experiment of Word, Zanna, and Cooper (1974), White Princeton students (perceivers) interviewed targets for a job. In fact, however, targets were confederates who had been carefully trained to engage in the same set of behaviors with each subject. Half of the confederate targets were African-American and half were White. The main dependent variables were interviewers' nonverbal behavior. Consistent with a self-fulfilling prophecy, the perceivers were colder to African-American targets than to White targets. In comparison to White targets, interviewers sat farther away from African-American targets, had more speech dysfluencies when talking to them, and comforted them less during the interview.

In their second experiment, Word et al. (1974) showed that this treatment undermined the performance of interviewees. Confederates were trained to interview subject-applicants in either of two ways: 1) the cold style comparable to that received by the African-American interviewees in Study 1, or 2) the warm style comparable to that received by the White interviewees in Study 2. All subject-applicants in this study were White. Results showed that the applicants treated coldly (i.e., the African-American applicants in Study 1) actually performed more poorly in the interview (as rated by independent judges) than did the applicants who were treated warmly. The type of treatment accorded African-American applicants in Study 1 undermined the actual interview performance of White applicants in Study 2.

Word et al. (1974) was an important landmark because it was the first experimental attempt to establish the potentially self-fulfilling effects of a social stereotype. However, even in this study, ethnic stereotypes were never measured. Perhaps the self-fulfilling prophecy was triggered, not by perceivers' stereotypes, but by their prejudice (disliking) of African-Americans. In fact, Word et al. (1974) ran a pilot study which documented that other Princeton students were prejudiced against African-Americans. Further, a recent study (Jussim, Nelson, Munis, & Soffia, 1995) found that if a group was often a more potent source of biases in person perception than were stereotypes (beliefs about the groups). Alternatively, the source of the self-fulfilling prophecy may be neither stereotypes nor prejudice. It may be society. People often feel anxious when interacting with members of a different ethnic group, especially when the groups have a long history of conflict (e.g., Stephan & Stephan, 1985). Clearly, the source of the self-fulfilling prophecy in this study remains to be pinned down more convincingly.

Regardless of the source, however, the basic finding requires replication. Would the same pattern of results hold up today, in colleges other than Princeton (where the study was conducted) and among nonstudent samples? Would African-American interviewees respond to the differential treatment received by interviewees in Word et al.'s (1974) second study in the same self-fulfilling manner as did the White interviewees? In addition, we cannot help but wonder whether other ethnic stereotypes are self-fulfilling. The answers to these questions are currently unknown.

B. SOCIAL CLASS STEREOTYPES

Abundant evidence shows that people hold higher expectations for individuals from middle-class backgrounds than for those from lower class backgrounds (Duske & Joseph, 1983; Jussim, Cohen, & Lerch, 1987). Nonetheless, we are aware of only two studies that have examined whether these expectations are self-fulfilling.

Perhaps the most dramatic and well-known study of social class-based self-fulfilling prophecies was performed by Rist (1970). Rist observed that by the eighth day of school, a kindergarten teacher had divided her class into two groups: students supposed to smart, average, or dumb. Each group sat at its own table (Tables A, B, and C, respectively). However, the main difference between the two groups was not intelligence—it was social class. Compared to the others, the students at Table A came from homes that had greater incomes, were less likely to be supervised by welfare, and were more likely to have both parents present; the children themselves were cleaner and more likely to dress appropriately. There were comparable differences between the students at Table B and C. Table A was positioned closest to the teacher, and she proceeded to direct nearly all of her time and attention to those students. In addition, she was generally friendlier and warmer to the students at Table A. Consequently, Rist (1970) interpreted his study as documenting strong self-fulfilling prophecies.

The differences Rist (1970) observed in teacher treatment of middle-class versus poor students would be inappropriate and unjustified, even if there were real differences in the intelligence of the children at the different tables. Nonetheless, despite Rist's (1970) conclusions, the study provided no evidence of self-fulfilling prophecy. Although Rist provided a wealth of observations concerning teacher treatment, he provided few regarding student performance. The differential treatment alone is not evidence of self-fulfilling prophecies. Differences in student outcome measures are also needed. The one student outcome measure that Rist (1970) provided was student IQ scores. In contrast to the self-fulfilling prophecy hypothesis, there were no IQ differences between the students at the different tables at the end of the school year. Thus, although the teacher may have held very different expectations for middle versus lower class students, and even though the teacher may have treated students from different backgrounds
very differently, the students' IQ scores were not affected (see Jussim & Eccles, 1995, for a more detailed critique of this case study).

A naturalistic study that included more than 10,000 high school students (Williams, 1976) provided a much more rigorous analysis of the role of social class in educational self-fulfilling prophecies. Williams (1976) used path analytic techniques to assess relationships between teacher expectations and students' previous and future achievement and social class. Consistent with most studies examining social class, Williams (1976) found that teachers held higher expectations for students from upper socioeconomic backgrounds. However, differences in teacher expectations for middle- and lower-class students evaporated after Williams controlled for students' previous levels of performance. This means that, rather than student social class influencing teacher expectations, teachers accurately perceive genuine differences in achievement among students from differing socioeconomic backgrounds. Of course, accurate expectations do not create self-fulfilling prophecies.

A colleague once described the Rist (1970) article as "a real near-jerk," and we cannot help but agree. Nonetheless, the less well-known Williams (1976) study is much stronger than Rist's (1970) study on almost all important scientific grounds: Rist relied primarily on his own subjective and potentially biased observations, whereas Williams relied on school records and questionnaires; Rist focused on 30 students, whereas Williams focused on more than 10,000 students; Rist claimed to provide strong evidence of self-fulfilling prophecy but actually provided none, whereas Williams (1976) rigorously tested for self-fulfilling prophecies and failed to find any. Although social class may sometimes lead to self-fulfilling prophecies, in terms of drawing scientific conclusions based on evidence, Williams (1976) deserves dramatically more weight than Rist (1970).

C. GENDER STEREOTYPES

Converging evidence from experimental and naturalistic studies shows that gender stereotypes create self-fulfilling prophecies. One series of experiments showed that when women believed they would be interviewed by more sexist or traditional men, they arrived wearing more traditionally feminine clothing (e.g., more makeup and accessories); and, if they believed he was attractive, they expressed more traditional gender-role attitudes on questionnaires and actually performed worse on anagrams test (von Baeyer, Sherk, & Zanna, 1981; Zanna & Pack, 1975). Another experiment showed that when targets (who were in a room isolated from perceivers and communicating only by using an electronic signaling system) were labeled as male (all were actually females), they took on more masculine and fewer feminine tasks than when they were labeled as female (Skyrms & Snyder, 1980).

Naturalistic studies, too, often find evidence of gender stereotypes leading to self-fulfilling prophecies. When first-grade teachers believe girls are smarter than boys, girls actually achieve more highly (Doyle, Hancock, & Kifer, 1972; Palfry, 1969). Another naturalistic study focused on the self-fulfilling effects of more than 1000 mothers' gender stereotypes on their children's self-perception of their ability in math, sports, and social activities (Jacobs & Eccles, 1993). This study showed that the children's sex interacted with their mothers' gender stereotypes. The children felt that they had more ability when their sex corresponded to the sex that their mother believed was generally superior. For example, among mothers who believed that boys were better at math, boys evaluated their math ability more highly than girls evaluated their own math ability. (This pattern was reversed among the minority of mothers who felt that girls were better at math.) These effects held even after the study controlled for prior achievement levels.

D. PHYSICAL ATTRACTIVENESS STEREOTYPES

Snyder, Tanke, and Berscheid (1977) showed that erroneous beliefs about another person's attractiveness may be self-fulfilling. Men and women were located in different rooms and communicated via telephone. Women believed to be attractive (a variable manipulated by photographs presented to the male perceivers) were treated more warmly than women believed to be unattractive. The women believed to be more attractive also responded with more friendliness.

A subsequent study, however, failed to replicate Snyder et al.'s (1977) findings, although it did yield highly qualified support for an attractiveness-based self-fulfilling prophecy. Anderson and Bem (1981) had androgynous or sex-typed men and women perceivers interact with men and women targets. In contrast to the Snyder et al. (1977) study, Anderssen and Bem (1981) did not find that male perceivers influenced women whom they believed were attractive to respond in more pleasant and socially skilled ways. Some allegedly attractive targets did respond more warmly than allegedly unattractive targets—but only when perceivers were sex-typed women. In contrast, androgynous female perceivers created a "boomerang" effect: Unattractive targets interacting with them were actually rated more favorably than were the attractive targets! These two experiments (Anderssen & Bem, 1981; Snyder et al., 1977) do not seem to provide a particularly strong
basis for broad statements about the self-fulfilling power of the physical attractiveness stereotype.

One naturalistic study showed that sometimes more attractive MBAs earn more income than their less attractive peers (Frieze, Olson, & Russell, 1991). Although this study was interpreted as showing self-fulfilling effects of the attractiveness stereotype, such an interpretation seems premature. Research consistently shows that physically attractive adults are more socially skilled than less attractive adults (e.g., Gollman & Levine, 1977; see a meta-analysis by Feingold, 1992). It seems likely that more socially skilled MBAs would deserve and actively receive higher pay than less socially skilled MBAs. Thus, attractiveness may predict MBAs' income because it is a proxy for social skill, rather than because of self-fulfilling prophecies (see Wusim & Eccles, 1998, for a more detailed critique).

Although the development of individual differences in social skill is beyond the scope of this paper, one may wonder where these differences originate. Is it not possible that self-fulfilling prophecies created a difference where none previously existed? However, the mere existence of social skill differences provides neither empirical evidence nor logical justification for supporting a self-fulfilling prophecy explanation (or any other explanation). There are many plausible alternative ways to explain why social skill differences between the attractive and unattractive exist. Furthermore, current evidence indicates that the expectancy explanation is one of the weakest explanations for those differences at any one point in time (see Feingold's [1992 meta-analysis]).

E. STEREOTYPE-BASED SELF-FULFILLING PROPHECIES AS EXPLANATIONS FOR SOCIAL INEQUALITIES

As we see it, the evidence that stereotypes lead to self-fulfilling prophecies that exacerbate or perpetuate social inequalities is currently extremely weak. To begin with, except for gender stereotypes, there just is not much evidence of any type—lab or naturalistic—showing that stereotypes actually do lead to self-fulfilling prophecies. Second, most of the studies showing stereotype-based self-fulfilling prophecies are experiments, which only demonstrate that stereotypes may be self-fulfilling. They provide no evidence that stereotypes actually are self-fulfilling in daily life. Only naturalistic studies are capable of documenting that stereotypes actually do create self-fulfilling prophecies. Except for gender studies, there are very few such studies that are well controlled (see reviews by Wusim & Eccles, 1995; Wusim & Fleming, in press).

The existence of social and economic inequalities is a phenomenon to be explained, but their existence does not provide prima facie evidence that all, or even most, ethnic or social class differences result from self-fulfilling prophecies. Social scientists seem to be committing at least one of two errors when they interpret the existence of inequalities as reflecting the effects of self-fulfilling prophecies (Snyder, 1984; Stangor, 1995; von Hippel et al., 1995).

The first error can be illustrated with a faulty syllogism. Premise 1: If minorities are genetically inferior to Whites intellectually, then minorities should, on the average, have lower educational and occupational achievement. Premise 2: Minorities have lower educational and occupational achievement than do Whites. Conclusion: Therefore, minorities are genetically inferior. Clearly, despite the currently popular claims in the Bell Curve (Herrnstein & Murray, 1994), this conclusion is unfounded. But the logic is no less inappropriate when the preferred explanation for inequality is a social one, such as self-fulfilling prophecy, rather than a biological one. Premise 1: If stereotypes are self-fulfilling, then one should find stereotype-consistent differences between various groups. Premise 2: There are stereotype-consistent differences between some groups. Conclusion: Therefore, stereotypes are self-fulfilling. Both premises are clearly true. The conclusion, however, does not follow from these premises and is an example of the classic error in logic known as "affirming the consequent." The second error involves a tendency to generalize too readily from artificial experimental laboratory studies to daily life. Laboratory experiments are extremely well-suited for testing theoretically driven hypotheses, identifying causality, and assessing conditions under which phenomena such as self-fulfilling prophecies are most likely to occur. However, lab experiments can only suggest possible explanations for real-life social phenomena. Whether such explanations hold true under naturalistic conditions is itself an empirical question that cannot possibly be addressed by experimental laboratory research. Thus, we agree that the experimental laboratory research does suggest that self-fulfilling prophecies might contribute to some social inequalities. However, in the absence of converging evidence from naturalistic studies, we also believe that it is premature and unjustified to conclude that self-fulfilling prophecies actually do make a major contribution to social inequalities.

The claim that stereotypes are self-fulfilling includes an occasionally explicit, but more often implicit, assumption: that stereotypes are, at least initially, inaccurate. This is because self-fulfilling prophecies, by definition, refer to erroneous expectations leading to their own fulfillment. Before empirically assessing the extent to which stereotypes bias teacher perceptions, we must first evaluate the validity of the assumption that stereotypes are generally inaccurate.
VI. Are Stereotypes Inaccurate?

A long tradition of social scientific research has assumed that stereotypes are generally inaccurate (see, e.g., Allport, 1954; LaPiere, 1936; Marger, 1991; Miller & Turnbull, 1986; Ottati & Lee, 1995; Snyder, 1988; Stagner, 1995). In this section, we examine some of the conceptual and empirical underpinnings of this assumption. First, we address the accuracy issue regarding people's beliefs about groups; second, we address the role of stereotypes in leading to biases and errors in person perception.

A. BELIEFS ABOUT GROUPS

An assumption or definition requiring stereotypes to be inaccurate quickly becomes mired in a swamp of conceptual and empirical troubles. Such a definition creates an undue burden on researchers interested in stereotypes: They must first document inaccuracy before they can consider a belief to be a stereotype. Unfortunately, there is enough research to determine the accuracy of most stereotypic beliefs. Consider the belief that librarians are introverted. If the definition of a stereotype requires it to be inaccurate, then this belief could not be construed as a stereotype. Because there have been no studies assessing the introversion of librarians, we are in no position to evaluate the validity of this stereotype.

A definition of stereotypes as inaccurate would also prevent researchers from considering demonstrably valid beliefs about groups as stereotypes. For example, stereotype conservatives could not study people who believe that girls do better in school than boys, that Asians are wealthier than most other ethnic groups, or that the majority of people on welfare are ethnic minorities. All of these beliefs are true (Deparole, 1994; Kimbell, 1989; Marger, 1991), and, therefore, would not qualify as stereotypes if stereotypes are, by definition, inaccurate.

The assumption that stereotypes are inaccurate is also empirically problematic, as stereotypes are defined as people's perceptions of the attributes of social groups (e.g., Ashmore & Del Boca, 1981). Most reviews of stereotyping conclude that there is very little scientific evidence regarding the reality or invalidity of many beliefs about groups. Moreover, the little empirical evidence that does exist provides a decidedly mixed picture (e.g., Brigham, 1973; Jud & Padt, 1993; Justine, 1990; Fiumon, McCauley, & Lee, 1995; McCauley, Stitt, & Segal, 1980; Ottati & Lee, 1995). Of course, validity is not just an issue of mean differences between groups; it also involves perceptions of the distributions of group members (e.g., Jud & Park, 1993; McCauley, 1995). However, the accuracy of people's beliefs about the distribution of social group on particular attributes has also been assessed rarely (see Lee & Ottati, 1993, for an exception). Because we know so little about whether perceptions of distributions are accurate or inaccurate, we are in no position to assume that stereotypes as perceived distributions are necessarily inaccurate.

There is also a peculiar irony in the claim the stereotypes are both inaccurate and self-fulfilling. If stereotypes are inaccurate (i.e., if people's beliefs about social groups do not correspond to the attributes of members of those groups), then we know that stereotypes are not self-fulfilling. If they were, the beliefs would be accurate (in that they would correspond to group members' attributes). Thus, a broad sweeping generalization, claims that stereotypes are both inaccurate and self-fulfilling are mutually exclusive.

Of course, stereotypes may be inaccurate, and through self-fulfilling prophecies become "accurate," which would create both social and conceptual problems. On the social problems side, it means that, even when two groups have similar distributions of skills, interests, motivation, and so forth, self-fulfilling prophecies may lead members of one group to excel (e.g., in school, on jobs, etc.) while underestimating the motivation, skills, and performance of other groups. Obvious to the social biases of such group differences, people can then point to the "objective" differences between groups as justification for maintaining pernicious stereotypes.

Conceptually, as described articulately by Snyder (1984), such processes seriously cloud the meaning of "accuracy." The "validity" of group differences created by perceiving themselves as a very precious sort of "accuracy." However, there are at least two conditions necessary for demonstrating that this possibility actually occurs: 1) The perceivers' stereotype must be shown not to correspond to some criterion at some Time 1 and 2) the perceivers' stereotypes must be shown to correspond to the criterion at Time 2. To our knowledge no one has published such data. Eccles and her colleagues are currently trying to get it; this issue through longitudinal developmental studies of the socialization of gender differences in abilities, self-concepts, performance, interests, and participation. Among early elementary school-age children, they have assessed individual differences (and gender differences) in sports, instrumental music, math, reading, and peer relations. They have asked the parents and teachers of these children to rate how well the children perform, how interested the children are, and how hard the children are trying to improve in each domain. They also
have given the children standardized measures of their current competence in each of these domains. This research shows that there are larger gender differences between the children’s sport self-perceptions and the parents’ perceptions of their children in the sports domain than in the standardized sport competence measures (Eccles & Harold, 1991). The researchers will use the longitudinal data to model the extent to which the parents’ perceptions lead to self-fulfilling prophecy practices, which in turn, lead to increases in the gender differences in actual competence. In our opinion, it will take longitudinal, field-based studies like this one to really address the question of whether stereotypes begin as inaccurate, and then, through self-fulfilling prophecies, become “accurate.”

B. STEREOTYPES AND PERSON PERCEPTION

Presumably, however, erroneous stereotypes are a social problem primarily if they lead to biases and discrimination. (If some people hold inaccurate social beliefs, but do not act any differently than others who hold accurate social beliefs, inaccuracy would not appear to be a major problem.) Inaccuracy becomes a problem when perceivers treat or evaluate members of one group differently than members of another group. Furthermore, even when a particular stereotype is accurate in a broad generalization, many members of the target group will not fit the stereotype. Therefore, even a generally accurate stereotype may lead to false expectations for many targets. Thus, one of the most important aspects of accuracy and inaccuracy in stereotypes involves their role in person perception.

In this area, too, social psychological theoretical perspectives have emphasized error and bias (e.g., DeVoe, 1989; Greenwald & Banaji, 1989; Jones, 1986; Miller & Turnbull, 1986). Consequently, stereotypes are frequently accused of being the cognitive culprits in prejudice and discrimination (e.g., Fiske & Taylor, 1984; Stanger, 1985). Others, however, have argued that the empirical evidence supporting the conclusion that stereotypes are generally inaccurate (by any criteria: perceived mean differences, distributions, or correlations) and lead to bias and discrimination is actually sparse, weak, and equivocal (see reviews by Jussim, 1990, 1991; Jussim et al., 1995; McCaulley, 1995; McCaulley, Stitt, & Segal, 1980; Oakes, Haslam, & Turner, 1994). Even the link between stereotypes and prejudice is often weak (Daly & Mladinic, 1989; Haddock, Zanna, & Esses, 1993; McCaulley & Thangavelu, 1991), and recent research has shown that, at least sometimes, prejudice is a more potent source of bias in person perception than are stereotypes (Jussim et al., 1995).

Again, however, most social psychological research on the role of stereotypes in person perception has been done in experimental laboratory studies. Lab research probably dominates perspectives on stereotypes for several reasons. First, lab research has several important merits. Tightly controlled studies are particularly well-suited for identifying some of the social and psychological processes relating stereotypes to person perception (e.g., Bodenhausen, 1988; Darley & Groce, 1983; Fiske & Neuberg, 1990; Krueger & Rothbart, 1988; Libville, 1982; Locksley, Borgida, Brekke, & Hargrave, 1989). Furthermore, experiments provide a stronger basis for drawing causal inferences than do naturalistic studies.

However, we suspect that there is also a second class of reasons for performing experiments that have questionable scientific merit. Laboratory studies of stereotypes and person perception often are easier to conduct than naturalistic studies. Researchers can create artificial targets on paper, slides, or videotapes to test any hypothesis. In general, laboratory researchers intentionally create targets from different groups who have identical personal attributes or engage in identical behaviors. Thus, any mean differences in judgments regarding targets from differing groups must represent bias because of this context. If experimental studies are by no means easy, then imagine a naturalistic study of stereotypes and person perception in contexts where discrimination is a major social issue. The researcher must first gain the cooperation of an organization (school, workplace, etc.) and the individuals in that organization. The researcher must then arrange to survey perceivers’ (teachers, managers, admissions or hiring personnel, etc.) judgments about targets (students, employees, applicants, etc.). Of course, those targets must actually vary on stereotype-category relevant dimensions (ethnicity, social class, sex, attractiveness, etc.). The researcher must then obtain two types of information from targets: their social group membership and their relevant personal attributes (e.g., school or job performance). Demonstrating bias then requires showing that perceivers see greater differences than really exist between individuals from the differing groups. Given the various obstacles and logistic difficulties, it is understandable that such research is daunting to so many social psychologists.

However, even this brief analysis highlights a political difference between lab and naturalistic research that examines whether stereotypes bias person perception. Because lab researchers have typically “operationalized away” differences between groups, they successfully avoided the political fallout that might accompany identification of real differences. In contrast, naturalistic study requires comparing perceivers’ judgments to some criterion. Doing so always leaves open the possibility that the groups may really differ on that criterion. Identification of real differences (e.g., between men
and women, between middle class and poor, or between ethnic groups) is almost always a delicate situation (e.g., Eagly, 1995; Graham, 1992; Jussim et al., 1995)—perhaps sufficiently delicate to intimidate many researchers away from dealing with such differences at all.

However, the failure of experimental research to examine the role of stereotypes in person perception when social groups really do differ is not just politically more palatable. It represents a major substantive limitation to all existing research on the issue of how well perceivers judge individuals from groups that really do differ on the attribute being judged. This is unfortunate considering that groups often differ in many ways (see, e.g., Eagly, 1995; Marger, 1994; McCauley et al., 1980; Steele, 1992; Swim, 1994; or the data on educational income, and family status available on various racial, gender, or geographic groups in any U.S. Census report).

Failure to study stereotypes and person perception when the groups really differ characterizes every experimental study of which we are aware (e.g., Hecht & Mills, 1985; Duncan, 1976; Krueger & Rothbart, 1988; Linville, 1982; Linville & Jones, 1980; Locksley et al. 1980; Locksley, Hepburn, & Ortiz, 1982), including our own (Jussim, 1993; Jussim et al., 1987, 1995). This means that social psychology actually provides little information about, for example, bias and accuracy to people's perceptions of individual men and women's assertiveness, the academic achievement of individual African-American and White students, or the social skill of individual attractive or unattractive targets (those groups really do differ on these attributes).

Furthermore, operationalizing away real differences prevents studies from assessing perceivers' sensitivity to existing differences between groups. Because bias assumptions are not mutually exclusive (see, e.g., Jussim, 1989, 1991; Jussim & Eccles, 1992), the finding of bias in lab studies provides little or no information whatsoever about perceivers' accuracy in detecting real differences. Thus, experimental laboratory studies that operationalize away real differences between groups probably underestimate social perceptual accuracy. Fiske and Neuberg (1990) have argued that, because naturalistic research is often so difficult, and because it is almost always impossible to obtain representative samples of relevant situations (job interviews, college admissions evaluations, etc.), all one can do is perform laboratory studies and generate logical arguments for how and when results from laboratory studies might be applicable to real-life situations. We have done just that. We conclude that the laboratory research is restricted to a situation that may rarely occur in the real world—one in which there are no differences between groups.

Of course, there is no theoretical or practical obstacle to conducting research on accuracy and bias in perceptions of individuals from groups that really do differ. If researchers can intentionally construct artificial targets who do not differ, they can just as readily construct targets who do systematically differ by their social group membership. Researchers can still avoid political fallout ("Why did you assume that Group X was superior to Group Y? Are you some sort of group-ist?") by counterbalancing the differences within a single study or by performing a second study in which the differences are reversed. In one set of conditions targets from Group X may be somewhat superior to Group Y; in another set, Group Y can be somewhat superior to Group X. Such a design could get at both bias (perhaps perceivers see a huge difference between the groups when X is better than Y, but no difference when Y is better than X) and accuracy (perhaps the perceived differences covary with the actual differences). Degree of difference can be set to either match real, known differences (see, e.g., Eagly, 1995; McCauley & Stitt, 1978; Swim, 1994) or set "arbitrarily" as needed to test theoretically based hypotheses (Mook, 1983).

Especially if objective data (standardized test scores, wins and losses in competitive games, grade point average (GPA), sales figures, words typed per minute, likelihood of having a college degree, etc.) were used as targets' personal characteristics (in current fashion, individualizing information), assessing accuracy and bias would be straightforward.

Even for more fuzzy attributes (laziness, extravertedness, ambition, etc.), one could scale the differences between targets through the use of independent judges rating the behaviors (without any group label). We can only speculate that the political and academic zeitgeist since the mid-1960s (see also Brigham, 1971; Eagly, 1995; Jussim, 1992; Jussim et al., 1995; Mackie, 1975; Oatley & Lee, 1995; Winfree, 1982) has created an intellectual environment that facilitated the field's failure to "notice" this glaring gap in research on stereotypes and person perception.

Of course, this type of research is not without its own limitations. Processes contributing to social inequality may take more time to surface than is available in most experiments. Nonetheless, this research would still be valuable, and we conclude that there is no serious obstacle preventing it. The lab studies also suffer one more extremely important conceptual limitation: They focus exclusively on identifying social-cognitive processor involved in stereotyping. However, even if processes are high in experimen- tal realism, they are completely incapable of drawing inferences about the accuracy of the content of stereotypes. This requires comparing judgments.
to real targets, not to laboratory stimuli (see Funder, 1987, 1995; Idd & Park, 1993; McCauley, Jussim, & Loe, 1995).

Because of unknown external validity and important conceptual limitations, we are compelled to conclude that the implications of much of social psychology's knowledge base for understanding the role of social stereotypes in naturally occurring person perception is not clear. Note, however, we are not claiming that the experiments are useless or trivial. Experiments have provided a great deal of knowledge about the social psychological processes leading to biases in judgments of similar targets and about the conditions under which such biases are most likely to occur.

However, if we are to address the practical issues involving prejudice, discrimination, and sources of inequality, then identifying accuracy or inaccuracy in the content and use of social stereotypes will require at least some research that meets three conditions. First, it should examine perceivers' judgments regarding targets who are real people with real attributes (as opposed to artificially created social stimuli). Second, there must be some means of measuring targets' attributes (a criterion). Third, perceivers' judgments must be compared to the criterion. In the next section, therefore, we describe two studies that meet these conditions as they analyze the role of stereotypes in biasing teachers' perceptions of students.

VII. Are Teacher Expectations Biased by Students' Sex, Social Class, or Ethnicity?

We now return to the question that sparked our conceptual analysis of self-fulfilling prophecies and stereotypes—why were teacher expectations affects stronger among girls, African-American students, and students from lower SES backgrounds? Classic social psychology suggests that stereotypes are types, because they are inaccurate and lead to biased perceptions of targets, would lead teachers to develop erroneous expectations for these students, which would then create self-fulfilling prophecies. However, the previous discussion also pointed out that this perspective is based almost entirely on experimental laboratory studies of unknown ecological validity using a conceptually limited paradigm.

Two studies were performed to help redress this limitation by examining naturally occurring person perception and by comparing those perceptions to criteria (Jussim & Eccles, 1995). The first study addressed accuracy by comparing teacher perceptions of achievement and motivation differences among students from differing sex, socioeconomic, or ethnic groups to actual differences among those students. Thus, this first study focused on the content of teacher perceptions. The second study examined the processes leading to accuracy and inaccuracy in teachers' perceptions of students from differing groups.

A. THE CONTENT STUDY

1. Main Research Questions

The study of content addressed three main questions: 1) Do teachers perceive sex, social class, or ethnic differences in achievement and motivation? 2) How accurate are these perceived differences (or lack of differences)? and 3) Are teachers' perceptions of individual students biased by teachers' own sex, social class, and ethnic stereotypes?

We use the term bias to refer to teachers systematically evaluating two groups as differing on some criteria more or less than they really do differ. For example, if banks approve more loan applications for Whites than for equally qualified minorities, the banks would be biased. However, if there are real income differences between ethnic groups, overall loan approval rates may differ among differing groups, even if banking officials are completely unbiased. Of course, if the difference in loan approval rates exceeds what would be predicted on the basis of income differences, the banks are still being biased.

We think that such a response represents one of the most common and critical forms of bias. In the teachers-student situation, it means that teachers see more (or fewer) differences between students from differing groups than really exist. In the MSALT data, higher SES students, on the average, perform better than lower SES students (this is discussed later in detail). Therefore, if teacher perceptions are biased against lower SES students relative to higher SES students, we should find that teachers perceive a greater social class difference, for example, in talent, that actually shows up in students' standardized test scores. Operationally, this means that SES should correlate more strongly with teacher perceptions than with students' actual performance.

Alternatively, there are at least two patterns that show teachers to be biased in favor of lower SES students relative to higher SES students. First, teachers might perceive that high SES students perform better than low SES students, but the perceived difference might be smaller than the real difference in previous achievement. Operationally, this means that the correlation between teacher perceptions and student SES would be smaller than the correlation between previous achievement and student SES. Second, teachers would be biased if they reversed the direction of the difference
(i.e., if they viewed the performance of lower SES students more favorably than that of higher SES students). Operationally, this means that the correlation between teacher perceptions and student SES would be in the opposite direction (have the opposite sign) as the correlation between student SES and student achievement.

Again we used the MSALT data set (described earlier) to explore accuracy and bias in teacher perceptions of students from differing demographic backgrounds. We first determined whether teachers perceived achievement and motivation differences among students from the different demographic groups by correlating teachers' ratings of students' performance, talent, and effort with students' demographic characteristics. Teachers were asked to rate each student in their classes. These ratings, therefore, were person perception measures. They were not teachers' beliefs about the differing groups in general (which were not assessed). These correlations indicated the extent to which teachers judged students from one group (e.g., boys) more favorably on the average than they judged students from another group (e.g., girls).

The content study did not address the accuracy of teacher perceptions of individual students within each demographic group. This question is conceptually like seeking to discover in an experiment whether perceivers rely on individualizing information more when judging targets belonging to one group (e.g., men) than when judging targets belonging to another group (e.g., women). Although this is an interesting and important issue, it is largely irrelevant to the issue of whether perceivers are biased for or against individuals from different groups. Showing, for example, that perceivers are more accurate when judging men than when judging women would provide no information at all about whether perceivers view men or women more favorably. In this example, perceivers' less accurate judgments of women could be, on the average, more favorable or less favorable, more accurate judgments of men. Accuracy of perceptions within groups is uninformative with respect to identifying whether there is a general bias or tendency to favor one group over the other. Obviously, however, the accuracy of teacher perceptions (and social perception more generally) within demographic groups is an important issue that should be addressed in future research.

Assessing whether teachers perceived differences between differing demographic groups of students, and whether those perceptions are accurate or biased is the focus of the content study. This involves determining whether teacher perceptions of individual students, aggregated across all students in each of two groups, correspond to the actual aggregated differences among the students in those groups. For example, this research addresses whether the differences teachers perceive between boys' and girls' performances (if any) overestimate, underestimate, or correspond to the real differences (if any) in boys' and girls' performances.

We compared the differing groups on measures of achievement and motivation. Final grades in fifth-grade math classes were used as the criteria for teacher perceptions of performance. Scores on standardized tests taken in fifth grade or early sixth grade were used as the criterion for teacher perceptions of talent. Three student motivation variables were used as criteria for teacher perceptions of student effort: self-concept of ability, self-perceptions of effort, and time spent on homework. These measures were reliable and valid (for more detail about the measures, see Eccles, 1988; Eccles (Parsons), Adler, & Meece, 1984; Jussim, 1987, 1989; Jussim & Eccles, 1992; Parsons, 1980, for a more detailed discussion of the use of these variables as criteria, see Jussim & Eccles, 1995). We considered self-concept of ability to be a motivational variable because it leads to effort and persistence according to many motivational theories (e.g., Bandura, 1977; Eccles et al., 1983; Eccles & Wigfield, 1985; Weiner, 1979). Consistent with this perspective, our data shows the correlation of self-concept of math ability with self-perceptions of effort to be .28 (p < .0001). We concluded that teachers were accurate when the size of the difference they perceived approximately corresponded to the size of the actual difference among students. Teachers' perceptions were inaccurate when the differences they perceived between different groups of students substantially deviated from the actual differences. They could be inaccurate in either of two directions: 1) They might overestimate differences between groups in the extreme, they might see a difference where none existed; or 2) they might underestimate differences between groups in the extreme, they might perceive no difference, when one existed.4

2. Sex

These analyses were based on 942 girls (coded as 1) and 847 boys (coded as 2) from the MSALT study. This was the subsample with valid data on all 4 These analyses did not control for classroom for several reasons. The criteria, student grades, standardized test scores, and motivations, were obtained from separate individuals, and did not suffer the same nonindependence problem associated with teacher perceptions. Although we could have controlled for classroom in identifying student grades, standardized test scores, and motivation, doing so would have unnecessarily complicated our results. Simple correlations (e.g., between student sex and grades), indicate which sex has received higher grades. This is what we used to know. For example, some schools track by ability, and whole classrooms might achieve at different levels from one another. If a disproportionate number of lower-class or African-American students were in low track classrooms, and we controlled for classroom, results would tend to underestimate the actual demographic difference in grades or standardized test scores. To be comparable to these criteria, therefore, analyses using demographic, motivational, and achievement variables to predict teacher perceptions also did not control for classroom.
variables. Teachers perceived girls as performing slightly better (r = −.08, p < .001) and as trying harder (r = −.16, p < .001) than boys. They perceived no difference in boys' and girls' talents (r = .00).

Were these perceptions accurate? For the most part, they were. Girls received slightly higher final grades than did boys in fifth-grade math classes (r = −.07, p < .01), a real difference that corresponds closely to the small perceived difference in performance. There also was no evidence of standardized test scores (r = .00), which corresponds with teachers' perceptions of no difference in talent.

However, there was no evidence that teacher perceptions of sex differences in effort were accurate. Boys and girls reported exercising the same amounts of effort (r = .00) and spending the same amount of time on homework (r = −.03, ns). Furthermore, boys had higher self-concepts of math ability than did girls (r = .09, p < .001).

Were teachers biased by students' sex? For performance and talent, the answer is no; for effort, the answer is yes. Teachers evaluated girls as trying harder than boys, even though boys and girls reported working equally hard, and even though boys had higher self-concepts of ability. Were teachers biased for or against girls? Because high effort is generally viewed positively by teachers and others (Covington & Omelich, 1979; Schuman et al., 1985), and because teachers rewarded supposedly harder-working students with higher grades (Jussim, 1989; Jussim & Eccles, 1992), this bias seems to favor girls. Alternatively, however, according to compensatory attributional perspectives [as one attribution goes up, others must go down (e.g., Covington & Omelich, 1979)], this result could be construed as a bias against girls because it implies lower teacher perceptions of girls' math ability. However, our results showing that teachers viewed girls and boys as having similar levels of talent strongly argues against this interpretation. Of course, whether this influences girls favorably in the long run depends on the psychological consequences of perceived effort for students' learning and motivation.

3. Social Class

These analyses were based on 1066 students. The multiple correlation of parental education and family income with each of the three teacher perception variables (all r's reported in this section are multiple r's) assessed whether teachers perceived differences between students from differing socioeconomic backgrounds. Teachers perceived students from higher SES backgrounds as performing better (R = .21) and as more talented (R = .26, both p's < .01). There also were real social class differences in achievements. Family income and education correlated with fifth-grade final grades (R = .27) and previous standardized achievement test scores (R = .31, both p's < .001). Thus, the size of the perceived social class differences closely corresponded to the size of the actual differences.

Teachers also perceived social class differences in effort. They viewed students from higher SES backgrounds as trying harder (R = .18, p < .01). Were there real SES differences in effort? Although there were no SES differences in self-reported effort or time spent on homework (both Rs < .05, ns), students from higher SES backgrounds had higher self-concepts of math ability (R = .15, p < .01). Teacher perceptions of student effort corresponded reasonably well with student SES differences in self-concept of ability, but not with the student reports of effort. Therefore, results for effort provided mixed evidence regarding accuracy and bias.

Overall, there was little evidence that students' social class biased teachers' perceptions. There was no evidence at all that teachers' perceptions of talent or performance were biased against students from lower socioeconomic backgrounds, although the results regarding teacher perceptions of effort were mixed.

4. Ethnicity

Analyses of ethnicity focused on teacher perceptions of African-American and White students. Did teachers perceive differences between African-American and White students? Answering this question turned out to be more difficult than it may seem because of the continuing patterns of school segregation apparent in this study. Owing to the ethnic differences between districts, we performed two sets of analyses. The first analysis examined teacher perceptions in the ethnically homogeneous districts. The second analysis examined teacher perceptions in the ethnically mixed district.

5. The Homogeneous Districts

Three groups were compared: 1) White students in the predominantly White districts; 2) African-American students in the predominantly White districts; and 3) African-American students in the predominantly African-American district. In each of these districts, none of the differences in teachers' perceptions of African-American versus White students were statistically significant (all P's > .25, all P's > .05). Teachers perceived no differences in the performance, talent, and effort of African-American and White students.

Were the lack of perceived differences in performance or talent accurate? Table VII presents the mean previous grades and standardized test scores for students in these districts, and shows that teacher perceptions were


### TABLE VII

<table>
<thead>
<tr>
<th></th>
<th>White students in the White district</th>
<th>African-American students in the White district</th>
<th>African-American students in the African-American district</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardized test scores (percentile ranks)</td>
<td>60 (p = 0.05)</td>
<td>52 (p = 0.09)</td>
<td>38 (p = 0.15)</td>
</tr>
<tr>
<td>Fifth-grade final grades B - F (n = 2040)</td>
<td>C+ - B (n = 27)</td>
<td>C+ (n = 61)</td>
<td></td>
</tr>
</tbody>
</table>

In both analyses, the difference between African-American students in the predominantly African-American district and White students is statistically significant (p < .05). In both analyses, the difference between African-American students and White students in the predominantly White district is not statistically significant.

(Purpose Statement: Is there a difference in achievement between African-American and White students? The study analyzed standardized test scores and final grades to determine if there was a difference in achievement between the two groups.)

### Partially Justified

In the predominantly White districts, neither the standardized test score differences nor the grade differences were statistically significant (all r's < 1.4, all p's > .1). Therefore, these teachers were justified in perceiving no differences between African-American and White students.

In contrast, as Table VII shows, students in the predominantly African-American district performed significantly more poorly according to both standardized test scores and previous grades than did White students in the predominantly White districts (both r's > 3, both p's < .01). Therefore, teachers were not justified in giving equivalent ratings for the performance of these African-American and White students.

Teachers were also mostly justified in perceiving little difference in the effort exerted by the different groups of students. The differences among students on three motivation variables reached statistical significance for time spent on homework (r = 4.68, p < .01) and self-concept of math ability (r = 5.08, p < .02), and marginal significance for perceived effort (r = 2.81, p < .07). However, only one of the post hoc comparisons was significant—African-American students in the predominantly African-American district had higher self-concepts of math ability than the White students in the predominantly White district (r = 2.11, p < .05). Furthermore, all of the r's were below .07, indicating that although statistically significant, the differences were minor.

### B. Teacher Expectations and Stereotypes: Preliminary Conclusions

Some answers to the questions guiding the content study are now available. Teachers did often (but not always) perceive differences between boys and girls, middle- and lower-class students, and African-American and White students. These perceptions were mostly accurate. For all three demographic groupings, teacher perceptions of the performance or talent of students belonging to different groups generally corresponded closely to the actual differences or similarities in these groups' previous grades and standardized test scores. There was only one exception to this pattern: Teachers rated African-American students in the predominantly African-American district as favorably as other students, even though both their grades and standardized test scores were not as high as those of other students.

The pattern for teacher perceptions of effort was more mixed, providing evidence of both accuracy and inaccuracy. Teachers believed that girls tried harder than boys, but there was no difference between the sexes on the effort measures, and boys felt that they had more math ability than girls felt they had. Thus, there was a small bias in favor of girls. Teachers believed that middle-class students tried harder than lower-class students, which did not correspond to the lack of social class differences in students' reports of their effort or time spent on homework, but did correspond closely to
student social-class differences in self-concept of math ability. In both the ethnically mixed and homogeneous school districts, the teachers perceived few differences in the effort of African-American versus White students. These perceptions were reasonably accurate—few differences emerged on either the effort measures or on self-concept of ability.

C. WHY WAS THERE SO LITTLE EVIDENCE OF BIAS?

1. Teachers Held No Stereotypes

The simplest explanation is that teachers held no negative stereotypes about girls, students from lower social-class backgrounds, or African-American students. If they did not perceive, then there would be no stereotype to bias their perceptions. Unfortunately, this possibility cannot be tested directly, because teachers’ beliefs about groups were not assessed as part of the MSALT study.

However, this explanation that teachers held no negative stereotypes seems highly implausible. If even a substantial minority held the stereotypes and relied on them, we still should have obtained some evidence of bias. Furthermore, abundant research in the social sciences attests to the widespread existence and importance of these stereotypes (e.g., APA Brief, 1991; Darley & Gross, 1983; Duske & Joseph, 1985; Fiske & Taylor, 1991; Jones, 1990; Marger, 1991). Thus, the likelihood that this sample of teachers is so unique that virtually none held stereotypes seems slight.

Second, some researchers have argued that one does not need to subscribe to a stereotype for that stereotype to influence social perception (Devine, 1989; Sedikides & Skowronski, 1991). Knowledge of cultural stereotypes (regardless of whether one accepts them), they argue, is sometimes sufficient to produce biases. For this type of analysis to explain our results, nearly our entire sample of teachers would need to be oblivious to the prevailing cultural stereotypes regarding girls, lower SES people, and African-Americans. This, too, seems highly implausible.

2. Teachers Held Stereotypes But Did Not Use Them

A second explanation is that teachers did hold stereotypes regarding these groups, but did not use them in evaluating students. Research in education and social psychology suggests considerable plausibility for this explanation. Earlier in this chapter we reviewed our own research that shows considerable accuracy in teacher perceptions of students (see also Brophy, 1983; Brophy & Good, 1976; Jussim, 1991, 1993, for reviews). In addition, abundant research shows that, whether individual targets are men and women, African-Americans and Whites, or upper class and lower class, perceivers generally judge them far more on the basis of their personal characteristics than on their membership in these social groups. This occurs both in laboratory studies and in naturalistic studies (see Jussim, 1990, 1993; Jussim, Madon, & Chatman, 1994, for reviews). Thus, even if teachers subscribed to social stereotypes, they probably judged students primarily on their academic performance rather than on their own stereotypes.

3. Teachers Used Valid Stereotypes

A third explanation is that teachers do hold stereotypes regarding these groups, and that they did, at least partially, rely on these stereotypes when judging students. Then why was there so little evidence of bias? If teachers relied on valid stereotypes when judging students, they would justifiably favor one group of students over another. They would have no tendency to exaggerate differences between the groups of students (this issue is addressed in more detail later in this chapter; see also Jussim, 1991).

D. WERE TEACHER PERCEPTIONS INFLUENCED BY ACCURATE OR INACCURATE STEREOTYPIC?

The content study showed that teachers’ perceptions of differences between students in the various groups mostly coincided with actual differences between the groups on comparable indicators. A follow-up study more deeply probed the processes by which teachers arrived at their judgments regarding students. We first developed a simple conceptual model for identifying whether teachers were relying on accurate stereotypes, inaccurate stereotypes, or no stereotypes when evaluating individual students. Additional analyses using the same MSALT data were performed to address the following two questions: 1) Did relying on an accurate stereotype facilitate accuracy in teacher perceptions? and 2) When teachers were inaccurate, did they inappropriately rely on their stereotype? To address these questions, two subquestions were examined: a) Did teachers rely on stereotypes when judging students? and b) If so, did relying on stereotypes enhance or undermine their accuracy? Thus, whereas the previous study focused exclusively on issues of content (e.g., did teacher perceptions of students from different groups coincide with real group differences), the next study focused on issues of process.

How can one discover whether teachers relied on stereotypes if stereotypes were not assessed? One can do so indirectly by using the methods
first developed in experimental social psychological laboratory studies of stereotypes and person perception. Studies in this area typically include no assessment of stereotypes. Instead, social psychologists typically manipulate targets' social group membership, manipulate information about targets' personal characteristics or hold them constant, then assess whether perceivers judge targets from one group differently than targets from another group (e.g., Beckett & Park, 1995; Bodenhausen, 1988; Darley & Groen, 1983; Dunham, 1976; Krueger & Rothbart, 1988; Linville, 1982; Locksley et al., 1980; see reviews by Fiske & Neuberg, 1990; Fiske & Taylor, 1991; Hamilton et al., 1990; Jussim, 1990; Jussim et al., 1994). If perceivers do judge targets from different groups differently (holding targets' behavior or attributes constant), then perceivers are assumed to be relying on their group stereotypes when they are judging individual targets.

This is the strategy we used to determine whether teachers relied on stereotypes in evaluating their students. Analyses assessed whether teachers perceived differences based on student sex, social class, or ethnicity while students' achievement and motivation were held constant. Specifically, we performed a series of regressions in which students' performance and motivation, as well as their social group memberships, predicted teacher perceptions. Operationally, therefore, the "relied on stereotypes" hypothesis was that teacher perceptions would be based on student group membership, even after we controlled for student performance and motivation. First, however, we present a general conceptual model of the relationships between targets' attributes, targets' group memberships, and perceivers' judgments of targets.

E. MODELS OF THE ROLE OF STEREOTYPES IN ACCURACY AND INACCURACY IN PERSON PERCEPTION

The Basic Model in Figure 8 captures the main aspects of our approach to identifying the role of stereotypes in person perception. This model is a variation on the reflection—construction model proposed by Jussim (1991) as a general framework for identifying relations between social perception and social reality. The Basic Model is a flexible tool that can be used with experimental or naturalistic data to address one of the major theoretical issues concerning stereotypes and person perception: Are perceivers' judgments of the differences between individuals belonging to differing groups biased? This has been a paramount question since Locksley's et al., 1980) controversial studies showing no bias in perceivers' judgments of assertive male and female targets. In Locksley's studies, and much subsequent research (e.g., Baron, Malloy, & Albright, 1995; Beckett & Park, 1995; Darley & Groen, 1983; Jussim et al., 1987, 1995; Krueger & Rothbart, 1988; Linville, 1982; Linville & Jones, 1980), researchers have sought to discover whether and when perceivers' a priori beliefs about group differences bias their judgments regarding individual targets. Bias, in this research, refers to seeing targets from different groups as different, even when their personal attributes (individualizing information) are identical.
The Basic Model incorporates the main ideas of this experimental laboratory paradigm, but also goes beyond them to allow for assessment of bias even when the groups really are different. The model has three main components: 1) the correlation between targets’ attributes and their group membership (r3); 2) the influence of targets’ attributes (i.e., individualizing information) on perceptions of targets’ group membership (Path A); and 3) the influence of targets’ group membership (i.e., perceivers’ stereotype) on perceivers’ judgments (Path B). For simplicity, we assume that Path A and B are standardized.

This Basic Model can be used to determine whether perceivers’ reliance on a stereotype enhances versus undermines the accuracy of their judgments regarding differences between targets belonging to different groups. Assume that r3 (not shown in the figure) is the correlation between perceivers’ judgments and targets’ group membership. In this model, perceivers see a difference that corresponds to the actual group difference when r3 = r1. In all experimental laboratory research on stereotypes and personal perception of which we are aware, r1 is intentionally rendered zero. In this situation, if perceivers think that groups really differ, and rely on their stereotype when judging targets (i.e., if Path B > 0), then they will judge targets from differing groups differently even though they are, on the average, the same. For example, they may judge a man as more assertive than a woman, even though both targets engage in identical behaviors (Beckert & Park, 1995; Krueger & Worchel, 1978). Rare cases (i.e., when Path B > 0), perceivers “see” a difference between a particular man and a particular woman that does not exist (i.e., r2 > 0, even though r1 = 0).

This model goes beyond the experimental laboratory research because it shows that the comparison of r3 to r1 is a more general index of bias. Even if there is a real difference between groups (i.e., r1 ≠ 0), the model shows that bias occurs when r3 > r1. For example, even if men are, on the average, more assertive than women (Swim, 1994), if perceivers judge the difference between individual men and women targets to exceed the real difference, their judgments are biased.

This model also shows that there are two separate routes to accuracy in perceiving the differences between targets from differing groups. With a few exceptions (Funder, 1995; Judiesch, 1991) social psychology has, thus far, only emphasized one route—judging targets on their personal attributes (i.e., Path A). The experimental lab paradigm shows no difference between groups, and perceivers will accurately perceive no difference if they judge targets solely on their personal attributes. In terms of the Basic Model, this is true because r2 = Path B + r1 (Path A).

"Judging solely on their attributes" means that Path B (the stereotype effect) is zero, and that Path A is high. However, because r1 (the real difference) is zero, r2 (the perceived difference) will also be zero, meaning that perceivers see no difference between targets from differing groups.

This “judging targets solely on their attributes” (Path A) route will also lead to accurate perceptions of group differences when the groups really do differ (i.e., r3 > 0). Even if there is no stereotype effect (Path B is zero), r2 (the perceived difference) approaches r1 (the actual difference) as Path A becomes larger (see preceding equation). Sometimes, perceivers may accurately detect a difference between groups because there really is a difference between individuals from differing groups, not because perceivers are stereotyping.

The second and less well-known route to accuracy in perceptions of group differences is through Path B. Conceptually, Path B represents perceivers relying on a stereotype. How can relying on a stereotype lead to accurate perceptions of group differences? If Path A is zero (i.e., perceivers are oblivious to targets’ personal attributes), and if Path B = r3 (i.e., perceivers stereotype the groups as differing to the same extent that they actually differ), then r2 = r1 (i.e., the perceived difference corresponds to the actual difference; see preceding equation). This simple model shows that perceivers who rely on a valid stereotype will accurately judge the differences between targets from different groups, even if they completely ignore the targets’ individual, personal characteristics. To put it somewhat differently, perceivers relying on an accurate stereotype could make nonzero errors in judgments of individuals, yet still arrive at judgments that, when aggregated across the individuals in each group, correspond to the actual difference between the groups. Perceivers’ judgments would be influenced by the stereotype without them being led to a bias for or against either group because they would see no greater differences between the individuals of differing groups than really exists.

1. Hypothetical Examples

Models 1 through 5 in Figures 8 and 9 present some hypothetical examples involving teacher perceptions and student sex in order to illustrate how the Basic Model can be used to distinguish between different aspects of accuracy and inaccuracy. Models 1 through 4 assume that there is a real difference between the achievement of boys and girls of −2 (coding girls as 1 and boys as 2 results in a negative coefficient when girls perform better than boys, as they did on our grades measure). In these models r3 = Path B − 2(Path A), where r3 is the correlation of teacher perceptions with student sex.
Model 3: Inaccuracy - The Stereotype is in the Wrong Direction

\[
\begin{align*}
\text{Student Achievement} & \rightarrow .5 \text{ Path A} & \rightarrow \text{Teacher Perceptions} \\
.2 \text{ Path B} & \\
\text{Student Sex} & \rightarrow .2 \\
\end{align*}
\]

Model 4: Inaccuracy - The Stereotype Exaggerates Differences

\[
\begin{align*}
\text{Student Achievement} & \rightarrow .5 \text{ Path A} & \rightarrow \text{Teacher Perceptions} \\
.2 \text{ Path B} & \\
\text{Student Sex} & \rightarrow .2 \\
\end{align*}
\]

Model 5: Inaccuracy - The Stereotype Assumes a Difference Where None Exists

\[
\begin{align*}
\text{Student Achievement} & \rightarrow .5 \text{ Path A} & \rightarrow \text{Teacher Perceptions} \\
.2 \text{ Path B} & \\
\text{Student Sex} & \rightarrow .2 \\
\end{align*}
\]

Fig. 9. Hypothetical models of inaccuracy in teacher perceptions of differences between boys and girls. Teacher perceptions of sex differences are inaccurate when the size of the perceived difference between boys and girls does not equal the actual difference between boys and girls. The perceived difference (not shown) equals the correlation between teacher perceptions and student sex, which equals Path B + \((.5 - \text{student sex, student achievement}) \times \text{Path A})

These models show that there are two ways for teachers to perceive an achievement difference between boys and girls that corresponds to the actual achievement difference. The first is shown as Model 1 in Figure 8. If teachers do not stereotype at all, but instead rely completely on achievement, then their perceptions will correlate \(-.2\) with student sex. In Model 1, Path A = 0 (indicating no stereotyping at all) and Path B = \(-2\) (indicating complete reliance on achievement information), so that

\[r_1 = -2 + (.5 \times 0) = -2\]

A second route to correspondence, however, is displayed in Model 2 of Figure 8. Even if teachers are oblivious to individual differences in student achievement, if they stereotype girls as performing better than boys (to an extent of \(-2\)), then they may perceive a sex difference that corresponds in size to the actual sex difference. In Model 2, Path A = 0 (indicating teacher obliviousness to achievement information) and Path B = \(-2\) (indicating a stereotype effect), so that

\[r_2 = -2 - 2 = -4\]

Figure 9 displays three models showing how this approach can also identify teachers' reliance on an inaccurate stereotype. In Model 3, the stereotype is in the wrong direction (Path B is positive rather than negative). In this model, Path A = -5, and Path B = 2, so that

\[r_3 = 2 - (-2 \times .5) = 3\]

In other words, teachers' stereotypes lead them to judge boys as performing better than girls, which is clearly incorrect because, in this example, girls perform better than boys.

In Model 4, teachers' stereotypes lead them to exaggerate real differences between the groups. In this model, Path A = -5, and Path B = -5, so that

\[r_4 = -5 - (-2 \times .5) = -4.5\]

The perceived difference between boys and girls (\(r_5 = 6\)) greatly exceeds the actual difference (\(r_1 = 2\)). In Model 5, the stereotype involves perceiving a difference (Path A is -2) when none exists (the correlation of achievement with sex, \(r_5 = 0\)),

\[r_5 = -2 + (.5 \times 0) = -2\]

Although these models are quite simple, involving only three variables, the same principles apply when correlations are decomposed in more complex models (e.g., Alwin & Hauser, 1975). Similarly, although we chose to focus on student sex in these examples, identical principles and processes hold for other groups. Obviously, these models could also be used to assess conditions under which bias is more likely to occur, either by including product–interaction terms, or by estimating the models separately in each different condition.
2 A Surprising Implication

These models inexorably lead to an implication that runs counter to an assumption implicit in much published research on stereotypes (e.g., Becket & Park, 1995; Bodenhausen, 1984; Darley & Gross, 1983; Kruger & Bodenhausen, 1988) showing that the influence of stereotypes on perceptions and judgments of individual targets is sufficient to demonstrate bias for or against a group. The models displayed in Figures 8 and 9, however, show that this is not necessarily so. The influence of stereotypes on judgments of individual targets is systematic group bias that stems from differences in over- or underestimate the differences between individuals belonging to different groups (or when it leads perceivers to over- or underestimate how much individuals differ). Even when behavior or attributes are not perfectly diagnostic, relying on targets' category will usually produce many errors in judging individual targets.

Laboratory studies skirt this problem by holding targets' behavior constant or by rendering it orthogonal to group membership, which forces the correlation between group membership and behavior to zero. However, as discussed previously, the correlation between group membership and individual behavior often is not zero in real life. For example, studies of the role of sex stereotypes in person perception (e.g., Becket & Park, 1995; Kruger & Bodenhausen, 1988) have rendered aggressive behaviors orthogonal to sex in an experimental setting. But even in these studies, sex is not independent of group membership. Of course, on the average, men really are more aggressive than women; in the experiments of Kruger and Bodenhausen, the effects of sex were not orthogonalized. However, with regard to sex and aggression, there are naturals conditions, \( r_{ij} = 0 \).

Consider the implications when this finding is combined with another frequent finding in the stereotype and person perception area. Virtually every study that manipulates targets' group membership and individualizing information finds that individuating information influences judgments more powerfully than does group membership (see, e.g., Juniss, 1986, 1991; Juniss, Madon, & Chaitman, 1994, for reviews). Consider the following hypothetical situation: 1) there is a real sex difference (e.g., \( r_{ij} = 2 \)); 2) targets' behavior influences judgments (e.g., Path A = 5); and 3) sex directly influences judgments in a small extent (e.g., Path B = 1). In this example, equation 1 is:

\[
\sigma = \beta + 2x(5) = 2 .
\]

Does the effect of sex on judgments represent bias in this situation? Holding all other aspects of this situation constant, would perceivers' judgments about the differences between boys and girls be more accurate if they did not use sex as a basis of judgment? The answer to both questions, in the hypothetical situation, is no. The real sex difference is 2, and the perceived difference is 2. If perceivers did not use sex as a basis for judgment (I.e., if Path B was 0), the perceived difference would be only .1, which, of course, underestimates the real difference.

We hope to avoid being misinterpreted here. We are not denying that stereotypes are generally accurate or that beliefs of unknown validity should influence social judgment. Furthermore, we doubt that there are any accurate stereotypes that are applicable to all members of the stereotyped group. Reliance on group membership, even when it correlates with target individuals' behavior and attributes, will always lead to less accurate (or less accurate assessments of individuals than will judging targets solely on the basis of their behavior or attributes if those behaviors or attributes are completely diagnostic of the characteristic being judged. Even when behavior or attributes are not perfectly diagnostic, relying on targets' category will usually produce many errors in judging individual targets.

However, error and bias are not always the same phenomenon. For example, perceivers may see some boys as more aggressive than they really are, and some girls as less aggressive than they really are. These are clear errors. Of course, perceivers also may see some boys as less aggressive than they are and some girls as more aggressive than they are. If perceivers' errors are systematic—if they are more likely to overestimate boys' aggressiveness and underestimate girls' aggressiveness—we would consider this to be a bias. However, if they are no more likely to underestimate than to overestimate boys' and girls' aggressiveness, there is no bias for or against either group (even if there are many within group errors). This whole analysis requires comparison of perceivers' judgments to some criterion—in the absence of a criterion, there can be no determination about the existence of systematic bias for or against a group. Once one has criteria, though, there will be a possibility that the groups will actually differ in some ways. With these principles in mind, the process study examined the role of stereotypes in enhancing or undermining teachers' perceptions of students from the different demographic groups.

F. THE PROCESS STUDY

The content study showed that teachers' perceptions of students belonging to different demographic groups generally corresponded well with the real differences and similarities between those groups, although teachers' perceptions were sometimes biased. The process study went a step further, and examined how teachers came to judge students from differing groups as being similar or different.
Did teachers rely on students' demographic group memberships when judging them? If so, did such reliance influence the extent to which teacher perceptions of differences corresponded to the actual differences between the groups? These questions were addressed in a series of analyses assessing the influence of student sex, social class, and ethnicity on teacher perceptions. Except where otherwise stated, all analyses used students' previous grades and standardized test scores, self-concept of math ability, time spent on homework, and self-perceptions of effort to predict teacher perceptions (intrinsic and extrinsic value were also used in preliminary analyses but were almost completely unrelated to teacher perceptions). In addition, each analysis also included students' sex, social class, or ethnicity as predictors.

1. When Teacher Perceptions Were Biased

First, we consider the interpretations of the process analyses when the content study showed that teacher perceptions were biased. Teacher perceptions of effort favored girls, even though there was no evidence of sex differences in effort. Therefore, teachers could not have based their perceptions on individualizing information. If teachers were basing their perceptions on stereotypes, the process study would have yielded a path coefficient similar in magnitude to, and of the same sign as, the zero-order correlation between teacher perceptions of effort and student sex.

Teachers also perceived the performance and talent of students in the predominantly African-American school district as favorably as they did those of other students, even though those other students received higher grades and higher standardized test scores. The process analyses should also have yielded small positive path coefficients linking student ethnicity to teacher perceptions of performance and talent. They should have been small because previous analyses showed bias only among a subset of African-American students (those in the predominantly African-American district). Moreover, the path coefficients should have been positive because African-American students were coded as 2, and Whites as 1. Positive path coefficients mean that given their performance, African-American students were viewed more favorably than others.

2. When Teacher Perceptions Were Accurate

Results from the content study have shown that many of the teachers' perceptions of the differences between students belonging to the differing groups were accurate. This held true for teachers' perceptions of boys' and girls' performance and talent; teachers' perceptions of SES differences in performance and talent; and teachers' perceptions of ethnic differences in performance, talent, and effort in the predominantly White districts. The process study is not intended to assess the accuracy of these perceptions—accuracy is an issue of content, not process, and has already been demonstrated. Drawing on the models in Figures 5 and 9, the process study is intended to assess how and why teacher perceptions became accurate. One possibility is that teachers relied on an accurate stereotype (a belief that groups differed to about the same extent that they actually differed). This possibility would be reflected in a path coefficient linking the student demographic variable to teacher perceptions that is similar in both sign and magnitude to the zero-order correlation between that demographic variable and teacher perceptions. Such a path coefficient would show that teachers by judging students according to their group membership had arrived at a perception of group differences that corresponded to existing actual group differences (of course, such a result is mute on the issue of accuracy and error within groups).

A second possibility is that teachers arrived at accurate perceptions of differences between students belonging to different demographic groups without relying on stereotypes. Instead, they may have relied exclusively on students' personal characteristics and accomplishments (individualizing information). Because the groups really did differ on some of these variables, teacher perceptions of the groups would also differ if teachers used the individualizing information. This possibility would be reflected in near-zero path coefficients linking student demographics to teacher perceptions, and high coefficients linking other student variables (grades, standardized test scores, effort, etc.) to teacher perceptions.

3. Gender Stereotypes

The main questions here were: 1) Would student sex predict teacher perceptions, independent of the other variables? and 2) If so, did the student sex effect enhance or undermine the accuracy of the teachers' perceptions of differences between boys and girls? Table VIII summarizes the results from these analyses. The results showed that teachers seemed to be relying on an accurate stereotype when judging students' performance. The beta relating student sex to teacher perceptions of performance was .49, which closely corresponded to the sex differences in grades of .07 (found in the content study). Although teachers did judge students on the basis of their performance, doing so was not the main source of the correlation between teacher perceptions and student sex, despite the fact that girls did get better grades than boys. The small independent effect of student sex on teacher perceptions accounted for most of the small correlation between sex and
**TABLE VIII**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Beta relating to teacher perceptions of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance $R^2 = .41^*$</td>
</tr>
<tr>
<td>Student sex</td>
<td>-.06</td>
</tr>
<tr>
<td>Fifth-grade final marks</td>
<td>.32*</td>
</tr>
<tr>
<td>Standardized test scores</td>
<td>.36*</td>
</tr>
<tr>
<td>Self-concept of math ability</td>
<td>.22*</td>
</tr>
<tr>
<td>Effort self-perceptions</td>
<td>.05*</td>
</tr>
<tr>
<td>Time spent on homework</td>
<td>-.06*</td>
</tr>
</tbody>
</table>

* $p < .05$

Beta are standardized regression coefficients.


Results for teacher perceptions of effort suggested reliance on an inaccurate stereotype. The beta relating student sex to teacher perceptions of effort was −.16 ($p < .001$), even though the correlations of student sex with self-concept of ability, time spent on homework, and self-perceptions of effort were .09, .03, and .06, respectively. Teachers apparently erroneously stereotyped girls as trying harder, despite the similarities between boys’ and girls’ actual effort, and despite boys’ higher self-concept of ability.

4. Social Class Stereotypes

Analyses were identical to those examining teachers’ sex stereotypes, except that instead of student sex, these analyses included parental education and income in the equations predicting teacher perceptions of performance, talent, and effort. Results are presented in Table IX.

These analyses provided no evidence that teachers relied on social class stereotypes. The $R^2$ increment associated with adding family income and education to the regression equations never exceeded .02 and was never statistically significant (all $F$'s $< 2.3$, all $p$'s $.1$). Of the six possible individ-

**TABLE IX**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Beta relating to teacher perceptions of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance $R^2 = .40^*$</td>
</tr>
<tr>
<td>Parental income</td>
<td>.02</td>
</tr>
<tr>
<td>Parental education</td>
<td>.03</td>
</tr>
<tr>
<td>Fifth-grade final marks</td>
<td>.21**</td>
</tr>
<tr>
<td>Standardized test scores</td>
<td>.35**</td>
</tr>
<tr>
<td>Self-concept of math ability</td>
<td>.20**</td>
</tr>
<tr>
<td>Effort self-perceptions</td>
<td>.00</td>
</tr>
<tr>
<td>Time spent on homework</td>
<td>-.06</td>
</tr>
</tbody>
</table>

* $p < .05$

Beta are standardized regression coefficients.


The lack of a perceived difference for talent may seem to conflict with our previous research (Justen, 1989; Justen & Elices, 1992) showing that student gender significantly predicts teacher perceptions of talent. However, there is no real conflict. The current analyses differ from others that we have reported previously in one crucial way. The older studies also allowed teacher perceptions of performance to predict teacher perceptions of talent and effort, whereas the current analyses do not. For example, in Justen and Elices (1992), the effect of gender on teacher perceptions of talent was .08; the effect of gender on teacher perceptions of performance was −.20; and the effect of teacher perceptions of performance on teacher perceptions of talent was .64. The total effect of gender on teacher perceptions of performance in Justen and Elices (1992), therefore, was .08 + (−.20)(.08) = .08, which corresponds exactly to the result reported here.
ual relations between income and education and the three teacher perception variables, only one was statistically significant (education predicted teacher perception of ability, p < .05). Apparently, the accuracy of teacher perceptions of social class differences in performance, talent, and effort occurred because teachers evaluated students on the basis of their achievement and motivation—factors which correlated with social class.

5. Ethnic Stereotypes

The main analysis included 1888 White students and 76 African-American students. (Again, because we used previous standardized test scores as a control, and because students in the ethnically mixed district did not take a standardized test in fifth or early sixth grade, they were not included in this analysis.) For this group, the correlations of ethnicity with grades and standardized test scores were -.12 and -.14, respectively (both p's < .001). Ethnicity was coded Whites = 1 and African-Americans = 2, so these small negative correlations mean that, overall, White students had somewhat (but not dramatically) higher grades and standardized test scores than did African-American students.

Table X presents the results of the regression analysis predicting teacher perceptions. These results consistently show small but positive relations between ethnicity and teacher perceptions (.05, .11, and .10, respectively, for teacher perceptions of performance, talent, and effort, all p's < .01). This means that teachers perceived African-American students slightly more favorably than they perceived White students with similar achievement histories and motivational patterns.

These results, however, do not necessarily represent bias in the sense of teachers evaluating the African-American students in their classes more favorably than similar White students. It seemed likely that this pattern largely reflected differences between the predominantly African-American district and the other districts. To test this possibility, we ran another set of regressions. These were identical to the first with one exception: They excluded the 53 students in the predominantly African-American district. Although there were only 23 African-American students remaining, this analysis directly tested whether teachers were biased in favor of African-American students over White students. In this analysis, all three bens relating ethnicity to teacher perceptions were reduced to .04 (although with the large overall N, they were still statistically significant or marginally significant). These results provide little evidence of bias in favor of African-American students over White students. Instead, they show that the main source of the apparent positive bias was the teachers in the predominantly African-American district (who evaluated their students as favorably as the other students, despite poorer performance).

To increase the number of African-American students, we performed one additional analysis. Still excluding the predominantly African-American school district, we did not use standardized test scores that had been used before as predictors of teacher perceptions. This allowed us to include the students in the ethnically mixed district. These analyses, which included 1873 White students and 37 African-American students, also yielded no evidence that teachers relied on ethnicity. The three bens relating ethnicity to teacher perceptions were all below .03, and none were even marginally significant.

G. ACCURACY IN TEACHER PERCEPTIONS OF STUDENTS FROM DIFFERING DEMOGRAPHIC GROUPS: CONTENT AND PROCESS

The present study provided some clear insights into the reasons why the results of the content study showed such minimal evidence of bias. With a few notable exceptions, teachers based their perceptions of students on those students' actual performance and motivation. Student social class did not influence teacher perceptions, after controls were used for students' actual achievement and motivation. Nor did ethnicity, except for the slightly