Invidious Comparisons: Imagining a Psychology of Race and Gender Beyond Differences

Elizabeth R. Cole
Center for Afroamerican and African Studies and Program in Women’s Studies, University of Michigan

Abigail J. Stewart
Department of Psychology and Program in Women’s Studies and Institute for Research on Women and Gender, University of Michigan

Within psychology there is a longstanding debate concerning whether large, fundamental differences between races and genders exist. Much of this research involves comparisons that are invidious (offensively discriminating) and supports a political ideology in which members of different groups are held to be fundamentally different, alien, and therefore destined to different spheres. In this paper, specific factors are identified that make research on group differences more likely to produce distorted and partial findings. In addition, studies are cited whose methodological strategies offer insight into the processes that create and maintain group differences. Such research may illuminate not only the differences between groups, but also the very meaning of group categories. The process of scholarly peer review should become sensitive to the features that make comparisons invidious, so as to incorporate them into the criteria used to evaluate research.

KEY WORDS: race comparisons, gender comparisons, research methods.

In the autumn of 1999, members of several divisions of the American Psychological Association (APA) were mailed a plain white envelope, without a return address or explanatory cover letter, containing a special abridged version of J. Philippe Rushton’s (1995a) Race, Evolution, and Behavior. Some psychologists who received the mailing later expressed their indignation on electronic listservs sponsored by the divisions.¹ Their comments suggested that Rushton’s work, which

¹ In a letter dated 23 December 1999, Jerry McGlaughlin, APA’s director of management information services, informed Transaction Publishers that they had violated APA’s policy for the purchase and
argues for an innate and essential hierarchy of the human races, was so far outside of acceptable scholarship that even to receive such a work in one’s mailbox, free of charge, is an outrage. Although we share these psychologists’ critical evaluation of Rushton’s work, in this paper we would like to consider seriously whether his methods—that is, seeking support for hypotheses involving deep and essential differences between racial and gender groups—are indeed beyond the pale. On the contrary, we contend that the study of deep difference thrives, uncensored, in many areas of psychology, and that many of these comparisons are invidious (or “offensively discriminating,” as defined in the Oxford English Dictionary).

Even a cursory examination of major psychology journals in the last 10 years reveals that psychologists continue to exchange blows over whether large and fundamental differences between races and genders exist (see, e.g., Duckitt, 1992; Eagly, 1995; Hyde & Plant, 1995; Zuckerman, 1990). This debate has continued since virtually the start of the discipline, in relatively unchanged form. For example, these two claims are virtually indistinguishable, but are separated by 71 years: “It is fairly well-established . . . that the intelligence of the White race is of a more versatile and complex order than that of the Black race” (Allport, 1924, p. 386); “With data from around the world it has been repeatedly found that the races differ not only in cognitive ability . . . but also in brain size” (Rushton, 1995b, p. 41).

Many psychologists have attempted to transcend this impasse by carefully delineating guidelines for improving research in these areas (e.g., Graham, 1992; McHugh, Koeske, & Frieze, 1986). Nevertheless, there’s little evidence that any of these eloquent entreaties have altered the contours of the debate. The lack of progress in the terms of the discussion is especially surprising in the context of the dramatic legal and social changes in racial and gender equality over the course of the last century.

There are at least two strong traditions among those who argue for the strength and centrality of differences in the study of both race and gender. The first and perhaps most obvious is exemplified by those who seek to document the existence of difference as evidence that women or minorities are innately and essentially different (e.g., Benbow, 1988; Benbow & Stanley, 1980) or inferior (e.g., Rushton, 1988). The second, and in our mind only somewhat less dangerous, is reflected in the impulse to identify and romanticize differences (e.g., Gilligan, 1982; Welsing, 1991). This type of “strategic essentialism” is one way that members of subordinate groups can stake a claim to the legitimacy of their experience, demanding recognition and accommodation for non-dominant perspectives. But however attractive this tactic may be, it is ultimately a two-edged sword. In a cultural/political context in which differences between groups are often presumed to imply that one group is superior (as in the case of race or gender) or “normal” (as in the case of sexual

use of mailing labels, as the publisher had obtained the labels by misrepresenting the material that would be sent under their cover. McGlaughlin stated that until the matter was resolved, Transaction Publishers would be prohibited from purchasing further labels from APA.
orientation or ability status), groups that emphasize their differences may unwittingly reinforce the rationale supporting the status quo that seeks to segregate and stigmatize them (the case for teaching “Ebonics” in schools is a case in point). A perhaps even more subtle and insidious trap, suggested by Bem (1998), results when subordinate groups practicing this type of strategic essentialism come to “oversolidify” these differences in group members’ own minds. In other words, when we argue for the existence of, for example, a “gay gene” or a woman’s way of making moral judgments, we buy into the “naturalness” of these social divisions, losing sight of the extent to which the meaning of these social categories is constructed in a particular social and historical context. When subordinate groups naturalize group distinctions, they lose an indispensable critical tool: the ability to see that the hierarchical value attached to difference is socially determined (see also Reicher & Hopkins, this issue).

Thus, although the essentialists and the “strategic essentialists” seemingly make completely different arguments motivated by different social and political values, both sets of arguments ultimately serve to support a political ideology and social worldview in which members of different groups are fundamentally different, alien, and therefore destined to different appropriate spheres. These ideas have found fertile ground not only in the popular imagination [e.g., Gray’s (1992) *Men Are From Mars, Women Are From Venus*], but also among academic psychologists [e.g., Shelley Taylor’s research on “tend and befriend” as the female analogue of the “fight or flight” stress response (Azar, 2000)].

At the same time, powerful arguments about the dangers of identifying, reifying, and reinforcing differences have been articulated repeatedly (e.g., Hare-Mustin & Marecek, 1994; Phinney, 1996). Many of these critiques take one of two forms: Some challenge the veridicality or accuracy of the “knowledge” produced by research focusing on differences (Yee, Fairchild, Weizmann, & Wyatt, 1993); others attend to the social harm that flows from uncritical belief in those findings (Betancourt & Lopez, 1993; Eccles & Jacobs, 1986; Lott, 1996). In fact, some psychologists have contended that we should never study race or gender differences (Baumeister, 1988).

We believe that despite the potential harm comparative research can do, there are good reasons to study the different experiences and circumstances associated with race and gender. In the rest of this paper, we explore the specific factors that make research on differences more likely to cause harm and produce distorted and partial findings, as well as the factors that make such research more likely to offer insight into the lived processes that create and maintain these differences.

**When Race and Gender Comparisons Are Invidious**

We argue that certain features distinguish comparisons that are made with either the explicit or unconscious purpose of showing one group in a negative light; such comparisons are, we think, invidious. Research resulting in invidious
comparisons is characterized by five features: dependence on an irrelevant norm; the assumption that differences may be ordered along a hierarchy of value; a focus on decontextualized outcomes; comparison of means without reference to distributions; and use of measures that are weak tests of the relevant ability. Some of these characteristics may seem to be simply characteristics of flawed research, and to some degree they are. However, even when that is most true, the fact that race and gender comparisons are accepted—even when made in ways that are generally viewed as “flawed”—highlights the fact that consensus on general research standards and norms is not sufficient to prevent invidious comparisons. We examine why this is true after discussing each of these characteristics in turn.

**Dependence on an irrelevant norm.** Critics of research on race and gender differences have routinely pointed out the frequent use of implicit standards of comparison of women to men and Blacks to Whites (Duckitt, 1992; Hare-Mustin & Marecek, 1994; Hyde, 1994; Sherif, 1979). These implicit standards are often irrelevant, and are invoked to suggest that women or African Americans are deficient or inferior. One of the earliest studies uncovering psychologists’ use of men as the implicit standard for mental health was the study of clinicians’ stereotypes conducted by Broverman, Vogel, Broverman, Clarkson, and Rosenkrantz (1972). They found that the stereotype or description of the mentally healthy man and the mentally healthy person were nearly identical, and that both were different from the stereotype of the mentally healthy woman. Similar themes are present in the controversy among clinicians about the legitimacy of diagnoses that appear to pathologize behavior consistent with women’s gender role socialization, such as self-defeating personality disorder (Caplan & Gans, 1991; Caplan, McCurdy-Myers, & Gans, 1992). Unfortunately, researchers are also vulnerable to the use of men as the implicit standard. For example, Gervasio and Crawford (1989) observed that there has been a great deal of research advocating the effectiveness of assertiveness training for women. In their cogent critique of this literature, they argued that despite a wealth of evidence that women are more competent than men in both verbal and nonverbal social communication, the emphasis on assertiveness implies that women are “deficient in communication competence and blamed for ineffective speech styles” (p. 10). This has occurred because masculinity is the implicit standard for mental health, and assertive behavior closely resembles the positive stereotype of masculine behavior. As a result of this male norm, women are often sent to assertiveness training, but analogous interventions are seldom prescribed for inappropriately aggressive men.

Rushton and Bogaert’s work on racial differences in sexuality (1987) represents an especially invidious example of such comparisons. Research in population biology suggests that across species, there is a continuum of r/K reproductive strategies; species adopting an r strategy produce many offspring but invest little care in each, whereas the K strategy entails few offspring and great parental investment. Rushton and his colleagues proposed that within the human species there are racial differences along this same continuum, such that Blacks are the
most r, Asians are the most K, and Whites fall somewhere between. In support of this claim, they marshaled evidence that Blacks reach sexual maturity earlier, and have greater interest in sex, shorter menstrual cycles, and larger penises, to conclude that they have evolved a predisposition to reproduce at a great rate; despicably, they argued that Blacks’ earlier average mortality indicates less parental investment over time. Elsewhere (Rushton & Bogaert, 1988) they made similar claims for non–college-educated Whites (although they found this group to be more K than Blacks). Although this line of research has been criticized by many scholars (for one example, see Lewontin, Rose, & Kamin, 1984), it nevertheless found publication in a major peer-reviewed psychology journal. The comparison is particularly invidious, as the claim for Blacks’ lesser investment in their children is made in the absence of any data concerning actual child-rearing practices. No mention is made of the sociological literature documenting the extended responsibility for child-rearing among adults in African American communities (Collins, 1990).

Ordering differences hierarchically. Where similarity between groups is found, this implicit hierarchy of value is also sometimes evident in the explanations that lay people and scientists offer for the findings. For example, when asked to make attributions about why athletes are successful, college students attribute Blacks’ achievement to physicality (innate athletic ability, such as extra muscle and speed) and Whites’ achievement to personality and intellect (hard work, leadership, and intelligence for the sport) (Johnson, Hallinan, & Westerfield, 1999; Stone, Perry, & Darley, 1997). Hoberman (1997) described in detail scientists’ attempts to identify biological and evolutionary differences between the races that might explain the success that Blacks have enjoyed in athletics. He argued that the search for biological difference as an explanation has two pernicious side effects. First, because of what he terms the “Darwinian law of compensation,” the assumption of an “inverse relationship between brain and brawn” (p. 189), these attempts to establish evidence for Blacks’ physical superiority are freighted with an assumption of Blacks’ intellectual inferiority. Thus, in the comparison of Blacks’ and Whites’ athletic ability, an advantage attributed to Blacks is paradoxically interpreted as evidence of their deficiency—evidence of how truly invidious this comparison is. Second, the biological explanation for African Americans’ athletic achievement obscures the fact that many other conventional arenas of success have been virtually closed to them in American society, and this is particularly true for those who are economically disadvantaged.

Decontextualized outcomes. As the Rushton and Bogaert studies make clear, research that produces invidious comparisons also tends to focus on decontextualized outcomes. Thus, for example, there is a wealth of evidence that the race of the tester makes a substantial difference in the tested IQ of both African American and European American children (Morawski, 1997). Nevertheless, many of those concerned with the question of race differences in intelligence simply discuss average scores, without regard to the contexts in which those scores are produced. Bowen and Bok’s (1998) landmark assessment of the life outcomes of Black and
White students who attended elite academic institutions in the early 1970s offers an important contrast. By carefully comparing students who differed in terms of initial SAT scores and race, but who experienced equivalent college educations, Bowen and Bok were able to assess the relative impact of test scores versus academic experience and credentials in predicting outcomes. Using this approach, they showed that outcomes for Black and White students with similar educations but different backgrounds and different entrance test scores were substantially the same.

Similarly, considerable research on gender and cognitive performance has simply focused on males’ higher scores relative to those of females on certain tests (for very different accounts of this literature, see Caplan, Crawford, Hyde, & Richardson, 1997; Kimura, 1999). Some researchers have shown, though, that these differences do not exist prior to adolescence (Hyde, Fennema, & Lamon, 1990); others have suggested that multiple-choice test formats may disadvantage females, who perform better on written tests (see Bridgeman & Moran, 1996; Halpern, 1997). Because there is now widespread “knowledge” about sex differences, merely naming a task or domain may affect expectations and performance (Eccles & Jacobs, 1986); indeed, Sharps, Welton, and Price (1992) found that changing the name of a task (from “a test of spatial abilities” to “a test of how people think about objects”) eliminated sex differences. Finally, a variety of researchers have suggested that experience relevant to tests may predict test performance. Baenninger and Newcombe (1989) concluded from a meta-analysis that both unplanned experience and direct training were associated with better performance on spatial tests for both sexes. The point here is that a variety of contextual factors—features of life experience and of the testing situation—matter, and researchers not aiming to “differentially discriminate” attend to these factors when they make comparisons.

Comparing means without reference to distributions. Simple comparisons of means may be invidious when they fail to take the distribution of scores for the two groups into account. A significant null hypothesis test indicates that two groups have different mean scores; however, all too often this test is interpreted to mean that there is some general difference between the two groups, such that individuals’ performances are either advantaged or disadvantaged by their group membership (Favreau, 1997). Note that this feature of invidious comparisons would not be remedied merely by requiring that researchers report effect sizes (although to do so would help set some difference findings in context). In an elegantly simple demonstration, Favreau showed how considerably overlapping groups may still show significant mean differences. She began with two samples, each with 25 observations, with significantly different means. She then added 85 more subjects with identical scores to each sample; although two-thirds of the subjects in each sample now had identical scores, the difference between the two groups was still significant, \( p < .01 \). For this reason, Epstein (1997) recommended making a distinction between differences that are statistically significant and those that have
social meaning. Favreau (1997) cautioned against the practical implications of research based on comparisons of mean differences; if, for example, girls (and only girls) were assigned to remedial math classes on the basis of such findings, many girls would receive training from which they could not benefit, and many boys would miss the chance for needed remediation.

**Comparisons-based questionable operationalizations.** Many studies of race and gender differences suffer from the use of measures with weak or dubious validity as indicators of the construct in question. Many questions have been raised about the real meaning, for example, of “intelligence,” and about the relationship between any given operation to assess it (e.g., IQ, SAT scores, etc.) and meaningful real-world indicators of performance (see, e.g., Cantor & Kihlstrom, 1987; Halpern, 1997; McClelland, 1973). This point was underscored by a recent report from a task force charged with creating an authoritative response to the debate on intelligence stimulated by the publication of *The Bell Curve* (Herrnstein & Murray, 1994). In that report, Neisser et al. (1996) noted that studies demonstrate that farmers in Liberia and children in Botswana performed better than comparable Americans at quantity estimation and storytelling tasks, respectively. It is clear that these tasks measure capacities that are more useful and relevant to success in those cultures than in ours. Should we then conclude that the (different) indicators valued in our culture are simply relevant to success in our culture? If so, why would women’s advantage over men in many tasks involving manual dexterity be viewed as irrelevant to the selection of good candidates for medical specialization in surgery (Halpern, 1994)?

The dubious validity of particular indicators reflects many studies’ dependence on an implicit (or explicit) value hierarchy of discovered race and gender differences. Halpern (1997) documented a number of areas in which females excel (knowledge of literature and foreign languages), as well as the fact that the areas in which males excel (math and science knowledge) are much more highly valued socially. These social values shape even the labeling of differences themselves. Thus, we think of men as being more vulnerable to “color blindness,” suggesting that it is “normal” to be sensitive to a full range of colors. We might, instead, think that women have an extraordinarily high level of color sensitivity. If what men excel at are the tasks that are viewed as important or difficult, then in those areas women will inevitably be seen as failing to measure up; if what women excel at are tasks that are viewed as trivial, then failure to measure up is clearly not important.

As we have pointed out, the five features of invidious comparisons discussed here can be viewed as examples of poor practice in research. For example, no social scientist would dispute that measures should be valid indicators, or that norms used in comparisons should be relevant. The problem is that consensus on general values—the importance of validity, or the importance of relevant standards for comparison—does not produce consensus on any particular evidence of validity or relevance. Although we believe that some of the research that has made invidious
comparisons does not stand up to any level of scrutiny on the criteria we have enumerated, in many cases the problem lies not with faulty reviewing judgments based on consensual values, but with disciplinary values themselves, and with the lack of a more routine critical perspective within the field. In the first instance—disciplinary values themselves—we join many others in pointing to the problematic satisfaction of our field with the use of decontextualized outcomes. The lack of attention to important features of context is more damaging in domains of social and political consequence than in some other areas, but it is common to a great deal of psychological research; indeed, it flows directly from the disciplinary embrace of the experimental laboratory method. And it is reinforced and rewarded by media enthusiasm for simple stories (which invidious comparisons surely are!) and aversion to complex ones (which contextualized findings always produce).

Even more troubling, though, some of the features we have identified—particularly judgments of the relevance of comparisons, the adequacy of operationalizations, or the hierarchical ordering of differences—depend on a relatively uncritical reliance on “the scientific method” to safeguard research quality. As we have seen in the examples we’ve discussed, these judgments often require us to notice features of our social world that violate common sense or our expectations; it is often difficult critical labor to notice those features and to question our unexamined assumptions (although, of course, new theoretical developments outside the field—feminist theories and critical race theories—may help us to do so). To the extent that we believe that the scientific method somehow automatically protects us from bias, we are quite likely not to notice problems with norms and operationalizations whenever they affirm common sense or the status quo. In contrast, research that violates common sense or the status quo will automatically be held to a higher standard precisely because it runs counter to our expectations.

It is, then, not difficult to understand how the research based on invidious comparisons we have documented here was published in peer-reviewed journals. Poor research on group differences need not be explicitly motivated by hatred. Certainly some disciplinary practices and values, if consistently applied, would provide better protection against invidious comparisons. Others, though, actually create a receptive climate for them (e.g., the preference for precise, but decontextualized, indicators; lack of critical stance toward affirmations of common sense or the status quo).

In sum, the five features we’ve discussed have at least two things in common, captured succinctly by feminist theorist Catharine MacKinnon’s (1987) assertion that “differences are inequality’s post hoc excuse” (p. 8). First, research that poorly operationalizes constructs or decontextualizes outcomes makes it easy to believe that the documentation of differences is undertaken to justify social inequalities. MacKinnon asked, “Why should women have to be ‘like’ men to be treated as equal citizens?” (1987, p. 9), arguing that it is wrong for legal and social equality to depend on demonstrations of sameness. The use of White men as the norm and the use of implicit value hierarchies, as well as a focus on mean differences, illustrate
her point that differences are taken as indications of inequality. Justice Blackmun, in his comments on the Bakke case (Regents v. Bakke, 1978), provided an answer to the parallel question in the domain of race: He suggested that “in order to get beyond racism, we must first take account of race. There is no other way . . . in order to treat persons equally, we must treat them differently” (pp. 311–312).

When Race and Gender Comparisons Are Not Invidious

Although in some quarters the race or gender difference debates have merely waxed and waned, in others researchers have created ways to “take account of race” and gender, and to “treat persons . . . differently” while avoiding invidious comparisons. These exemplars offer us models of how research can help us understand difference rather than exploiting it for political ends. They incorporate the following strategies: The lived experiences and social processes that define race or gender categories are identified and assessed independently; once these processes are identified, a wide range of relevant settings or interactions is sampled; the phenomenology of race and gender is captured in the voices of individuals on both sides of the dichotomy; appropriate control or comparison groups are used; and differences within race and gender groups are foregrounded.

Treating social categories as lived experiences and social processes. Some researchers assess the lived experiences and social processes that define race or gender categories, and thereby avoid treating them as though their meaning is self-evident and fully explanatory. Phinney (1996) pointed out that the impact of a static categorical variable like “minority status” shapes psychological outcomes only through factors such as “the history and present status of one’s ethnic group in society, personal experiences with prejudice, and one’s response to perceptions of stereotypes and discrimination” (p. 924). In this vein, Claude Steele and his colleagues (Steele, 1992, 1997) have explored how race differences in academic performance may result from a process of stereotype threat. They have shown that Black students in academic testing situations are affected by the salience of negative stereotypes about Black academic performance. Equally, differential stereotype threat also helps account for gender differences in math achievement (Shih, Pittinski, & Ambady, 1999). Interestingly, this research has also been extended beyond the context of academic performance. Stone, Lynch, Sjomeling, and Darley (1999) examined how stereotype threat might affect race differences in performance on an athletic task. They showed that in performance on the same task, White males were vulnerable to the stereotype that they lack “natural athletic ability,” whereas Black males were vulnerable to the stereotype that they lack “sports intelligence.” This thoughtful line of research into differences permits us to see how the same psychological process differently constructs race and gender for members of different groups.

Sampling a wide range of settings or interactions. Steele and his colleagues were able to discover that stereotype threat is a general process only because they
explored the possibility that the same (or similar) psychological processes were active in apparently disparate settings and groups. Failure to explore such possibilities may exaggerate differences; in other cases, it may leave a false impression of similarity. For example, Krieger and Sidney (1996) found that African American men and women reported roughly equivalent levels of racial discrimination in school, work, housing, and medical care. However, Black men experienced more discrimination on the street, from police, in the legal system, and in getting a job. Using open-ended narratives of urban residents’ experience of violence, Fine and Weis (1998) also found that Black men reported that police were the perpetrators of violence against them. Black women in this study reported some of the same kinds of police harassment, but they differed from Black men in reporting other kinds of street and domestic violence as well. Because these researchers included White and Latino men and women in this study, they were able to show that Latino men had similar experiences to Black men, whereas White men reported feeling victimized by street violence from men of color, and White women and Latinas reported domestic violence by men of their own ethnic group (although for different reasons). If Fine and Weis had only examined one kind of violence, these very different perspectives on these groups’ feelings of vulnerability could not have emerged.

**Attending to experiences of individuals on both sides of social dichotomies.** In addition to examining multiple contexts, these research programs captured the experiences of individuals on both sides of one or more social dichotomies (e.g., race/ethnicity, gender). Another way to avoid assuming that we already understand the meanings of categories is to examine (rather than take for granted) the meaning that respondents assign to their answers, even in forced-choice surveys. This was vividly illustrated by Landrine, Klonoff, and Brown-Collins (1992). They showed that Black and White women received comparable scores on Bem Sex-Role Inventory items; however, when women were asked to choose from alternative definitions for each item, group differences in meaning emerged. When rating themselves on the adjective “passive,” women of color most often meant “don’t say what I really think,” whereas White women most often meant “am laid-back/easy-going.” In this case, traditional psychological methods would fail to uncover an important difference—not an invidious comparison—that has implications for the validity of the measure and the interpretation of the results.

**Choosing appropriate control or comparison groups.** As long ago as 1923, Crane observed that the answer to the question “Is there a race difference?” requires “demonstrating that the difference is . . . due to the fact of race, and race alone and . . . involve[s] the testing of two groups of subjects—Black and White—who have been living under like environmental and cultural conditions” (p. 73, italics in original). He pointed, then, to the critical importance of having adequate comparison or control groups or indicators when attempting to attribute group differences to “race” or to “gender” (see also Parlee, 1981). This was demonstrated recently in the area of gender in Eagly and Wood’s (1999) reanalysis of data assembled by David Buss and his colleagues. Buss et al. (1990) argued that these
data indicated that men and women differ in fundamental ways based in evolution, because they showed a consistent cross-cultural pattern of gender differences in characteristics used in mate selection. Using United Nations data to control for the status of women, Eagly and Wood demonstrated that in countries where women had access to education and equal treatment under the law, these gender differences were dramatically reduced. These findings are consistent with the view that the characteristics men and women use in mate selection are at least partly socially constructed.

*Foregrounding differences within race and gender groups.* In the midst of the debate over racial differences in intelligence, Frumkin (1997) pointed out that some creative researchers had made more subtle and complex comparisons than can be made between a heterogeneous sample of “Blacks” and an equally heterogeneous sample of “Whites.” He noted comparisons as early as 1921 that attended to regional or migration differences between Blacks associated with quality of schools (Yerkes, 1921), as well as Peterson and Lanier’s (1929) demonstration that some (Northern) Blacks performed better than some (Southern) Whites on IQ tests. Klineberg (1935) incorporated attention not only to region or migration, but also to time in a given setting, by comparing IQ scores of Blacks who had migrated recently or longer ago with those of Blacks who were native to the South and had not migrated. These studies, like more recent efforts to incorporate statistical controls for the many environmental correlates of race (e.g., income, education, neighborhood characteristics, etc.), recognize that in order to establish differences attributable to race, they must heed Crane’s admonition to “[test] two groups . . . under like environmental and cultural conditions.”

Many of the studies we have discussed in this section have foregrounded differences within race and gender groups. This is important because attention to within-group differences highlights the fact that these categories are not unitary or homogeneous. For example, Hughes and Hertel (1990) explored “the significance of color” among Black Americans for ethnic consciousness and life outcomes; they showed that the magnitude of the difference in socioeconomic status between light- and dark-skinned Blacks was comparable to that between Blacks and Whites. Cole and Stewart (1996) found that the correlates of midlife political participation for midlife Black and White women were both similar and different. For both groups, social responsibility was related to political activity; however, political consciousness was associated for White but not Black women. The latter finding was due in part to Black women’s higher, and more homogeneous, scores on political consciousness. Studies like these illuminate not only the differences within groups, but the very meaning of each category.

**Conclusions**

Review of the study of race and gender differences provides ample evidence of polarized and inflammatory themes recurring in predictable patterns, producing
a dispiriting sense of stagnation. We found evidence of invidious comparisons between men and women and between Blacks and Whites, not only in the controversial work of Rushton, but throughout the field of psychology. Our point here has not been to argue that Rushton’s work is not wrong-headed, racist, and dangerous. Rather, we believe that it is instructive to view his work within the larger context of psychology’s longstanding interest in group differences. Within this context, his theoretical approach and the conclusions he drew from his data represent an extreme example of how invidious comparisons may be used to legitimate profound social and economic inequality between men and women, Blacks and Whites. However, as this survey of the literature makes clear, his methods and hypotheses are fully within the parameters of normal science as practiced within psychology today. Treating Rushton’s work as though it were an isolated case of racism within psychology, or as though he alone had an ideological or political agenda that necessarily distorted his findings, blinds us to the ways in which ideology and “common sense” assumptions about human groups influence all areas of psychology.

However, analysis of key features of these studies reveals the elements of research that result in invidious race and gender comparisons, permitting us both to notice them in published and unpublished studies and to avoid them in our own work. It is our hope that by identifying the features of invidious comparisons, the process of scholarly peer review may become sensitive to these features and can incorporate them into the criteria used to evaluate research. Moreover, outside of this maelstrom, creative researchers have developed strategies for truly illuminating race and gender differences. Unlike those whose research provides “scientific” legitimacy for stereotypes, these psychologists provide evidence that variously challenges stereotypes, provides explanations for differences that fit preexisting stereotypes, and uncovers differences in lived experience outside of stereotypes and our awareness.

Both those who seek evidence of deep difference and those who challenge it reveal that psychology is political—in our view, that is inescapable and inevitable. The former, though, simply confirm the status quo and produce a fruitless debate over how big or how bad the differences are. The latter focus not on the existence, size, or value of the differences, but on their meaning, sources, and place in the context of other kinds of distinctions that matter. This should not be surprising: As Harding (1991) reminded us, all science is a social endeavor that takes place in a particular historical and political context, and is necessarily influenced by that context. This is as much true of “good” science as it is of “bad” science. If, in the interests of objectivity, we ignore the political implications of our work, we are certain to legitimate and reproduce the prejudices of the social contexts in which our research takes place. If instead we aim to challenge and transcend these prejudices, it is crucial that we make ourselves aware of the characteristics of invidious comparisons and the methodological strategies to guard against them.
AUTHOR’S ADDRESS

Correspondence concerning this article should be sent to Elizabeth R. Cole, Women’s Studies Program, University of Michigan, 204 S. State St., Ann Arbor, MI 48109-1290 (Email: ecole@umich.edu).

REFERENCES

Invidious Comparisons


