the field of evaluation who identifies as White:

> I hope this finds everything well at your end.
>
> I was looking over the AEA conference program over coffee this morning and I noticed that you are participating as discussant in an interesting Think Tank, The Revolution will not be Culturally Competent. I hope that this catchy title doesn't position critical theory in opposition to cultural competence. I think that would be a mistake and also a serious misreading of cultural competence literature. Critical theory is in fact one epistemology from which cultural competence can be approached (see Williams, 2006). Many authors have argued persuasively that interrogation of power needs to be part of cultural competence, and we tried to make this explicit in the AEA Public Statement, so I hope that the creator of this session isn't building a straw person argument against cultural competence.
>
> Lots of good stuff on the program this year. I think it will be a really lively conference. I just wanted to weigh in with my two cents, since I think people often react to the language of cultural competence (which has its limitations) in a way that gives others permission to ignore it.
>
> Can't wait to see you in Minneapolis.

The advisory board member never engaged with me directly about the session. The discussant, her protégé, later served on AEA's Board of Directors.

This year, with the luxury of a full day as opposed to the 45-minute conference session I had in 2012, I sandwiched scenes of the play in between content from my dissertation. As in 2012, assembling the cast, rehearsing, and performing the play such that workshop participants caught the subtleties required engaging in considerable group discussion with the cast and co-facilitators—most of whom identified as BIPOC evaluators. The discussions were life-giving and possibly life-changing.

Cast members said they did not want them to end. They asked for an additional session after the workshop took place to debrief. At it, they announced that they wanted to convene local evaluators who identified as BIPOC. We successfully gathered more than 20 such evaluators in person and remotely. Many more expressed interest, but were unavailable. The group who participated that day came up with three pages of topics that they would like the group to address collectively and had scheduled the next meeting within a week of the first, despite the holidays.

Meanwhile, an hour into the professional development workshop, after we had introduced ourselves, but before we had covered any content, we went on our first break. During the break, I was accosted in the bathroom by one of the participants. After having paid nearly $200 for the workshop, the abstract of which detailed its content, she asked me—in the bathroom—why I was talking about race. "Eth-ni-ci-ty," she enunciated as she slowly shook her head back and forth with a charitable smile and note of finality. I told her that she could stay for the workshop to find out or she could change her workshop registration and turned to leave. She followed me out of the bathroom and kept trying to talk to me. Later when I saw her name on the sign-in sheet, I learned that she works at the director level for the National Science Foundation.

Cultural Imperialism

Although I was drained by the end of the generally well-received workshop, I felt compelled to spend as much of Tuesday evening and Wednesday morning as possible at the meeting of the International Society of Evaluation Education (ISEE), which has the potential to shape evaluation education significantly. It was scheduled immediately before the annual conference itself. Once there, I noticed that while the group was called "international," those present overwhelmingly represented European settlers of colonial states. Potential future meetings were also proposed to be on settler colonial states. Participants struggled to imagine how to teach students of evaluation about social justice, seemingly unaware of the literature on justice that is available in neighboring disciplines as well as the necessity to produce such literature within evaluation. I wondered if my face revealed my identification with community organizers and community organizing as

I heard during one table's share-back that many people in evaluation shy away from the prospect of advocacy, because advocacy is "loud and boisterous." Advocacy need not be loud and boisterous, though, the person reporting back said. It can be soft.

On Wednesday, I hosted a meeting I had intentionally arranged between a highly respected African American evaluation scholar and the younger, emerging evaluators—all of whom identify as BIPOC—who work at Rainbow Research. In it, the scholar shared what he described as an important lesson he gained from one of his professors: Evaluation is not his life. Evaluation can be a vehicle for achieving what he pursues in life. The scholar's pursuits, like mine, included liberation. I wondered if the ISEE meeting participant would have found those of us in that room, whose survival has depended on advocacy, loud and boisterous.

On Thursday, after presenting a session on how evaluation firms can build racial, economic, gender, and disability justice—using frameworks from this dissertation—I spent the evening at the Advocacy and Policy Change (APC) TIG meeting, on their invitation. The TIG had made a public statement about white nationalism in direct and explicit response to my efforts on EvalTalk. They had further urged other TIGs and AEA more broadly to do the same. TIG members were going to spend the meeting discussing how white nationalism shows up in their work. With my sleepy child in tow, I left the meeting early, and as I walked out, the TIG chair told me individually what he said he planned to say publicly at the meeting's closing: One person can create change. As grateful as I was to hear that, and for my daughter to hear it, my gratitude did not prevent me from noting that change required not just one person to say something, but also one person to hear it. Thinking of many of those I interviewed for this dissertation, I said that people have been saying things long before me. Until now, they have gone unheard. I silently wondered if it took the extreme manifestation of white nationalism to be heard, and if so, what exactly was it that people were hearing. Efforts to pre-empt the current extreme manifestation of white nationalism, by questioning the conflation of "American" with "white," were apparently neither loud nor boisterous enough to be heard in years past.

Exploitation

Finally, Friday morning came. The three speakers, including the discussant from 2012, shared their top priority areas, inviting audience members to re-think their own paths and fuel discussions about the evaluation community's collective vision of the future of evaluation. As I listened to the speeches, awaiting my cue to go on stage, I heard "structural change." Then I heard a presumably illustrative success story wherein impoverished African American communities received $100 each to implement community-led projects in response to high rates of police violence and maternal mortality. The communities had decided to root their work in love. This was the moral of the story: Love could serve as a theory of change. And $100 could go far.

We singers immediately followed, and my verse ended up being first:

> Prophets of doom give white nationalists room
> Evaluators, help liberation bloom!
> Together we can decolonize,
> Honor Black lives,
> Listen for truth through history's lies.

Despite being set to probably the most memorable beat of all time, Queen's "We Will Rock You," I had trouble feeling it in my body and, apparently, it showed. I came off the stage to see a text from a new friend. She saw right through my performance and offered her support.

Meanwhile, Twitter was afire with video, still shots, and quotations from the #Eval19 Friday plenary. "Structural change" and "love" were everywhere. Black love—even with $100—cannot stop bullets, though. Black love and $100 do not constitute structural change unless that love is organized and directed against the carceral state. Structural change requires a change in the way that capital flows through the privatization of justice and health. It does not require a change in Black people. Black love, while palatable for White consumption, and a mere $100 demand almost no change in existing structures. They demand nothing of Whites. Indeed, they feed the idea that Black people just need to do more with less. Only then did it occur to me that "structural change" had become co-opted, as Bob Covert noted happens routinely in his interview for this

dissertation. “Structural change”—particularly when juxtaposed by a Black woman with “Black” and “love”—allows evaluators to feel revolutionary without changing a thing.

Marginalization

On this realization, I went to take my friend up on her offer of support. She and I had initially met through the interview process for my dissertation and had become better acquainted at the Culturally Responsive Evaluation and Assessment (CREA) conference, the theme for which was intersectionality, where we attended each other’s presentations.

CREA’s call for proposals specifically invited “critical reflection” sessions that “deconstruct dynamics of power.” My session, entitled “Identity and the Oppression Olympics: Intersectional Evaluation as a Vehicle for Building Solidarity and Collaboration”—having drawn heavily from the content of this dissertation—barely got in. It received the following “likely reject” review:

> This proposal is interesting. In some areas, it has promise, yet it mainly begins by criticizing just about everything related to evaluation, except for the author’s own ideas…. This does not mean that the author does not have potentially good ideas, but there is no reason to basically condemn everything that does not speak to the author’s points of view. It is also possible to discuss the dearth of wide perspectives, without being condescending or overly criticizing. Black Feminist thought [*sic*] is absolutely critical, and it should exist, but it can do so without demonizing other perspectives.

I sometimes wonder if the reviewer expected me to feel relieved that they thought Black Feminist Thought should exist and to take note of how they thought it should.

My friend framed her session in terms of decolonization. Having been invited to share predicaments we faced in our work, participants soon filled the space with tearful stories of undisguised racism in the context of evaluation that we had never before uttered in an evaluation space. Participants lovingly acknowledged each other’s pain and power. We all asserted the need for healing spaces in all evaluation fora.

Thus, I worked with her, among others, to create a healing and respite space for BIPOC attendees at AEA’s annual conference this year. Some organizers of the space were motivated in part by a performance in the previous year’s opening plenary, which attendees from racially otherized groups found (re)traumatizing. While AEA’s board

justified the performance as "research-based" and provided a link to the performer's website without taking the opportunity to educate its membership on the history of black-face and minstrel shows, AEA did agree to provide a space at this year's conference, which allowed not only for healing and respite but also for connection and creativity to grow organically among members of differently otherized groups.

Violence

My friend and I were motivated not necessarily by the previous year's plenary, however, but rather by years of similar experiences in numerous evaluation spaces, where we found no appropriate opportunity or critical mass of politicized evaluators of color with whom to discuss them. The annual conference of 2019 was the tenth AEA conference I had attended. My first year, I helped interview one of evaluation's founding fathers. He commenced the interview reminiscing about his start in evaluation, illustrated with jovial memories of the days that he explicitly noted were before Affirmative Action. He ended his interview by explaining his lack of interest and involvement in a related professional association's publication.

This evaluation forefather, who was part of the original invitation-only May 12 Group—so named explicitly to prevent evaluators who were not invited from feeling entitled to be part of it—explained his disinterest by lamenting that that professional association—which he noted was now led by an African American man—insisted on an open call for its publications. The forefather had been doing it as a service—he said, off the record—and was not going to start competing now. Although I had prepared a spreadsheet full of color-coded conference sessions to attend, I ended up spending much of the remainder of that year's conference alone in my hotel room wondering if he thought that as an Asian American, I agreed with him about Affirmative Action. Or did he not even notice that I was in the room? Or did he notice but not care? In retrospect, it would appear that Affirmative Action—as a form of structural change—is far more revolutionary, and poses a far greater threat to the White evaluation community, than Black love and $100 investments are or do.

As noted in Chapter Seven of this dissertation, it is phrases like "Affirmative Action" that bear the imprint of struggle crystallized in the Civil Rights Movement. In 2013, I participated in the Thought Leader series—an electronic discussion with featured leaders in the field of evaluation. The topic that the featured thought leader raised was my favorite: revolution. I contributed to the discussion, which by then included references to Rosa Parks, by asking what role evaluation could play in community organizing. The thought leaders participating—some of whom were living legends—responded that evaluation would lose its credibility if applied to community organizing efforts. Confused, I posted:

> Unless I'm misunderstanding something, I'm unclear how evaluating a constituent-led organizing effort against material inequity changes our role or compromises our credibility any more than evaluating a foundation- or government-funded program or service intended to achieve another goal (or even the same one)? Likewise, how does contributing research and evaluative thinking for social betterment (finding out if something is doing good—however and whoever defines it) make us advocates any more than contributing the same skills to a business's bottom line does? Ricardo Millett posed this question directly at a session at AEA 2012—asking specifically about the field's role in addressing racial and economic inequality.
>
> Again, no choice about where to invest one's energies is apolitical.

To this, the featured thought leader responded:

> Thank you for your message. I was glad to see that the topic of revolution engaged you and a couple others. Even if I end up remembering that I have to be satisfied with incremental change, I don't want to forget the outrage at injustice that is part of the fuel for the revolution that I'd like to see. I appreciate your thoughtful, forceful comments.

Through this interaction, I concluded that Rosa Parks, the Civil Rights Movement, and—indeed—revolution itself were abstract concepts, used loosely, among the thought leaders and discussion participants alike. This differs considerably from many of my dissertation interviewees, who participated in such movements themselves. I further noted the thought leader's use of the word "forceful." How does written text demonstrate force? Condescension? An "overly criticizing" tone?

I realized that there is no way of addressing structural change that is acceptable in tone, except not really addressing it at all. Even (or perhaps especially) well-researched, well-reasoned academic language—with no personal attacks, profanity, capital letters, or exclamation points—is considered loud and boisterous if it draws direct attention to systemic oppression. Gatekeepers routinely apply gendered and racially-coded labels like "loud and boisterous," in gendered and racially differentiated ways, to dismiss the points that those of us who speak with analytical specificity make about white supremacy, cis-hetero-patriarchy, capitalism, and ableism. They do not, however, characterize the analytically specific writing about race by Bob Covert and Ernie House—both of whom identify as White men—as "forceful," "overly criticizing," or "condescending."

This weaponization of tone makes navigating academia and knowledge-based fields like evaluation especially difficult for those of us who choose not to internalize and adopt—or, in some cases, who choose to unlearn—the oppressive patterns of super-ordinated groups and who choose instead to honor the spirit of interdependence and intersubjectivity with which we were raised. How do we value our emotional, cognitive, and political labor and acknowledge our multiple ways of knowing even as we engage the possibility that we may be wrong with curiosity and critical reflexivity?

Justice

The second time I attended AEA's annual conference, I fortified my ability to participate more fully by bringing my partner, thanks to a scholarship that my essay had been awarded by the MIE TIG. It covered my attendance. I delivered a presentation entitled "True Knowledge Confers Humility," in which I projected an excerpt from a reading that was assigned during my second semester of Evaluation coursework:

> As described by Shapiro, "the history of medical treatment until relatively recently is the history of the placebo effect."[11] Frank points out that, "Until the last few decades most medications prescribed by physicians were pharmacologically inert. That is, physicians were prescribing placebos without knowing it."[12] Today, this is still the probable basis for the so-called "faith" cures reported by almost all known religious or mystical sects—and *non-Western medicine in general*. (Suchman, 1967, p. 97, emphasis added)

I also projected an excerpt from materials distributed at a session during the previous year's conference. The handout read, "Community Context: The countries funded through the Africa Grants Program have the following characteristics" among which it was listed that "women have almost no power domestically or politically."

On the title slide of my presentation was written in Tamil: "A full pot does not gurgle," the English interpretation of which was the title of my presentation. I attempted to demonstrate the difference between such a worldview—in which knowledge is humility—and one rooted in the European Enlightenment, in which "knowledge is power." I drew from Hazel Symonette's work to clarify that without critical reflexivity, our phenotype, culture, and identity were insufficient to prevent ourselves from committing epistemic violence through evaluation discourse. I offered alternative conceptualizations of knowledge from Hindu and Buddhist philosophy that are not just culturally, but more importantly ideologically rooted in the notion that individual human beings' knowledge is inherently limited:

> What we know is a drop;
> what we do not know is an ocean.
>
> What we know is a grain of sand;
> what we do not know is the earth.
>
> When young, the stalks of wheat stand stiff;
> when mature, they bow with the weight of the grain.

In closing the session, the chair—a woman of color—summarized the point of my presentation as the importance of shared lived experience between program participants and program evaluators.

Counter-Narrative

Almost ten years ago exactly, I discussed my initial dissertation ideas with my first adviser, who is a person of color. I had wanted to look carefully at the racism, as I called it at the time, that I was experiencing through my evaluation readings—including Suchman's piece, which I named as an example—and evaluation interactions. How

pervasive was it? He assured me that that was not a dissertation topic. I was later told that you could find racism anywhere if you went looking for it.

I struggled for two years to find something within evaluation that approximated the type of study that I wanted to do. I started seeking disconfirming evidence for the pattern that I was seeing, wherein the most critical literature was farthest in the past—defying notions of "progress" rooted, once again, in the European Enlightenment. I first started summarizing the literature that I found that directly addressed race. Because I could not trust my perception—rightly, I believe—without some way to check it, I started tracking data about all the peer-reviewed evaluation literature. I also started investigating stratification economics—to see whether racial disparities are simply a function of socio-economic class as I was repeatedly told they were. I needed a problem statement but first I had to prove to myself and others that what I saw as a problem was indeed a problem.

I finally revealed the results of these efforts to my adviser in 2016. I had been working on them in my spare time while working full time to direct the research and evaluation efforts of a large, multimillion-dollar social service agency whose clients were disproportionately of color and where I was the highest ranking staff of color. As part of its balanced score card, it had decided to measure staff confidence in their program data through a survey every six months. Having felt hopeless for years, staff expressed hope for the first time—explicitly as a result of the inventory of programmatic data collection procedures that I had conceptualized and led. The day that survey results showed staff's increased confidence in their data, I was demoted. When I realized that I had been sacrificing my dissertation for an organization that had little interest in maintaining staff confidence in their data quality, I decided to quit and shift focus.

My adviser told me that my on-the-side efforts constituted essentially a dissertation's worth of research. My biggest question, however, was why: How could we make sense of these patterns regarding race? What factors contributed to them? How did the authors whose names repeatedly rose to the surface in the literature and AEA documents feel about them? Fortunate that so much of that history was still living, I wanted those closest to it to have a chance to tell the story before they retired.

As I increasingly honor multiple ways of knowing, a Malayalee saying that I grew up hearing came back to me today by way of a sister Malayalee artist-evaluator and acquaintance:

> You can wake
> someone who is
> sleeping, but
> you can never wake
> someone who is
> pretending to sleep.

Amid the many examples of people pretending to sleep within this Epilogue and the dissertation as a whole—whom we need not bother with—are examples of those who are continually waking up. As such, this dissertation's results—wherein existing institutional mechanisms were tied to both interpersonal relationships as well as structural forces—express hope rather than despair. All evaluation scholars and practitioners have the power to examine, name, and change how we are institutionally and structurally related. Regardless of phenotype, upbringing, or social group identification, we can all exercise agency to circulate capital within the collective action and knowledge production efforts of BIPOC evaluators.We all have the power to decenter whiteness and cultivate relations of interdependence and intersubjectivity with existing TIGs, CREA, and other groups that represent the interests of otherized evaluators. Regardless of our individual or institutional role, we can all act in solidarity with evaluators who are otherized—racially and along additional dimensions—through the group forming in Minneapolis, healing and respite spaces at evaluation events, or the APC TIG.

REFERENCES

Abrams, L. S., & Moio, J. A. (2009). Critical race theory and the cultural competence dilemma in social work education. *Journal of Social Work Education, 45*(2), 245-261.

AEA. (2004a). Cultural Reading of the Program Evaluation Standards. Retrieved from www.eval.org/p/cm/ld/fid=74.

AEA. (2004b). Guiding Principles for Evaluators. Retrieved from Respect for People https://www.eval.org/d/do/4204.

AEA. (2004c). Guiding Principles for Evaluators. Retrieved from Competence https://www.eval.org/d/do/4204.

AEA. (2004d). Guiding Principles for Evaluators. Retrieved from Responsibilities for General and Public Welfare https://www.eval.org/d/do/4204.

AEA. (2004e). Guiding Principles for Evaluators. Retrieved from Preface https://www.eval.org/d/do/4204.

AEA. (2011). Statement on Cultural Competence in Evaluation. Retrieved from http://www.eval.org/p/cm/ld/fid=92.

AEA. (2012). Evaluation 2012. Retrieved from http://archive.eval.org/search12/session.asp?sessionid=2225&presenterid=2408.

AEA. (2018a). Evaluator Competencies Brochure. Retrieved from https://www.eval.org/p/do/sd/topic=12&sid=14371.

AEA. (2018b). Guiding Principles for Evaluators. Retrieved from https://www.eval.org/p/cm/ld/fid=51.

AEA. (2018c). MIE TIG Background. Retrieved from http://comm.eval.org/multiethnicissuesinevaluation/new-item3

Aguilar, O., Dueñas, T., Flores, B., Godinez, L., Joy, H., & Zavala, I. (2005). *Fairness in Philanthropy Part I: Foundation Giving to Minority-Led Nonprofits*. Retrieved from Greenlining Institute http://greenlining.org/wp-content/uploads/2013/02/FairnessinPhilanthropyPartIFoundationGivingtoMinorityledNonprofits.pdf.

Ahmed, S. (2006). The nonperformativity of antiracism. *Meridians: Feminism, Race, Transnationalism, 7*(1), 104-126.

Ahmed, S. (2007a). A phenomenology of whiteness. *Feminist Theory, 8*(2), 149-168.

Ahmed, S. (2007b). The language of diversity. *Ethnic and Racial Studies, 30*(2), 235-256.

Ahmed, S. (2009). Embodying diversity: Problems and paradoxes for Black feminists. *Race Ethnicity and Education, 12*(1), 41-52.

Alkin, M. C. (2004) *Evaluation Roots*. Thousand Oaks, CA: Sage Publications.

Aponte-Soto, L., Ling Grant, D. S., Carter-Johnson, F., Colomer, S. E., Campbell, J. E., & Anderson, K. G. (2014). Championing culturally responsive leadership for evaluation practice. *New Directions for Evaluation, 2014*(143), 37-47.

Association for the Study and Development of Community (ASDC, now Community Science). (2002). *Building Diversity Plan*. Retrieved from American Evaluation Association http://www.eval.org/d/do/59.

Baer, W. C. (1992). *Race and the Shadow Market in Housing*. Los Angeles, CA: University of Southern California Press.

Bain, T., & Barnett, B. (2016). Insights into Nonprofit Compensation and Diversity. *Trends.* Retrieved from Guidestar.com https://trust.guidestar.org/insights-into-nonprofit-compensation-and-diversity.

Bandyopadhyay, S. K., & Pardasani, M. (2011). Do quality perceptions of health and social services vary for different ethnic groups? An empirical investigation. *International Journal of Nonprofit and Voluntary Sector Marketing, 16*(1), 99-114.

Bannerji, H. (2000). *The Dark Side of the Nation: Essays on Multiculturalism, Nationalism and Gender*. Toronto, ON: Canadian Scholars Press.

Barrett, K. H., & George, W. H. (2005). *Race, Culture, Psychology, and Law*. Thousand Oaks, CA: Sage.

Behrens, T. R., & Kelly, T. (2008). Paying the piper: Foundation evaluation capacity calls the tune. *New Directions for Evaluation, 2008*(119), 37-50.

Besag, F. P. (1981). Social Darwinism, race, and research. *Educational Evaluation and Policy Analysis, 3*(1), 55-69.

Bheda, D. (2011). En 'gendering' evaluation: Feminist evaluation but 'I am NOT a feminist!'. *New Directions for Evaluation, 2011*(131), 53-58.

Birman, D. (2007). Sins of omission and commission: To proceed, decline, or alter? *American Journal of Evaluation, 28*(1), 79-85.

Birnbaum, H., & Weston, R. (1974). Home ownership and the wealth position of Black and White Americans. *Review of Income and Wealth, 20*(1), 103-118.

Bivens, D. (1995). *Internalized Racism: A Definition*. Retrieved from Women's Theological Center http://www.racialequitytools.org/resourcefiles/bivens.pdf.

Bivens, D. (2005). What is internalized racism? In M. Potapchuk, S. Leiderman, D. Bivens, & B. Major (Eds.), *Flipping the Script: White Privilege and Community Building* (pp. 43-52). Retrieved from MP Associates, Inc. and Center for Assessment and Policy Development http://www.mpassociates.us/uploads/3/7/1/0/37103967/flippingthescriptmostupdated.pdf.

Blau, F. D., & Graham, J. G. (1990). Black-White differences in wealth and asset composition. *The Quarterly Journal of Economics, 105*(2), 321-339.

Bledsoe, K. L. (2014). Truth, beauty, and justice: Conceptualizing House's framework for evaluation in community-based settings. *New Directions for Evaluation, 2014*(142), 71-82.

Bledsoe, K. L. (2018). Introduction to the Section on Race and Evaluation. *American Journal of Evaluation 39*(4), 511-513.

BoardSource. (2012). *CEO Survey of BoardSource Members* (1). Retrieved from Nonprofit Governance Index: http://leadingwithintent.org/past-reports.

Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review, 62*(3), 465-480.

Bonilla-Silva, E. (2010). *Racism without Racists: Color-blind Racism and the Persistence of Racial Inequality in the United States* (3rd ed.). Lanham: Rowman & Littlefield Publishers.

Boris, E. T., & Steuerle, C. E. (2006). Scope and dimensions of the nonprofit sector. In W. W. Powell & R. Steinberg (Eds.), *The Nonprofit Sector: A Research Handbook* (2nd ed., pp. 66-88). New Haven, CT: Yale University Press.

Brown v. Board of Education. 347 U.S. 483 (1954).

Bryan, M. L., & O'Sullivan, R. (2014). Shaping a new generation of culturally responsive evaluators: Program director perspectives on the role of the internship experience. *New Directions for Evaluation, 2014*(143), 67-81.

Byng, M. D. (2013). You can't get there from here: A social process theory of racism and race. *Critical Sociology, 39*(5), 705-715.

Campbell, D. (1991). Methods for the experimenting society. *Evaluation Practice, 12*(3), 223-260.

Castagno, A. E., & Brayboy, B. M. (2008). Culturally responsive schooling for indigenous youth: A review of the literature. *Review of Educational Research, 78*(4), 941-993.

Cavino, H. M. (2013). Across the colonial divide: Conversations about evaluation in indigenous contexts. *American Journal of Evaluation, 34*(3), 339-355.

Chan, S. (1991). *Asian Americans: An Interpretive History*. New York, NY: Twayne.

Chen, H. T., & Turner, N. C. (2012). Formal theory versus stakeholder theory: New insights from a tobacco-focused prevention program evaluation. *American Journal of Evaluation, 33*(3), 395-413.

Chiteji, N., & Hamilton, D. (2005). Family matters: Kin networks and asset accumulation. In M. Sherraden (Ed.), *Inclusion in the American Dream: Assets, Poverty, and Public Policy* (pp. 87-111). New York, NY: Oxford University Press.

Choudhury, S. (2001). *Racial and Ethnic Differences in Wealth and Assets Choices* (4). Retrieved from Social Security Bulletin https://www.ssa.gov/policy/docs/ssb/v64n4/v64n4p1.pdf.

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation, 30*(4), 457-494.

Chouinard, J. A., & Hopson, R. (2016). A critical exploration of culture in international development evaluation. *Canadian Journal of Program Evaluation, 30*(3), 313-328.

Chouinard, J. A., & Milley, P. (2016). Mapping the spatial dimensions of participatory practice: A discussion of context in evaluation. *Evaluation and Program Planning, 54*(1), 1-10.

Christie, C. A., & Vo, A. T. (2011). Promoting diversity in the field of evaluation: Reflections on the first year of the Robert Wood Johnson Foundation Evaluation Fellowship Program. *American Journal of Evaluation, 32*(4), 547-564.

Civil Rights Act, 42 USCS § 2000e (1964).

Clayson, Z. C., Castañeda, X., Sanchez, E., & Brindis, C. (2002). Unequal power-changing landscapes: Negotiations between evaluation stakeholders in Latino communities. *American Journal of Evaluation, 23*(1), 33-44.

Cohen, R. (2012). Foundation portrait: Demographics and compensation. *Nonprofit Quarterly*. Retrieved from http://nonprofitquarterly.org/2012/03/22/foundation-portrait-demographics-and-compensation.

Cohen, R. (2014). Data snapshot on racial justice grantmaking. In Critical Issues Forum (Vol. 5, pp. 38-42). Philanthropic Initiative for Racial Equity: Moving Forward on Racial Justice Philanthropy. Retrieved from http://www.racialequity.org/criticalissues.html - critical5.

Collins, P. H. (2000). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment* (2nd Edition). New York, NY and London, UK: Routledge

Collins, P. M. & Hopson, R. K. (2014). Editors' Notes. *New Directions for Evaluation, 2014*(143), 1-7.

Collins, P. M., & Hopson, R. K. (2007). Building leadership development, social justice, and social change in evaluation through a pipeline program. In K. Hannum, J. Martineau, C. Reinelt, & C. f. C. Leadership (Eds.), *The Handbook of Leadership Development Evaluation* (1st ed., pp. 173-198). San Francisco, CA: Jossey-Bass.

Collins, P. M., Kirkhart, K. E., & Brown, T. (2014). Envisioning an evaluation curriculum to develop culturally responsive evaluators and support social justice. *New Directions for Evaluation, 2014*(143), 23-26.

Competence. (2016). OED Online. [Def. I. 1., II. 2. and II. 4a.]. Retrieved from http://www.oed.com.

Conley, D. (1999). *Being Black, Living in the Red. Race, Wealth, and Social Policy in America*. Berkeley, CA: University of California Press.

Conley, D. (2000). The racial wealth gap: Origins and implications for philanthropy in the African American community. *Nonprofit and Voluntary Sector Quarterly, 29*(4), 530-540.

Conner, R. F. (1989a). Is diversity an inherent part of good evaluation? *American Journal of Evaluation, 10*(3), 56-58.

Conner, R. F. (1989b). President's corner. *American Journal of Evaluation, 10*(4), 75-79.

Conner, R. F. (1989c). President's corner: AEA's power source. *American Journal of Evaluation, 10*(2), 93-97.

Conner, R. F. (1990). In praise of differences. *American Journal of Evaluation, 11*(1), 83-87.

Conner, R. F. (2004). Developing and implementing culturally competent evaluation: A discussion of multicultural validity in two HIV prevention programs for Latinos. *New Directions for Evaluation, 2004*(102), 51-65.

Conner, R. F., Fitzpatrick, J. L., & Rog, D. J. (2012). A first step forward: Context assessment. *New Directions for Evaluation, 2012*(135), 89-105.

Covert, R. W. (1987a). President's corner. *American Journal of Evaluation, 8*(1), 94-101.

Covert, R. W. (1987b). Ways of involving clients in the evaluation process. *American Journal of Evaluation, 8*(4), 83-87.

Covert, R. W. (1988). Ethics in evaluation: Beyond the Standards. *American Journal of Evaluation, 9*(1), 32-37.

Cram, F., Tibbetts, K. A., & LaFrance, J. (2018). Editors' Notes: A stepping stone in Indigenous Evaluation. *New Directions for Evaluation, 2018*(159), 7-16

Crazy Bull, C. (1997). A Native conversation about research and scholarship. *Tribal College Journal, 9*, 17-23.

Crenshaw, K. W. (1988). Race, reform, and retrenchment: Transformation and legitimation in antidiscrimination law. *Harvard Law Review, 101*(7), 1331-1387.

Crenshaw, K. W. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *The University of Chicago Legal Forum, 1989*, 139-167.

Creswell, J. W. & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., & ... Weiner, S. S. (1980). *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey-Bass.

Crusto, M. F. (2005). Blackness as property: Sex, race, status, and wealth. *Stanford Journal of Civil Rights & Civil Liberties, 1*, 51-169.

Dahler-Larsen, P., & Schwandt, T. A. (2012). Political culture as context for evaluation. *New Directions for Evaluation, 2012*(135), 75-87.

Dahler-Larsen, P., Abma, T., Bustelo, M., Irimia, R., Kosunen, S., Kravchuk, I., … Tshali, C. K. (2017). Evaluation, language, and untranslatables. *American Journal of Evaluation, 38*(1), 1-12.

Darity, W. A. (2002). Intergroup disparity: Why culture is irrelevant. *Review of Black Political Economy, 29*(4), 77-90.

Darity, W. A. (2005). Stratification economics: The role of intergroup inequality. *Journal of Economics and Finance, 29*(2), 144-153.

Darity, W. A., & Nicholson, M. J. (2005). Racial wealth inequality and the Black family. In V. C. McLoyd, N. E. Hill, & K. A. Dodge (Eds.), *African American Family Life: Ecological and Cultural Diversity* (pp. 78-85). New York, NY: Guilford Press.

Darity, W. A., Hamilton, D., & Stewart, J. B. (2015). A tour de force in understanding intergroup inequality: An introduction to stratification economics. *Review of Black Political Economy* (42), 1-6.

Datta, L.-E. (2011). Politics and evaluation: More than methodology. *American Journal of Evaluation, 32*(2), 273-294.

Davis, J. B. (2014). Stratification economics with identity economics. In. https://ssrn.com/abstract=2512370: Forthcoming Cambridge Journal of Economics.

Davis, J. E. (1992). Reconsidering the use of race as an explanatory variable in program evaluation. *New Directions for Program Evaluation, 1992*(53), 55-67.

Davis, J. E. (1999). Commentary: Advocacy, care, and power. *American Journal of Evaluation, 20*(1), 119-122.

Deaux, K., & Ullman, J. (1983). *Women of Steel*. New York, NY: Praeger.

DeGroff, A., & Cargo, M. (2009). Policy implementation: Implications for evaluation. *New Directions for Evaluation, 2009*(124), 47-60.

Delgado, R. (2002). Explaining the rise and fall of African American fortunes: Interest convergence and Civil Rights gains. *Harvard Civil Rights-Civil Liberties Law Review, 37*(2), 369-387.

Delgado, R., & Stefancic, J. (2001). *Critical Race Theory: An Introduction*. New York, NY: New York University Press.

Deloria, V. (1999). Ethnoscience and Indian realities. In B. Deloria, K. Foehner, & S. Scinta (Eds.), *Spirit and Reason: The Vine Deloria Reader*. Golden, Colo.: Fulcrum.

DeSilver, D. (2013). *Black Unemployment Rate is Consistently Twice that of Whites*. Retrieved from Fact Tank: News in the Numbers: http://www.pewresearch.org/fact-tank/2013/08/21/through-good-times-and-bad-black-unemployment-is-consistently-double-that-of-whites.

DeSilver, D. (2015). *On MLK Day, a Look at Black and White America*. Retrieved from Pew Research Center http://www.pewresearch.org/fact-tank/2015/01/19/on-mlk-day-a-look-at-black-and-white-america.

Desivilya Syna, H., Rottman, A., & Raz, M. (2015). Social justice in action: The contribution of evaluation to employment integration of a vulnerable population—The case of college graduates with learning disabilities. *New Directions for Evaluation, 2015*(146), 45-55.

Desmond-Harris, J. (2016, December 12). A pollster on the racial panic Obama's presidency triggered—and what Democrats must do now. *Vox Media*. Retrieved from https://www.vox.com/identities/2016/12/12/13894546/obama-race-black-white-house-cornell-belcher-racism.

Dillard, C. B. (2016). We are still here: Declarations of love and sovereignty in Black life under siege. *Educational Studies, 52*(3), 201-215.

Diverse. (2016). OED Online. [Def. 2]. Retrieved from http://www.oed.com.

Doane, A. W. (2003). Rethinking whiteness studies. In A. W. Doane & E. Bonilla-Silva (Eds.), *White Out: The Continuing Significance of Racism* (pp. 2-18). New York, NY: Routledge.

Doane, A. W. (2006). What is racism? Racial discourse and racial politics. *Critical Sociology, 32*(2-3), 255-274.

Doane, A. W., & Bonilla-Silva, E. (Eds.). (2003). *White Out: The Continuing Significance of Racism*. New York, NY: Routledge.

Doherty, C., Tyson, A., & Weisel, R. (2014). *Few Say Police Forces Nationally Do Well in Treating Races Equally: Most Have At Least 'Fair Amount' of Confidence in Local Police*. Retrieved from http://www.people-press.org/2014/08/25/few-say-police-forces-nationally-do-well-in-treating-races-equally/2.

Dueñas, T., Cano, E., & Mayorga, E. (2006). *Investing in a Diverse Democracy: Foundation Giving to Minority-Led Nonprofits*. Retrieved from Greenlining Institute http://greenlining.org/wp-content/uploads/2013/02/InvestinginaDiverseDemocracyFoundationGivingtoMinorityLedNonprofits.pdf.

Dunaway, K. E., Morrow, J. A., & Porter, B. E. (2012). Development and validation of the cultural competence of program evaluators (CCPE) self-report scale. *American Journal of Evaluation, 33*(4), 496-514.

Ellmer, R., & Olbrisch, M. E. (1983). The contribution of a cultural perspective in understanding and evaluating client satisfaction. *Evaluation and Program Planning, 6*(3-4), 275-281.

Espiritu, Y. L. (1992). *Asian American Pan-ethnicity: Bridging Institutions and Identities* (1st ed.). Philadelphia, PA: Temple University Press.

Evergreen, S. D. H., & Cullen, A. (2010). Moving to genuine: Credible cultural competence and stakeholder believability. *Journal of Multidisciplinary Evaluation, 6*(13), 130-139.

Evergreen, S. D. H., & Robertson, K. N. (2010). How do evaluators communicate cultural competence? Indications of cultural competence through an examination of the American Evaluation Association's Career Center. *Journal of MultiDisciplinary Evaluation, 6*(14), 58-67.

Fairclough, N. (2003). *Analysing Discourse: Textual Analysis for Social Research*. London, UK and New York, NY: Routledge.

Fairclough, N. (2012). Critical discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 9-20). Oxford; New York: Routledge.

Feagin, J. R. (2013). *The White Racial Frame: Centuries of Racial Framing and Counter-Framing.* New York, NY: Routledge.

Fein, E., Staff, I., & Kobylenski, S. (1993). The evaluator as power merchant. *Evaluation Practice, 14*(1), 9-15.

Fine, M. (1998). Working the hyphens: Reinventing self and others in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Landscape of Qualitative Research: Theories and Issues* (pp. 130-155). Thousand Oaks, CA: Sage Publications.

Fitzpatrick, J. L. (2012). An introduction to context and its role in evaluation practice. *New Directions for Evaluation, 2012*(135), 7-24.

Flood, R. L. (2010). The relationship of 'Systems Thinking' to Action Research. *Systemic Practice and Action Research, 23*, 269-284.

Fredericks, K. A., Deegan, M., & Carman, J. G. (2008). Using system dynamics as an evaluation tool: Experience from a demonstration program. *American Journal of Evaluation, 29*(3), 251-267.

Freeman, M., Preissle, J., & Havick, S. (2010). Moral knowledge and responsibilities in evaluation implementation: When critical theory and responsive evaluation collide. *New Directions for Evaluation, 2010*(127), 45-57.

Gargani, J. (2011). More than 25 years of the American Journal of Evaluation: The recollections of past editors in their own words. *American Journal of Evaluation, 32*(3), 428-447.

Gill, C. J. (1999). Invisible ubiquity: The surprising relevance of disability issues in evaluation. *American Journal of Evaluation, 20*(2), 279-287.

Gilmore, R. W. (2007). In the shadow of the shadow state. In INCITE! (Ed.), *The Revolution Will Not be Funded: Beyond the Non-profit Industrial Complex* (pp. 41-52). Cambridge, Massachusetts: South End Press.

Gitlin, T. (2018, November 16). The great race panic. *Dissent Magazine*. Retrieved from https://www.dissentmagazine.org/blog/the-great-race-panic.

Gittleman, M., & Wolff, E. N. (2004). Racial differences in patterns of wealth accumulation. *The Journal of Human Resources, 39*(1), 193-227.

Gong, J., & Wright, D. (2007). The context of power: Young people as evaluators. *American Journal of Evaluation, 28*(3), 327-333.

Gooden, S. (2014). *Race and Social Equity: A Nervous Area of Government*. Armonk, New York: M.E. Sharpe.

Greene, J. C., & Walker, K. (2001). Review of Evaluation: An integrated framework for understanding, guiding, and improving policies and programs. *Evaluation and Program Planning, 24*, 368-371.

Greene, J. C., Millett, R. A., & Hopson, R. K. (2004). Evaluation as a democratizing practice. In M. T. Braverman, N. A. Constantine, & J. K. Slater (Eds.), *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy* (pp. 96-118). San Francisco, CA: Jossey-Bass.

Grønbjerg, K. A. (1987). Patterns of institutional relations in the welfare state: Public mandates and the nonprofit sector. *Nonprofit and Voluntary Sector Quarterly, 16*(1-2), 64-80.

Grønbjerg, K. A. (2001). The U.S. nonprofit human service sector: A creeping revolution. *Nonprofit and Voluntary Sector Quarterly, 30*, 276-297.

Grossberg, L. (1996). Identity and cultural studies: Is that all there is? In S. Hall & P. Du Gay (Eds.), *Questions of Cultural Identity* (pp. 87-107). London; Thousand Oaks, CA: Sage.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Technology Research and Development, 29*(2), 75-91.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage.

Guinier, L. (2004). From racial liberalism to racial literacy: Brown v. Board of Education and the interest-divergence dilemma. *The Journal of American History, 91*(1), 92-118

Guo, B. (2010). Beyond the public safety net: The role of nonprofits in addressing material hardship of low-income households. *Nonprofit and Voluntary Sector Quarterly, 39*(5), 784-801.

Gupta, A. & Ferguson, J. (1992). Beyond "Culture": Space, Identity, and the Politics of Difference. *Cultural Anthropology, 7*(1), 6-23.

Hall, B., & Howard, K. (2008). A synergistic approach conducting mixed methods research with typological and systemic design considerations. *Journal of Mixed Methods Research, 2*(3), 248-269.

Hall, P. D. (2004). A historical perspective on evaluation in foundations. In M. T. Braverman, N. Constantine, & J. K. Slater (Eds.), *Foundations and evaluation:*

Contexts and practices for effective philanthropy (1st ed., pp. 27-50). San Francisco: Jossey-Bass.

Hall, P. D. (2006). A historical overview of philanthropy, voluntary associations, and nonprofit organizations in the United States, 1600-2000. In W. W. Powell & R. Steinberg (Eds.), *The Nonprofit Sector: A Research Handbook* (2nd ed., pp. 32-65). New Haven, CT: Yale University Press.

Hall, P. D. (2013). Philanthropy, the nonprofit sector & the democratic dilemma. *Daedalus, 142*(2), 139-158.

Hall, S., & Open University. (1997). *Representation: Cultural Representations and Signifying Practices*. London: Sage in association with the Open University.

Hamilton, D., & Darity, W. A. (2010). Can 'Baby Bonds' eliminate the racial wealth gap in putative post-racial America? *Review of Black Political Economy, 37*, 207-216.

Hancock, A.-M. (2013). Empirical intersectionality: A tale of two approaches. *University of California Irvine Law Review, 3*, 259-296.

Hankivsky, O., & Grace, D. (2015). Understanding and emphasizing difference and intersectionality in multimethod and mixed methods research In S. N. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (pp. 110-127). Oxford; New York: Oxford University Press.

Hansen, N. D., Pepitone-Arreola-Rockwell, F., & Greene, A. F. (2000). Multicultural competence: Criteria and case examples. *Professional Psychology: Research and Practice, 31*, 652-660.

Hanson, B. (2001). Systems theory and the spirit of feminism: Grounds for a connection. *Systems Research and Behavioral Science, 18*, 545-556.

Harrington, M. (1962). *The Other America: Poverty in the United States*. New York, NY: Macmillan.

Harris, C. L. (1995). Whiteness as property. In K. W. Crenshaw, N. Gotanda, G. Peller, & K. Thomas (Eds.), *Critical Race Theory: The Key Writings that Formed the Movement* (pp. 276-291). New York, NY: The New Press.

Hendricks, M. (1985). How can evaluators become more involved in the important funding and budgeting discussions that determine a program's scope and operations? *American Journal of Evaluation, 6*(1), 58-60.

Henretta, J. C. (1979). Race differences in middle class lifestyle: The role of home ownership. *Social Science Research, 8*(1), 63-78.

Henry, G. T. (2000). Why not use? *New Directions for Evaluation, 2000*(88), 85-98.

Henry, G. T., & Julnes, G. (1998). Values and realist evaluation. *New Directions for Evaluation, 1998*(78), 53-71.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24*(3), 293-314.

Hesse-Biber, S. (2013). Thinking outside the randomized controlled trials experimental box: Strategies for enhancing credibility and social justice. *New Directions for Evaluation, 2013*(138), 49-60.

Hesse-Biber, S. N., Rodriguez, D., & Frost, N. A. (2015). A qualitatively driven approach to multimethod and mixed methods research. In S. N. Hesse-Biber & B. Johnson

(Eds.), *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (pp. 3-20). Oxford, UK and New York, NY: Oxford University Press.

Hilliard, A. G. (1989). Kemetic (Egyptian) historical revision: Implications for cross-cultural evaluation and research in education. *American Journal of Evaluation, 10(*2), 7-23.

Hollinsworth, D. (2013). Forget cultural competence; ask for an autobiography. *Social Work Education: The International Journal, 32*(8), 1-13.

Hollway, W. (2010). Relationality: The intersubjective foundations of identity. In M. Wetherell & C. Talpade Mohanty (Eds.), *The SAGE Handbook of Identities* (pp. 216-232). London: Sage Publications Ltd.

Hood, S. (2000). Commentary on deliberative democratic evaluation. *New Directions for Evaluation, 2000*(85), 77-83.

Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. *New Directions for Evaluation, 2001*(92), 31-44.

Hood, S. (2004). A journey to understand the role of culture in program evaluation: Snapshots and personal reflections of one African American evaluator. *New Directions for Evaluation, 2004*(102), 21-37.

Hood, S. (2014). How will we know it when we see it? A critical friend perspective of the Graduate Education Diversity Internship (GEDI) program and its legacy in evaluation. *New Directions for Evaluation, 2014*(143), 109-121.

Hood, S. (2017). African Americans in the early history of evaluation in the United States: Contributions of Ambrose Caliver in the U.S. Office of Education. *American Journal of Evaluation*, 38(2), 262-274.

Hood, S., Hopson, R. K., & Kirkhart, K. (2015). Culturally Responsive Evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (eds). Handbook of Practical Program Evaluation (4th edition). San Francisco : Jossey-Bass & Pfeiffer Imprints, Wiley.

Hoover, H. (1922). *American Individualism*. Garden City, NY: Doubleday, Page & Company.

Hopson, R. K. (2001). Global and local conversations on culture, diversity, and social justice in evaluation: Issues to consider in a 9/11 era. *American Journal of Evaluation, 22*(3), 375-380.

Hopson, R. K. (2009). Reclaiming knowledge at the margins: Culturally Responsive Evaluation in the current evaluation moment. In K. Ryan, F. & B. Cousins, J. (Eds.), *The Sage International Handbook of Educational Evaluation* (pp. 429-446). Thousand Oaks, CA: Sage.

House, E. R. (1976). Justice in evaluation. In G. V. Glass (Ed.), *Evaluation Studies* (Vol. 1, pp. 75-100). Beverly Hills, CA: Sage Publications.

House, E. R. (1983). How we think about evaluation. *New Directions for Program Evaluation, 1983*(19), 5-25.

House, E. R. (1985). Undisciplined social science. *Evaluation News, 6*(2), 28-31.

House, E. R. (1987a). Evaluators, libs, and neocons. *Evaluation Practice, 8*(1), 64-67.

House, E. R. (1987b). The evaluation audit. *Evaluation Practice, 8*(2), 52-56.

House, E. R. (1990). Methodology and justice. *New Directions for Program Evaluation, 1990*(45), 23-36.

House, E. R. (1991). Evaluation and social justice: Where are we? In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and Education at Quarter Century* (Vol. Ninetieth Yearbook of the National Society for the Study of Education, pp. 233-247). Chicago, IL: University of Chicago Press.

House, E. R. (1995a). Principled evaluation: A critique of the AEA Guiding Principles. *New Directions for Program Evaluation, 1995*(66), 27-34.

House, E. R. (1995b). Putting things together coherently: Logic and justice. *New Directions for Evaluation, 1995*(68), 33-48.

House, E. R. (1997). Evaluation in the government marketplace. *Evaluation Practice, 18*(1), 37-48.

House, E. R. (1998). *Schools for Sale: Why Free Market Policies Won't Improve America's Schools, and What Will*. New York, NY: Teachers College Press.

House, E. R. (1999a). Evaluation and people of color—A response to Professor Stanfield. *American Journal of Evaluation, 20*(3), 433-435.

House, E. R. (1999b). Race and policy. *Education Policy Analysis Archives, 7*(16), 1-15.

House, E. R. (2001). Responsive Evaluation (and its influence on Deliberative Democratic Evaluation). *New Directions for Evaluation, 2001*(92), 23-30.

House, E. R. (2011). Conflict of interest and Campbellian validity. *New Directions for Evaluation, 2011*(130), 69-80.

House, E. R. (2014). Origins of the ideas in Evaluating with Validity. *New Directions for Evaluation, 2014*(142), 9-15.

House, E. R. (2017). Evaluation and the framing of race. *American Journal of Evaluation*, *38*(2), 1-23.

House, E. R. (Ed.) (1983b). *Philosophy of Evaluation* (Vol. 19). San Francisco, CA: Jossey-Bass.

House, E. R., & Howe, K. R. (1998). The issue of advocacy in evaluations. *American Journal of Evaluation, 19*(2), 233-236.

House, E. R., & Howe, K. R. (1999). *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.

House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation *New Directions for Evaluation, 2000*(85), 3-12.

Immigration and Nationality Act, H.R. 2580; Pub.L. 89-236; 79 Stat. 911. 89th Cong., 2nd sess. (1965).

INCITE! (Ed.) (2007). *The Revolution Will Not be Funded: Beyond the Non-profit Industrial Complex*. Cambridge, Massachusetts: South End Press.

Industry. (n.d.). U.S. Bureau of Labor Statistics. *Handbook of Methods*. Retrieved from Glossary: https://www.bls.gov/opub/hom/glossary.htm.

Irving, S. K., & Loveless, T. A. (2015). *Dynamics of Economic Well-Being: Participation in Government Programs, 2009-2012: Who Gets Assistance*. Retrieved from U.S. Census Bureau http://www.census.gov/2010census/data.

Jackman, M. R., & Jackman, R. W. (1980). Racial inequalities in home ownership. *Social Forces, 58*(4), 1221-1234.

Jani, J. S., Pierce, D., Ortiz, L., & Sowbel, L. (2011). Access to intersectionality, content to competence: Deconstructing social work education diversity standards. *Journal of Social Work Education, 47*(2), 283-301.

Jenkins, J. C. (2006). Nonprofit organizations and policy advocacy. In W. W. Powell & R. Steinberg (Eds.), *The Nonprofit Sector: A Research Handbook* (pp. 307-332). New Haven, CT: Yale University Press.

Johnson, W. & Packer, A. (1987). *Workforce 2000: Work and Workers for the Twenty-first Century*. Indianapolis, IN: Hudson Institute.

Johnston-Goodstar, K. (2012). Decolonizing evaluation: The necessity of evaluation advisory groups in indigenous evaluation. *New Directions for Evaluation, 2012*(136), 109.

Julnes, G., Mark, M. M., & Henry, G. T. (1998). Epilogue: Experience informing theory, supporting practice. *New Directions for Evaluation, 1998*(78), 103-104.

Karim, L. (2008). Demystifying microcredit: The Grameen Bank, NGOs, and neoliberalism in Bangladesh. *Cultural Dynamics, 20*(1), 5-29.

Keister, L. A. (2000). Race and wealth inequality: The impact of racial differences in asset ownership on the distribution of household wealth. *Social Science Research, 29*, 477-502.

Kelle, U. (2007). The development of categories: Different approaches in grounded theory. In A. Bryant & K. Charmaz (Eds.), *The Sage Handbook of Grounded Theory* (pp. 191-213). London, UK and Thousand Oaks, CA: Sage.

Kerr, S. (2006). First person, first peoples: A journey through boundaries. *American Journal of Evaluation, 27*(3), 360-369.

Khosla, N. (2009). The ready-made garments industry in Bangladesh: A means to reducing gender-based social exclusion of women? *Journal of International Women's Studies, 11*(1), 289-303.

Kim, H., & Andersen, D. F. (2012). Building confidence in causal maps generated from purposive text data: Mapping transcripts of the Federal Reserve. *System Dynamics Review, 28*(4), 311-328.

King, J. A. (2016). Taking what action for change? Dare evaluation build a new social order? *Visitor Studies, 19*(1), 3-11.

King, J. A., Cousins, J. B., & Whitmore, E. (2007). Making sense of participatory evaluation: Framing participatory evaluation. *New Directions for Evaluation, 2007*(114), 83-105.

King, J. A., Nielsen, J. E., & Colby, J. (2004). Lessons for culturally competent evaluation from the study of a multicultural initiative. *New Directions for Evaluation, 2004*(102), 67-80.

Kirkhart, K. E. (1995). Seeking multicultural validity: A postcard from the road. *Evaluation Practice, 16*(1), 1-12.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation, 31*(3), 400-413.

Kirkhart, K. E. (2011). Culture and influence in multisite evaluation. *New Directions for Evaluation, 2011*(129), 73.

Kochhar, R., Fry, R., & Taylor, P. (2011). *Wealth Gaps Rise to Record Highs Between Whites, Blacks, Hispanics*. Retrieved from Pew Research Center https://www.pewsocialtrends.org/2011/07/26/wealth-gaps-rise-to-record-highs-between-whites-blacks-hispanics

Kothari, U. (2005). From colonial administration to development studies: A postcolonial critique of the history of development studies. In U. Kothari (Ed.), *A Radical History of Development Studies: Individuals, Institutions and Ideologies* (pp. 47-66). London, UK: Zed.

Kothari, U. (2006a). An agenda for thinking about 'race' in development. *Progress in Development Studies, 6*(1), 9-23.

Kothari, U. (2006b). Critiquing 'race' and racism in development discourse and practice. *Progress in Development Studies, 6*(1), 1-7.

Kothari, U. (2006c). From colonialism to development: Reflections of former colonial officers. *Commonwealth and Comparative Politics, 44*(1), 93-107.

Kovach, M. (2010). *Indigenous Methodologies: Characteristics, Conversations, and Contexts*. Toronto, Ontario, Canada: University of Toronto Press.

LaFrance, J. (2004). Culturally competent evaluation in Indian Country. *New Directions for Evaluation, 2004*(102), 39-50.

LaFrance, J., Nichols, R., & Kirkhart, K. E. (2012). Culture writes the script: On the centrality of context in indigenous evaluation. *New Directions for Evaluation, 2012*(135), 59-74.

Lane, D. C. (2001a). Rerum cognoscere causas: Part II—Opportunities generated by the agency/structure debate and suggestions for clarifying the social theoretic position of system dynamics *System Dynamics Review, 17*(4), 293-309.

Lane, D. C. (2001b). Rerum cognoscere causas: Part I—How do the ideas of system dynamics relate to traditional social theories and the voluntarism/determinism debate? *System Dynamics Review, 17*(2), 97-118.

Laszlo, A., & Krippner, S. (1998). Systems theories: Their origins, foundations, and development. In J. S. Jordan (Ed.), *Systems Theories and A Priori Aspects of Perception* (pp. 47-74). Amsterdam: Elsevier Science.

Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange, 17*(4), 63-84.

Lawrence, S. (2009). *Social Justice Grantmaking II*. Retrieved from The Foundation Center foundationcenter.org/gainknowledge/research/pdf/socialjustice2009_highlights.pdf.

Lee, B. (1999). The implications of diversity and disability for evaluation practice: Commentary on Gill. *American Journal of Evaluation, 20*(2), 289-293.

Lee, E. (1997). Overview: The assessment and treatment of Asian American families. In E. Lee (Ed.), *Working with Asian Americans: A Guide for Clinicians* (pp. 3-36). New York, NY: Guilford Press.

Lee, K., & Gilbert, B. (2014). Embedding the Graduate Education Diversity Internship (GEDI) program within a larger system. *New Directions for Evaluation, 2014*(143), 97-108.

Lee, Y.-F., Altschuld, J. W., & Lee, L.-S. (2012). Essential competencies for program evaluators in a diverse cultural context. *Evaluation and program planning, 35*(4), 439-444.

Leiderman, S., Potapchuk, M., & Major, B. (2005). What is racism? In M. Potapchuk, S. Leiderman, D. Bivens, & B. Major (Eds.), *Flipping the Script: White Privilege and Community Building* (pp. 31-42). Retrieved from MP Associates, Inc. and Center for Assessment and Policy Development http://www.mpassociates.us/uploads/3/7/1/0/37103967/flippingthescriptmostupdated.pdf.

Leonardo, Z. (2004). The color of supremacy: Beyond the discourse of 'white privilege'. *Educational Philosophy and Theory, 36*(2), 137-152.

LeRoux, K. (2009). The effects of descriptive representation on nonprofits' civic intermediary roles: A test of the "racial mismatch" hypothesis in the social services sector. *Nonprofit and Voluntary Sector Quarterly, 38*(5), 741-760.

Lester, W. (1999). U.S. ethnic diversity grows. *Associated Press*.

Letichevsky, A. C., & Penna Firme, T. (2012). Evaluating with at-risk communities: Learning from a social program in a Brazilian slum. *New Directions for Evaluation, 2012*(134), 61-76.

Letiecq, B. L., & Bailey, S. J. (2004). Evaluating from the outside: Conducting cross-cultural evaluation research on an American Indian reservation. *Evaluation Review, 28*(4), 342-357.

Levine, A. & Levine, M. (1977). The Orient as an object of knowledge. *Evaluation Review*, 1(4), 515-542.

Levine, A., & Levine, M. (1977). The social context of evaluative research: A case study. *Evaluation Review, 1*(4), 515-542.

Lincoln, Y. & Guba, E. (2004). The roots of fourth generation evaluationl: theoretical and methodological origins. In Alkin, M. C. *Evaluation Roots* (pp. 226-241). Thousand Oaks, CA: Sage Publications.

Lincoln, Y. S. (1991). The arts and sciences of program evaluation. *Evaluation Practice, 12*(1), 1-7.

Lincoln, Y. S. (1994). Tracks toward a postmodern politics of evaluation. *Evaluation Practice, 15*(3), 299-309.

Lui, M., Robles, B., Leondar-Wright, B., Brewer, R., & Adamson, R. (2006). *The Color of Wealth: The Story Behind the U.S. Racial Wealth Divide*. New York, NY: The New Press.

Luna-Reyes, L. F., & Andersen, D. L. (2003). Collecting and analyzing qualitative data for system dynamics: Methods and models. *System Dynamics Review, 19*, 271-296.

Lynn, M., & Parker, L. (2006). Critical race studies in education: Examining a decade of research on U.S. schools. *The Urban Review, 38*(4), 257-290.

Maack, S. C., & Upton, J. (2006). Collaborative relationships in evaluation consulting. *New Directions for Evaluation, 2006*(111), 57-66.

Madison, A.-M. (1992a). Editor's Notes. *New Directions for Program Evaluation, 1992*(53), 1-4.

Madison, A.-M. (1992b). Primary inclusion of culturally diverse minority program participants in the evaluation process. *New Directions for Program Evaluation, 1992*(53), 35-43.

Madison, A.-M. (2007). New Directions for Evaluation coverage of cultural issues and issues of significance to underrepresented groups. *New Directions for Evaluation, 2007*(114), 107-114.

Manning, C., Bachrach, E., Tiedemann, M., McPherson, M., & Goodman, I. (2008). *Internal Scan Report to the Membership*. Goodman Research Group. Retrieved from American Evaluation Association http://www.eval.org/d/do/69.

Mark, M. M., & Henry, G. T. (2004). The Mechanisms and Outcomes of Evaluation Influence. *Evaluation, 10*(1), 35-57.

Mark, M. M., & Shotland, R. L. (1985). Stakeholder-based evaluation and value judgments. *Evaluation Review, 9*(5), 605-626.

Mark, M., M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An Integrated Framework for Understanding, Guiding and Improving Policies and Programs*. San Francisco, CA: Jossey-Bass.

Martinez-Moyano, I. J., & Richardson, G. P. (2013). Best practices in system dynamics modeling. *System Dynamics Review, 29*(2), 102-123.

Mathie, A., & Greene, J. C. (1997). Stakeholder participation in evaluation: How important is diversity? *Evaluation and Program Planning, 20*(3), 279-285.

Mathison, S. (2000). Deliberation, evaluation, and democracy. *New Directions for Evaluation, 2000*(85), 85-89.

Mathison, S. (2011). Internal evaluation, historically speaking. *New Directions for Evaluation, 2011*(132), 13-23.

McGill, L., T., Bryan, B., & Miller, E. D. (2009). *Benchmarking Diversity: A First Look at New York City Foundations and Nonprofits*. New York, NY: The Foundation Center.

Meadows, D. H. (1999). *Leverage Points: Places to Intervene in a System*. Hartland, VT: The Sustsainability Institute.

Menendian, S., & Watt, C. (2008). *Systems Primer*. Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity.

Merryfield, M. M. (1985). The challenge of cross-cultural evaluation: Some views from the field. *New Directions for Evaluation, 1985*(25), 3-17.

Mertens, D. M. (1995). Identifying and respecting differences among participants in evaluation studies. *New Directions for Program Evaluation, 1995*(66), 91-97.

Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation, 20*(1), 1-14.

Mertens, D. M. (2007). Transformative considerations: Inclusion and social justice. *American Journal of Evaluation, 28*(1), 86-90.

Mertens, D. M. (2013). What does a transformative lens bring to credible evidence in mixed methods evaluations? *New Directions for Evaluation, 2013*(138), 27-35.

Mertens, D. M. (2016). Assumptions at the philosophical and programmatic levels in evaluation. *Evaluation and Program Planning, 2016*(59), 102-108.

Mertens, D. M., & Hopson, R. K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. *New Directions for Evaluation, 2006*(109), 35-51.

Mertens, D. M., & Russon, C. (2000). A proposal for the International Organization for Cooperation in Evaluation. *American Journal of Evaluation, 21*(2), 275-283.

Mertens, D. M., Farley, J., Madison, A.-M., & Singleton, P. (1994). Diverse voices in evaluation practice: Feminists, minorities, and persons with disabilities. *Evaluation Practice, 15*(2), 123-129.

Midgley, J. (1998). Colonialism and welfare. *Journal of Progressive Human Services, 9*(2), 31-50.

Mills, C. W. (1994). Revisionist ontologies: Theorizing white supremacy. *Social and Economic Studies, 43*(3), 105-134.

Mohanty, C. T. (2003). Feminism Without Borders: Decolonizing Theory, Practicing Solidarity. Durham, NC: Duke University Press.

Moynihan, D. P. (1965). *The Negro Family: The Case for National Action*. Retrieved from http://web.stanford.edu/~mrosenfe/Moynihan's The Negro Family.pdf.

Mukherjee-Reed, A. (2014). Rights and development: A social power perspective. In R. M. Buchanan & P. Zumbansen (Eds.), *Law in Transition: Human Rights, Development and Transitional Justice*. Oxford, UK: Hart Publishing.

Myers, S. L., Jr. (2002). Analysis of race as policy analysis *Journal of Policy Analysis & Management, 21*(2), 169-190.

Myrdal, G. (1944). *An American Dilemma: The Negro Problem and Modern Democracy*. New Brunswick, NJ and London, UK: Transaction Publishers.

Nagarjuna & Inada, K. K. (1993). *Nagarjuna: A Translation of his Mūlamadhyamakakārikā, with an Introductory Essay*. Delhi, India: Sri Satguru Publications.

Nam, Y., Hamilton, D., Darity, W. A., & Price, A. E. (2015). *Bootstraps Are for Black Kids: Race, Wealth, and the Impact of Intergenerational Transfers on Adult Outcomes*.

National Urban Fellows. (2012). *Diversity Counts: Racial and Ethnic Diversity Among Public Service Leadership*. Retrieved from http://www.nuf.org/sites/default/files/Documents/NUF_diversitycounts_V2FINAL.pdf.

Nevo, D. (1982). The international context for research on evaluation. *Evaluation News, 3*(4), 73-75.

Noblit, G. W., & Jay, M. (2010). Against the majoritarian story of school reform: The Comer schools evaluation as a critical race counternarrative. *New Directions for Evaluation, 2010*(127), 71-82.

Nylund, D. (2006). Critical multiculturalism, whiteness, and social work. *Journal of Progressive Human Services, 17*(2), 27-42.

O'Keeffe, A. (2012). Media and discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 441-454). Oxford, UK and New York, NY: Routledge.

Occident. (2016). OED Online. [Def. A. 1, A. 2, and B]. Retrieved from http://www.oed.com.

Oliver, M. L., & Shapiro, T. M. (1989). Race and wealth. *Review of Black Political Economy, 17*(4), 5-25.

Oliver, M. L., & Shapiro, T. M. (1990). Wealth of a nation: At least one-third of households are asset poor. *American Journal of Economics and Sociology, 49*, 129-151.

Oliver, M. L., & Shapiro, T. M. (1995). *Black Wealth/White Wealth: A New Perspective on Racial Inequality* (1st ed.). New York, NY: Routledge.

Oliver, M. L., & Shapiro, T. M. (1997). *Black Wealth/White Wealth: A New Perspective on Racial Inequality*. New York, NY and London, UK: Routledge.

Oliver, M. L., & Shapiro, T. M. (2001). *Wealth and racial stratification* (N. R. C. U. S. C. o. B. a. S. S. a. Education Ed.). Washington, DC: National Academy Press.

Oliver, M. L., & Shapiro, T. M. (2006). *Black Wealth/White Wealth: A New Perspective on Racial Inequality* (10th ed.). New York, NY: Routledge.

Omi, M., & Winant, H. (1993). On the theoretical status of the concept of race. In C. McCarthy & W. Crichlow (Eds.), *Race, Identity, and Representation in Education* (pp. 3-10). New York: Routledge.

Omi, M., & Winant, H. (2014). *Racial Formation in the United States*. New York, NY: Routledge.

Ong, P. M., & Grigsby, J. E. (1988). Race and life-cycle effects on home ownership in Los Angeles 1970 to 1980. *Urban Affairs Review, 23*(4), 601-615.

Orient. (2016). OED Online. [Def. A. 1a, A. 2, A. 3, B. 2, B. 3a]. Retrieved from http://www.oed.com.

Orwin, R. G., Campbell, B., Campbell, K., & Krupski, A. (2004). Welfare reform and addiction: A priori hypotheses, post hoc explorations, and assisted sensemaking in evaluating the effects of terminating benefits for chronic substance abusers. *American Journal of Evaluation, 25*(4), 409-441.

Ostrander, S. A. (1989). Private social services: Obstacles to the welfare state? *Nonprofit and Voluntary Sector Quarterly, 18*(1), 25-45.

Ostrower, F. (2007). *Findings on Performance and Accountability from the First National Representative Study*. Washington, DC: The Urban Institute Center on Nonprofits & Philanthropy.

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review Sociology., 34*, 181-209.

Parcel, T. L. (1982). Wealth accumulation of Black and White men: The case of housing equity. *Social Problems, 30*(2), 199-211.

Park, Y. (2006). Constructing immigrants: A historical discourse analysis of the representations of immigrants in U.S. social work, 1882-1952. *Journal of Social Work, 6*(2), 169-203.

Parker, L. (2004). Commentary: Can critical theories of or on race be used in evaluation research in education? *New Directions for Evaluation*, 2004(101), 85-93.

Patton, M. Q. (1985). Cross-cultural nongeneralizations. *New Directions for Program Evaluation, 1985*(25), 93-96.

Peak, G. L., Peters, L., & Fishman, S. (2007). *Tracking Transformation: Evaluating the American Evaluation Association's Building Diversity Initiative (Report of Phase I Evaluation Findings)*. Two Gems Consulting. Retrieved from American Evaluation Association https://www.eval.org/p/cm/ld/fid=131.

Pearce, L. D. (2015). Thinking outside the Q boxes: Further motivating a mixed research perspective. In S. N. Hesse-Biber & B. Johnson (Eds.), *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (pp. 42-56). Oxford, UK and New York, NY: Oxford University Press.

Personal Responsibility and Work Opportunity Reconciliation Act, H.R. 2260; Pub.L. 104-193; 110 Stat. 2105. 104th Cong. (1996).

Pittz, W., & Sen, R. (2004). *Short Changed: Foundation Giving and Communities of Color*. Oakland, CA: Applied Research Center.

Pon, G. (2009). Cultural competency as the new racism: An anthology of forgetting. *Journal of Progressive Human Services, 20*(1), 59-71.

powell, j. a. (2010). Systems thinking, evaluation and racial justice. In Critical Issues Forum (Vol. 3, pp. 9-12). Philanthropic Initiative for Racial Equity: Marking Progress: Movement Toward Racial Justice. Retrieved from http://www.racialequity.org/criticalissues.html - critical3.

powell, j. a. (2013). Deepening our understanding of structural marginalization. *Poverty and Race, 22*(5), 3-13.

powell, j. a., Heller, C. C., & Bundalli, F. (2011). *Systems Thinking and Race: Workshop summary*. Los Angeles, CA: The California Endowment.

Prashad, V. (2000). *The Karma of Brown Folk*. Minneapolis, MN and London, UK: University of Minnesota Press.

Prest, M. J. (2014). Salaries on the Rise at U.S. Foundations. *News and Analysis.* Retrieved from the Chronicle of Philanthropy: https://www.philanthropy.com/article/Salaries-on-the-Rise-at-U.S./153049.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York, NY: Simon and Schuster.

Rabaka, R. (2007). The Souls of White Folk: W.E.B. DuBois's critique of white supremacy and contributions to critical white studies. *Journal of African American Studies* (11), 1-15.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.

Reichertz, J. (2007). Abduction: The logic of discovery of grounded theory. In A. Bryant & K. Charmaz (Eds.), *The Sage Handbook of Grounded Theory* (pp. 214-228). London, UK: Sage Publications.

Reskin, B. F. (2003). Including mechanisms in our models of ascriptive inequality. *American Sociological Review, 68*(1), 1-21.

Rich, C. G. (2016). Reclaiming the Welfare Queen: Feminist and critical race theory alternatives to existing anti-poverty discourse. *Southern California Interdisciplinary Law Journal, 25*(2), 257-288.

Richardson, V. (1990). At-risk programs: Evaluation and critical inquiry. *New Directions for Evaluation, 1990*(45): 61-75.

Roberts, D. J., & Mahtani, M. (2010). Neoliberalizing race, racing neoliberalism: Placing 'race' in neoliberal discourses. *Antipode, 42*(2), 248-257.

Roberts, R. E., & Attkisson, C. C. (1983). Assessing client satisfaction among Hispanics. *Evaluation and Program Planning, 6*(3-4), 401-413.

Robertson, P., Jorgenson, M., & Garrow, C. (2004). Indigenizing evaluation research: How Lakota methodologies are helping "raise the tipi" in the Oglala Sioux Nation. *The American Indian Quarterly, 28*(3-4), 499-526.

Robinson, E. (2019, July 18). Trump has become the voice of insecure white Americans. *The Washington Post*. Retrieved from Opinion https://www.washingtonpost.com/opinions/george-wallace-would-have-been-proud-of-trumps-latest-rally/2019/07/18/cc76dc5e-a997-11e9-9214-246e594de5d5_story.html.

Rodriguez, D. (2007). The political logic of the nonprofit industrial complex. In INCITE! (Ed.), *The Revolution Will Not be Funded: Beyond the Non-profit Industrial Complex* (pp. 21-40). Cambridge, Massachusetts: South End Press.

Rog, D. J. (2012). When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation, 2012*(135), 25-40.

Said, E. W. (1978). *Orientalism* (1st ed.). New York, NY: Pantheon Books.

Salamon, L. M. (2002). The Resilient Sector: The State of Nonprofit America. In L. M. Salamon (Ed.), *The State of Nonprofit America* (pp. 3-61). Washington, DC: Brookings Institution Press.

Salamon, L. M., & Sokolowski, S. W. (2006). *Employment in America's Charities: A Profile*. Baltimore: Johns Hopkins Center for Civil Society Studies.

Samuels, M., & Ryan, K. (2011). Grounding evaluations in culture. *American Journal of Evaluation, 32*(2), 183-198.

Sanders, J. R., JCSEE, & AASA. (1994). *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Schlueter, D. F. (2011). New evaluators addressing health disparities through community-based evaluation. *New Directions for Evaluation, 2011*(131), 59-64.

Schwandt, T. A. (2005). The centrality of practice to evaluation. *American Journal of Evaluation, 26*(1), 95-105.

Seigart, D. & Brisolara, S. (2002). Editors' Notes. *New Directions for Evaluation*, 2002(96), 1-2.

SenGupta, S., Hopson, R., & Thompson-Robinson, M. (2004). Cultural competence in evaluation: An overview. *New Directions for Evaluation, 2004*(102), 5-19.

Shalhoub, Z. K., & Al Qasimi, J. (2005). A soft system analysis of nonprofit organizations and humanitarian services. *Systemic Practice and Action Research, 18*(5), 457-473.

Shapiro, T. M. (2004). *The Hidden Cost of Being African American: How Wealth Perpetuates Inequality*. New York, NY: Oxford University Press.

Shapiro, T. M., Meschede, T., & Osoro, S. (2013). *The Roots of the Widening Racial Wealth Gap: Explaining the Black-White Economic Divide*. Waltham, MA: Brandeis University Institute on Assets & Social Policy.

Shiva, V. (2007). Biodiversity, intellectual property rights, and globalization. In B. d. S. Santos (Ed.), *Another Knowledge is Possible: Beyond Northern Epistemologies* (Vol. 3, pp. 272-287). London, UK: Verso.

Shohat, E., & Stam, R. (1994). *Unthinking Eurocentrism: Multiculturalism and the Media*. London, UK and New York, NY: Routledge.

Sielbeck-Bowen, K. A., Brisolara, S., Seigart, D., Tischler, C., & Whitmore, E. (2002a). Beginning the conversation. *New Directions for Evaluation, 2002*(96), 109-114.

Sielbeck-Bowen, K. A., Brisolara, S., Seigart, D., Tischler, C., & Whitmore, E. (2002b). Exploring feminist evaluation: The ground from which we rise. *New Directions for Evaluation, 2002*(96), 3-8.

Sirotnik, K. A. (Ed.) (1990). *Evaluation and Social Justice: Issues in Public Education* (Vol. 45). Hoboken, NJ: Wiley Periodicals.

Šlaus, I. & Jacobs, G. (2011). Human capital and sustainability. *Sustainability*, 3, 97-154

Smith, A. (2016). Heteropatriarchy and the three pillars of white supremacy: Rethinking women of color organizing. In INCITE! (Ed.), *Color of Violence: The INCITE! Anthology* (2nd ed., pp. 66-73). Durham, NC: Duke University Press.

Smith, A. L. (2016). #BlackWomenMatter: Neo-capital punishment ideology in the wake of state violence. *The Journal of Negro Education, 85*(3), 261-273.

Smith, L. T. (2012). *Decolonizing Methodologies—Research and Indigenous Peoples* (2nd ed.). London & New York: Zed Books.

Smith, N. L. (1985). Foundation support of evaluation. *Evaluation Review, 9*(2), 215-239.

Smith, N. L. (2002). International students' reflections on the cultural embeddedness of evaluation theory. *American Journal of Evaluation, 23*(4), 481-492.

Soss, J., Fording, R. C., & Schram, S. (2011). *Disciplining the Poor: Neoliberal Paternalism and the Persistent Power of Race*. Chicago: University of Chicago Press.

Spilerman, S. (2000). Wealth and stratification processes. *Annual Review of Sociology, 26*, 497-524.

Stake, R. E. (1974). *Program evaluation, particularly responsive evaluation*. [Occasional Paper No. 5]. Kalamazoo, MI: Western Michigan University Evaluation Center

Stanfield, J. H. (1985). *Philanthropy and Jim Crow in American Social Science* (Vol. 82). Westport, CT: Greenwood Press.

Stanfield, J. H. (1993). African American traditions of civic responsibility. *Nonprofit and Voluntary Sector Quarterly, 22*(2), 137-153.

Stanfield, J. H. (1999). Slipping through the front door: Relevant social scientific evaluation in the people of color century. *American Journal of Evaluation, 20*(3), 415-431.

Stanhope, V., Solomon, P., Pernell-Arnold, A., Sands, R. G., & Bourjolly, J. N. (2005). Evaluating cultural competence among behavioral health professionals. *Psychiatric Rehabilitation Journal, 28*, 225-233.

Stave, K., & Hopper, M. (2007). *What constitutes systems thinking? A proposed taxonomy*. Paper presented at the International Conference of the System Dynamics Society, Boston, MA.

Stec, J. (2007). Deconcentration of poverty as an example of Derrick Bell's interest-convergence dilemma: White neutrality interests, prisons, and changing inner cities. *Northwestern Journal of Law and Social Policy, 2*(1), 30-62.

Stephens, A. (2012). Feminist Systems Theory: Learning by praxis. *Systemic Practice and Action Research, 25*, 1-14.

Stephens, A., Jacobson, C., & King, C. (2010). Towards a Feminist-Systems Theory. *Systemic Practice and Action Research, 23*, 371-386.

Stufflebeam, D. L. (1968). *Evaluation as Enlightenment for Decision Making*. Columbus, OH: Ohio State University Evaluation Center.

Sturges, K. M. (2014). External evaluation as contract work: The production of evaluator identity. *American Journal of Evaluation, 35*(3), 346-363.

Sturges, K. M. (2015). Complicity revisited: Balancing stakeholder input and roles in evaluation use. *American Journal of Evaluation, 36*(4), 461-469.

Suchman, E. A. (1967). *Evaluative Research: Principles and Practice in Public Service & Social Action Programs*. New York: Russell Sage Foundation

Suedfeld, P., Paterson, H., Soriano, E., & Zuvic, S. (2002). Lethal stereotypes: Hair and eye color as survival characteristics during the Holocaust. *Journal of Applied Social Psychology, 32*(11), 2368-2376.

Super-ordinate. (2016). OED Online. [Def. A. 2a, B. 2]. Retrieved from http://www.oed.com.

Symonette, H. (2004). Walking pathways toward becoming a culturally competent evaluator: Boundaries, borderlands, and border crossings. *New Directions for Evaluation, 2004*(102), 95-109.

Symonette, H. (2015). Culturally responsive evaluation as a resource for helpful-help. In S. Hood, R. Hopson, & H. Frierson, (Eds.), *Continuing the Journey to Reposition Culture and Cultural Context in Evaluation Theory and Practic*e (pp. 109-130). Charlotte, NC: Information Age Publishing, Inc.

Symonette, H., Mertens, D. M., & Hopson, R. (2014). The development of a diversity initiative: Framework for the Graduate Education Diversity Internship (GEDI) program. *New Directions for Evaluation, 2014*(143), 9-22.

Szostak, R. (2015). Interdisciplinary and transdisciplinary multimethod and mixed methods research. In S. N. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (pp. 128-143). Oxford, UK and New York, NY: Oxford University Press.

Taylor, P., Kochhar, R., Fry, R., Velasco, G., & Motel, S. (2011). *Twenty-to-One Wealth Gaps Rise to Record Highs Between Whites, Blacks and Hispanics*. Washington, DC: Pew Research Center.

Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *New Directions in Mixed Methods Research, 13*(1), 12-28.

Tervalon, M. & Murray-Garcia, J. (1998). Cultural humility versus cultural competence: A critical distinction in defining physician training outcomes in multicultural education. *Journal of Health Care for the Poor and Underserved, 9,* 117-125.

The Institute for Democratic Renewal & Project Change Anti-Racism Initiative. (1999). *A Community Builder's Tool Kit—15 Tools for Creating Healthy, Productive Interracial/Multicultural Communities: A Primer for Revitalizing Democracy from the Ground Up*. Retrieved from Racial Equity Resource Guide https://www.racialequitytools.org/resourcefiles/idr.pdf.

The Oral History Project Team. (2004). The Oral History of Evaluation Part II: The Professional Development of Lois-ellin Datta. *American Journal of Evaluation, 25*(2), 243-253.

The Oral History Project Team. (2005). The Oral History of Evaluation, Part 3: The Professional Evolution of Michael Scriven. *American Journal of Evaluation, 26*(3), 378-388.

The Oral History Team. (2015). The oral history of evaluation: The professional development of Ernest House. *American Journal of Evaluation, 36*(2), 270-282.

Thomas, V. G. (2004). Building a contextually responsive evaluation framework: Lessons from working with urban school interventions. *New Directions for Evaluation, 2004*(101), 3-23.

Thomas, V. G., & Madison, A.-M. (2010). Integration of social justice into the teaching of evaluation. *American Journal of Evaluation, 31*(4), 570-583.

Thomas, V. G., & Parsons, B. A. (2017). Culturally responsive evaluation meets systems-oriented evaluation. *American Journal of Evaluation, 38*(1), 7-28.

Thompson, D. (2001). *Radical Feminism Today*. London: Sage.

Thompson-Robinson, M., Hopson, R. K., & SenGupta, S. (2004). Editors' Notes. *New Directions for Evaluation, 2004*(102), 1-4.

Tibbs, D. (2012). From Black Power to hip hop: Discussing race, policing, and the Fourth Amendment through the "War on" paradigm. *The Journal of Gender, Race, and Justice, 15*(1), 47-79.

Tilly, C. (1998). *Durable Inequality*. Berkeley: University of California Press.

Todd, S. (2011). "That power and privilege thing": Securing whiteness in community work. *Journal of Progressive Human Services, 22*(2), 117-134.

Toren, N., & Kraus, V. (1987). The effects of minority size on women's position in academia. *Social Forces, 65*, 1090-1100.

Tranquist, J. (2015). The oral history of evaluation: The professional development of Evert Vedung. *American Journal of Evaluation, 36*(4), 570-583.

Tuck, E. & Yang, K.W. (2012). Decolonization is not a metaphor. *Decolonization: Indigeneity, Education & Society, 1*(1), 1-40.

Valentinov, V. (2012). Toward a critical systems perspective on the nonprofit sector. *Systemic Practice and Action Research, 25*(4), 355-364.

van Dijk, T. A. (1993). Analyzing racism through discourse analysis: Some methodological reflections. In J. H. Stanfield & R. M. Dennis (Eds.), *Race and ethnicity in Research Methods* (pp. 92-134). Newbury Park, CA: Sage Publications.

van Dijk, T. A. (2015). Critical discourse analysis. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (2nd edition, Vol. 2, pp. 466-486). West Sussex, UK: John Wiley & Sons, Inc.

Vijaya, R. M., Eshleman, A., & Halley, J. (2015). The culture of class and its economic impact. *The Review of Black Political Economy, 42*(1), 7-18.

Vo, A. T. (2013). Visualizing context through theory deconstruction: A content analysis of three bodies of evaluation theory literature. *Evaluation and Program Planning, 38*(44-52).

Vo, A. T., & Christie, C. A. (2015). Advancing research on evaluation through the study of context. *New Directions for Evaluation, 2015*(148), 43-55.

Volpp, L. (1996). Talking "culture": Gender, race, nation, and the politics of multiculturalism. *Columbia Law Review, 96*(6), 1573-1617.

Ward Hood, D., & Cassaro, D. A. (2002). Feminist evaluation and the inclusion of difference. *New Directions for Evaluation, 2002*(96), 27-40.

Weisinger, J. Y., Borges-Méndez, R., & Milofsky, C. (2015). Diversity in the nonprofit and voluntary sector. *Nonprofit and Voluntary Sector Quarterly, 45*(IS), 3S-27S.

Weiss, C. H. (1993). Where politics and evaluation research meet. *Evaluation Practice, 14*(1), 93-106.

Western States Center. (2011, February 15). *The origin of the phrase "Women of Color."* [Video and transcript]. Retrieved from Youtube https://www.youtube.com/watch?v=82vl34mi4Iw&t=1s.

White, K. T. (2007). Initial conditions at Emancipation: The long-run effect on Black-White wealth and earnings inequality. *Journal of Economic Dynamics & Control, 31*, 3370-3395.

Whitmeyer, J. M. (1994). Why actor models are integral to structural analysis. *Sociological Theory, 12*(2), 153-165.

Wikler, D. (1999). Can we learn from eugenics? *Journal of Medical Ethics, 25*, 183-194.

Williams, B., & Hummelbrunner, R. (2011). *Systems Concepts in Action: A Practitioner's Toolkit*. Stanford, CA: Stanford Business Books.

Williams, C. (2008). Surveying the battlefield: Reflections on the reproductive dynamics of racism. *International Journal of Criminology and Sociological Theory, 1*(1), 62-79.

Wodak, R. (2001a). *The discourse-historical approach*. London, UK and Thousand Oaks, CA: Sage.

Wodak, R. (2001b). What CDA is about: A summary of its history, important concepts and its developments. In R. Wodak & M. Meyer (Eds.), *Methods of Critical Discourse Analysis* (pp. 1-12). London, UK and Thousand Oaks, CA: Sage.

Woelders, S., & Abma, T. (2015). A different light on normalization: Critical theory and responsive evaluation studying social justice in participation practices. *New Directions for Evaluation, 2015*(146), 9-18.

Wolfe, R. R. (2010). *Working (in) the gap: A critical examination of the race/culture divide in human services.* (Doctoral Thesis), University of Alberta, Edmonton, Alberta.

Wu, F. H. (2002). The Model Minority: Asian American "success" as a race relations failure. In *Yellow: Race in America Beyond Black and White* (pp. 39-77). New York, NY: Basic Books.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (3rd ed.). Thousand Oaks, CA: Sage.

Young, I. M. (2011). The five faces of oppression. In I. M. Young & D. S. Allen (Eds.), *Justice and the Politics of Difference* (Paperback reissue ed., pp. 39-65). Princeton, NJ: Princeton University Press.

Zaw, K., Bhattacharya, J., Price, A., Hamilton, D., & Darity, W. A. (2017). *Women, Race & Wealth*. Research Brief Series, Volume 1. Duke University Samuel DuBois Cook Center on Social Equity and Insight Center for Community Economic Development.

APPENDICES

Appendix A: Potential Interviewees

Name/Affiliations	Connection to discursive events and institutional mechanisms close in proximity or conceptually to shifts in discourse
1. Alden Loury, Director of Research and Evaluation, Metropolitan Planning Commission	• Dialogues on Race & Class
2. Anna-Marie Madison	• AEA/MIE • Scholarship/Literature
3. Astrid Hendricks Smith	• Cultural Competence in Evaluation Panel of Experts
4. Bertha Holliday, Ph.D., American Psychological Association	• Building Diversity Advisory Committee
5. Brandi Gilbert, Urban Institute	• GEDI participant • GEDI Program Director • Scholarship/Literature
6. Brandon Ledward	• Cultural Competence in Evaluation assistance
7. Carolyn Huie Hofstetter, Ph.D., Graduate School of Education, University of California at Berkley	• Building Diversity Advisory Committee
8. Celina Moreno, Legislative Attorney, Mexican American Legal Defense & Educational Fund (MALDEF)	• Dialogues on Race & Class
9. Charles Thomas, Ph.D., George Mason University	• Building Diversity Advisory Committee • Cultural Reading 2003
10. Cindy Crusto, Yale University	• Cultural Competence in Evaluation Task Force (Chaired 2008) • Cultural Reading 2004
11. Claude Bennett, U.S. Department of Agriculture	• Cultural Reading 2003

Name/Affiliations	Connection to discursive events and institutional mechanisms close in proximity or conceptually to shifts in discourse
12. Cornel Pewewardy, Ph.D., Teaching and Leadership Department, University of Kansas at Lawrence	• Building Diversity Advisory Committee
13. Corrine Kirchner	• Cultural Competence in Evaluation Panel of Experts
14. Craig Russon, Kellogg Foundation	• Cultural Reading 2004
15. David Chavis, Ph.D., Association for the Study and Development of Community/Community Science	• Building Diversity Advisory Committee
16. Davis Ja, Davis Y. Ja & Associates	• Building Diversity Advisory Committee
17. Debra Rog, Vice President, Westat	• Dialogues on Race & Class
18. Denice Cassaro, Cornell University	• Cultural Reading 2003 • Cultural Reading 2004 • Scholarship/Literature
19. Denice Ward Hood	• Scholarship/Literature
20. Diana Biro	• Cultural Competence in Evaluation assistance
21. Donna Mertens, Ph.D., Gallaudet University	• 1998 President (conference theme on social transformation) • Building Diversity Advisory Committee Co-Chair • GEDI Program Director • Cultural Reading 2003 • Cultural Reading 2004
22. Edith Thomas, Ph.D., U.S. Department of Agriculture, Guiding Principles Subcommittee	• Building Diversity Advisory Committee • Cultural Reading 2003
23. Elizabeth Whitmore, Carleton University	• Cultural Reading 2003 • Cultural Competence in Evaluation Task Force

Name/Affiliations	Connection to discursive events and institutional mechanisms close in proximity or conceptually to shifts in discourse
24. Elmima Johnson, Ph.D., Division of Research, Evaluation & Communication, National Science Foundation	• Building Diversity Advisory Committee • Cultural Reading 2003 • Cultural Reading 2004
25. Ernest House	• Scholarship/Literature
26. Frank Besag	• Scholarship/Literature
27. Gary Harper	• Cultural Competence in Evaluation Panel of Experts
28. Geri Lynn Peak	• Phase I evaluation of the BDI
29. Guadalupe Pacheco, President/CEO, Pacheco Consulting Group, LLC	• Dialogues on Race & Class • Cultural Competence in Evaluation Panel of Experts
30. Hazel Symonette, Ph.D., University of Wisconsin, Madison	• Building Diversity Advisory Committee Co-Chair • Cultural Competence in Evaluation Panel of Experts
31. Henry Frierson, Jr., Ph.D., Research Education Support Program, University of North Carolina at Chapel Hill	• Building Diversity Advisory Committee
32. James Davis, Ph.D., College of Education, Temple University	• Building Diversity Advisory Committee • Scholarship/Literature
33. Jennifer Greene	• Cultural Competence in Evaluation assistance
34. Jenny Jones	• Cultural Competence in Evaluation Task Force
35. Joan LaFrance, Owner of Mekinak Consulting and enrolled Citizen of the Turtle Mountain Band of Chippewa, Belcourt North Dakota	• Cultural Reading 2004 • Chaired Diversity Committee 2005 • Dialogues on Race & Class • AEA/MIE/IPE • Scholarship/Literature
36. Joanna Birckmayer, Ph.D., Research Associate, Association for the Study and Development of Community/Community Science	• Building Diversity Advisory Committee

Name/Affiliations	Connection to discursive events and institutional mechanisms close in proximity or conceptually to shifts in discourse
37. John Stanfield	• Scholarship/Literature
38. Jonathan Holmes, Race and Equity Program Coordinator, Chicago Urban League	• Dialogues on Race & Class
39. Juan Martinez, M.A., Research and Policy Division, Arizona Department of Education	• Building Diversity Advisory Committee
40. Karen Kirkhart. Ph.D., School of Social Work. Syracuse University	• 1994 President (conference theme on social justice) • Building Diversity Advisory Committee • Cultural Reading 2003 (Chaired Work Group/Task Force 2003) • Cultural Reading 2004 • Cultural Competence in Evaluation Task Force • Scholarship/Literature
41. Katherine Tibbetts	• Cultural Competence in Evaluation Task Force
42. Katrina Bledsoe	• Cultural Competence in Evaluation Task Force
43. Kien Lee, Principal Associate/Vice President, Community Science	• Dialogues on Race & Class
44. Kimberly Hall, M.A., Department of Psychology, Division of Community and Prevention Research, University of Illinois at Chicago	• Building Diversity Advisory Committee
45. Kristen Harper, Senior Policy Specialist, Child Trends	• Dialogues on Race & Class
46. Laura Rendon, Ph.D., Department of Educational Psychology and Administration, California State University at Long Beach	• Building Diversity Advisory Committee
47. Laurine Thomas, Ph.D., Senior Research & Evaluation Officer, Academy for Educational Development, Project Oversight Committee	• Building Diversity Advisory Committee
48. Lester Horvath, Evaluation Consultant, Connecticut	• Cultural Reading 2003

Name/Affiliations	Connection to discursive events and institutional mechanisms close in proximity or conceptually to shifts in discourse
49. Lisa Aponte-Soto	• AEA/MIE • GEDI • Scholarship/Literature
50. Lisa Gray	• Cultural Competence in Evaluation assistance
51. Liz Lutz, Executive Director of the Health Collaborative	• Dialogues on Race & Class
52. Lois-ellin Datta	• AEA
53. Lorraine Forte, Executive Editor, The Chicago Reporter	• Dialogues on Race & Class
54. Marian Secundy, Ph.D., Tuskegee University Center for Bioethics	• Building Diversity Advisory Committee
55. Mary Merryfield	• Scholarship/Literature
56. Matthew Corry	• Cultural Competence in Evaluation assistance
57. Melvin Hall, Northern Arizona University	• Cultural Reading 2003 • Cultural Reading 2004 (Chaired Diversity Committee 2004) • Cultural Competence in Evaluation assistance • Cultural Competence in Evaluation Task Force (Chaired 2005-2008)
58. Mike Lowe, San-Antonio leader of Black Lives Matter and co-founder of SATX4, a grassroots organization founded in San Antonio, TX.	• Dialogues on Race & Class
59. Nick Hart, President, Washington Evaluators	• Dialogues on Race & Class
60. Nick Smith	• Cultural Competence in Evaluation assistance
61. Pauline Brooks	• Cultural Competence in Evaluation Panel of Experts
62. Prisca Collins	• GEDI Program Director
63. RaeDeen Karasuda	• Cultural Competence in Evaluation assistance
64. Ricardo Millett	• Building Diversity Advisory Committee

Name/Affiliations	Connection to discursive events and institutional mechanisms close in proximity or conceptually to shifts in discourse
65. Robert Covert	• AEA/MIE
66. Roberta (Robbi) Ferron	• Cultural Competence in Evaluation Panel of Experts
67. Robin Lin Miller, Professor of Ecological-Community Psychology at Michigan State University	• Dialogues on Race & Class
68. Rodney Hopson, Ph.D., School of Education, Department of Foundations and Leadership, Duquesne University George Mason University CREA	• 2012 President (conference invitation to Linda Tuhiwai Smith) • Building Diversity Advisory Committee • GEDI Program Director • AEA/MIE • Cultural Competence in Evaluation assistance • Scholarship/Literature
69. Rogelo Sanz, Dean of the College of Public Policy and Mark G. Yudof Endowed Chair at the University of Texas at San Antonio	• Dialogues on Race & Class
70. Ross Connor	• 1989 President (conference theme on international & cross-cultural evaluation) • Cultural Competence in Evaluation Panel of Experts • Scholarship/Literature
71. Satish Verma, Louisiana State University	• Cultural Reading 2003 (Chaired Diversity Committee 2003)
72. Sharon Rallis, University of Connecticut	• Cultural Reading 2003
73. Stafford Hood, Ph.D., Division of Psychology in Education, Arizona State University Founding Director, Center for Culturally Responsive Evaluation & Assessment, University of Illinois at Urbana-Champaign	• Building Diversity Advisory Committee • Cultural Reading 2003 • Dialogues on Race & Class • Scholarship/Literature

Name/Affiliations	Connection to discursive events and institutional mechanisms close in proximity or conceptually to shifts in discourse
74. Stan Capela, Heart Share Human Services	• Building Diversity Advisory Committee
75. Theresa Singleton, Ph.D., Project Director, Association for the Study and Development of Community	• Building Diversity Advisory Committee
76. Veronica Thomas, Howard University Professor	• Dialogues on Race & Class
77. Virginia Richardson	• Scholarship/Literature

Appendix B: Invited Interviewees

Name	Scholarship	AEA BIPOC TIG Sponsorship/ Leadership	Building Diversity Advisory Committee	Cultural Reading of Standards	GEDI	Cultural Competence Statement	Dialogues on Race & Class
Anna-Marie Madison, Associate Professor	X	X					
Brandi Gilbert, Urban Institute	X				X		
Craig Russon, formerly at Kellogg Foundation				X			
David Chavis, Community Science			X				
Denice Cassaro, Cornell University	X			X			
Donna Mertens, retired from Gallaudet University	X		X	X	X		
Elmima Johnson, retired from Division of Research, Evaluation & Communication, National Science Foundation			X	X			
Geri Lynn Peak, Two Gems Consulting			X				
Hazel Symonette, retired from University of Wisconsin, Madison	X		X			X	
James Davis, College of Education, Temple University	X		X				

Name	Scholarship	AEA BIPOC TIG Sponsorship/ Leadership	Building Diversity Advisory Committee	Cultural Reading of Standards	GEDI	Cultural Competence Statement	Dialogues on Race & Class
Joan LaFrance, Owner of Mekinak Consulting and enrolled Citizen of the Turtle Mountain Band of Chippewa, Belcourt North Dakota	X	X		X			X
Johanna Birckmayer, formerly Research Associate, Community Science			X				
Karen Kirkhart, School of Social Work. Syracuse University	X		X	X		X	
Kien Lee, Principal Associate/ Vice President, Community Science	X						X
Lois-ellin Datta, Datta Analysis	X	X					
Melvin Hall, Northern Arizona University	X			X		X	X
Ricardo Millett, formerly at Kellogg Foundation			X				
Robert Covert; retired from UVA	X	X					
Satish Verma, Louisiana State University				X			

Name	Scholarship	AEA BIPOC TIG Sponsorship/ Leadership	Building Diversity Advisory Committee	Cultural Reading of Standards	GEDI	Cultural Competence Statement	Dialogues on Race & Class
Stafford Hood, Founding Director, Center for Culturally Responsive Evaluation & Assessment, University of Illinois at Urbana-Champaign	X		X	X			X
Theresa Singleton, formerly Project Director, Community Science			X				
Veronica Thomas, Howard University Professor	X						X

Appendix C: Interviewee Invitation

Dear Xx. Xxx:

I am contacting you because journals affiliated with the American Evaluation Association and documents available in AEA's online library suggest that your leadership, scholarship, and participation on specific AEA committees and task forces have contributed to the professional association and the field's understanding and language regarding racialized difference. I am writing to invite you to take part in my PhD dissertation research, which examines the evaluation field's construction of racialized difference and which I expect to advance understanding of the history, context, and interests that have influenced that construction. Jean King is my adviser and Rodney Hopson has been an invaluable member of my committee.

Taking part would involve participating in an interview whose questions would ask you to share your recollection of your experience in leadership roles, committees, and task forces as they pertain to the understanding and language regarding racialized difference within the evaluation field and its professional association. Questions will also ask you to interact with, expand upon, and help contextualize preliminary results of my analysis of evaluation literature and professional association documents in relation to AEA's development of remedies for the differences described in the literature between program participants and program evaluators.

The interview is expected to require no more than 90 minutes and ideally take place by Google Hangout in March or April. While I prefer Google Hangout or similar platform because it would facilitate the sharing of any relevant documents, including a visualization of the preliminary results mentioned above, we can conduct the interview by phone if you prefer. Instructions and links to any materials would be provided before the interview takes place. If you consent to participate, interviews will be audio-recorded and then transcribed. Information that could potentially identify participants will not be attached to the audio-recording or the transcript of their interviews in any way, nor will it be included with the reporting of results—particularly direct quotations.

Please respond to this email by March 6 to let me know if you would consider participating in this research or if you would prefer not to be contacted again. If you would consider participating, I will send you a consent form that includes additional information by March 13. Upon receipt of the consent form, please let me know by March 20 whether you would like to participate. If you choose to, I will contact you by March 27 to schedule an interview and will send the preliminary results and interview questions immediately thereafter.

Appendix D: Interview Outline

I. Grounding in the process

This interview is expected to take 60-90 minutes of your time, depending on the nature and extent of your involvement in leadership roles, committees, and task forces as they pertain to the understanding and language regarding racialized difference within the evaluation field and its professional association. Initial questions will ask you to interact with, expand upon, and help contextualize preliminary results of the analysis of evaluation literature and professional association documents in relation to AEA's development of remedies for the difference described in the literature between program participants and program evaluators. Depending on your role, you may also be asked to share your recollection of your personal experience.

II. Locating ourselves in relation to the field

A. How would you describe the similarities, differences, and relationships among program participants and program evaluators, including yourself?

B. Can you briefly describe the professional, social, political, cultural, or other groups that you belong to or identify with?

III. Orientation to and interaction with visualization of the preliminary results

As part of this research, I have conducted a diachronic analysis of the evaluation literature, which shows how the language used to refer to racialized difference in evaluation has changed over time and in relation to some of the pivotal moments in the evaluation field and professional association's understanding and programming around racialized difference.

A. Is the visualization clear?

How does it look to you?

Any initial responses to it?

B. Any surprises?

Is this the direction you expected the language and understanding to go in?

C. How do you interpret the trajectory?

D. What factors do you think contributed to it?
(Internalized, interpersonal, institutional, or structural?)
1. What role did you play in it (with participants for whom this is relevant)?
2. Who/What (else) influenced the trajectory?

E. Focusing on the discourse shift (with participants for whom this is relevant):
1. Can you describe your recollection of the sequence of events immediately surrounding the shift (between about 1999 and 2004)?
2. How did it affect or otherwise relate to your scholarship and practice?

IV. Considering the boundaries and dimensions of justice

A. Amid numerous references to "social justice," my search yielded only six articles that include the phrase "economic justice" and no instance of the phrase "racial justice" in the evaluation literature.
1. Do you have any thoughts about that?
Why do you think that is?
2. How do racial, economic, and gender justice converge with social justice and how do they diverge from it?

B. What would evaluation look like in a world that is just intersectionally—at the intersections of oppression defined by race, class status, gender, sexuality, ability status, etc.?
What would intersectional justice look like in evaluation?
- What would need to change for us to see either of those visions?

V. Final thoughts & closing