SOCIAL PERCEPTION, SOCIAL STEREOTYPES, AND TEACHER EXPECTATIONS: ACCURACY AND THE QUEST FOR THE POWERFUL SELF-FULFILLING PROPHECY

Lee Jussim

Jacquelynne Eccles

Stephanie Madon

I. Introduction

How are social perception and social reality related? Social psychology has long emphasized the power of beliefs to create reality, and the power of interpersonal expectancies to create social problems (e.g., Gage & Cronbach, 1955; Jones, 1986, 1990; Merton, 1948; Miller & Turnbull, 1986; Snyder, 1984). Social scientists have a longstanding interest in one particular source of expectations—stereotypes—largely because stereotypes may contribute to social inequalities and injustices. But are people so malleable that they readily fulfill others' inaccurate expectations? How accurate are interpersonal expectations? To what extent do stereotypes bias person perception and lead to self-fulfilling prophecies? Who is most vulnerable to self-fulfilling prophecies?

In this article, we address these questions as follows. First, we present a brief overview of research on accuracy, error, bias, and self-fulfilling prophecies. Second, we review our own research showing that teacher expectations predict student achievement mainly because they are accurate, although they do lead to small self-fulfilling prophecies and biases.

We subsequently embark on a quest to identify conditions under which self-fulfilling prophecies might be considerably more powerful. Third, therefore, we report the results of new research showing that teacher expectancy effects are more powerful among girls, students from lower socioeconomic status (SES) backgrounds or African-Americans.

Social psychological research on stereotypes suggested a possible explanation for this pattern: Teachers rely on stereotypes in developing expectations for students from stigmatized groups, and because such expectations will often be inaccurate, they are also more likely to be self-fulfilling. Therefore, we review the general literature on the role of stereotypes in creating self-fulfilling prophecies and on issues of accuracy and inaccuracy in stereotypes more generally.

We then address some of these issues empirically in two studies that examined whether teacher perceptions of differences among students belonging to different demographic groups (boys or girls, middle class or poor, African-American or White) were biased or accurate. Although these studies provided some evidence of bias (surprisingly, these biases were usually in favor of students from culturally stigmatized groups), they also showed that, in general, differences in teachers' perceptions of students from the differing groups corresponded well to actual differences between those same groups of students.

Although we found such results particularly interesting in light of the social sciences' emphasis on inaccuracy of stereotypes, they left us still unable to explain why self-fulfilling prophecies were stronger among students from stigmatized groups. We therefore speculate that students who feel devalued in school will be particularly susceptible to confirming teachers' expectations. Although we cannot test this idea directly, we summarize another new study that provides indirectly supportive evidence by showing that students with low self-concepts of ability or with previous records of low achievement were, much like students from stigmatized groups, considerably more vulnerable to self-fulfilling prophecies. We conclude this chapter by reviewing other moderators of expectancy effects, discussing the evidence showing whether self-fulfilling prophecies accumulate or dissipate over time, and making recommendations for future research on self-fulfilling prophecies.

II. Accuracy, Error, Bias, and Self-Fulfilling Prophecy

A. A BRIEF OVERVIEW

Research on accuracy, error, bias, and self-fulfilling prophecy have long traditions in social psychology. Error and bias research dates back at least to the emphasis in the 1930s on the inaccuracy, irrationality, and rigidity of social stereotypes (e.g., Katz & Braly, 1933; LaPiere, 1936). The idea that many social injustices and inequalities reflect self-fulfilling prophecies

was first suggested in the 1940s (Merton, 1948). There was lively interest in accuracy through the 1950s (e.g., Taft, 1955; Vernon, 1933), which came to an abrupt and premature end after Cronbach identified many seemingly difficult statistical and methodological problems involved in assessing certain types of accuracy (Cronbach, 1955; Gage & Cronbach, 1955). At the same time, the New Look in Perception, which emphasized a myriad of ways in which perceivers' goals, needs, fears, and motives could influence and undermine the veridicality of perception, initiated a revolution in approaches to perception, at least in social and personality psychology (e.g., Allport, 1955; Bruner, 1957).

Subsequently, from the 1960s through much of the 1980s, social and personality psychology emphasized a host of errors and biases in social perception (e.g., Kahneman & Tversky, 1973; Miller & Turnbull, 1986; Nisbett & Ross, 1980). Central to this effort was the work by Rosenthal demonstrating that experimenters and teachers can evoke expectancy-confirming behavior from both animals and people (see Rosenthal, 1974, for a review; Rosenthal & Jacobson, 1968; see Rosenthal & Rubin, 1978, for a meta-analysis). Numerous researchers then followed up this work with studies of the potentially self-fulfilling effects from expectancies of all sorts in and out of the laboratory (see reviews by Jones, 1977; Jussim, 1986; Snyder, 1984).

Although a few researchers did attempt to keep the study of accuracy alive after the 1950s (Archer & Akert, 1977; McCauley & Stitt, 1978), they were rare voices barely heard above the din of the zeitgeist emphasizing error, bias, and self-fulfilling prophecy. In the 1980s, however, four articles sparked the beginning of a renaissance of interest in accuracy. McArthur and Baron (1983) presented the first coherent theoretical alternative to the constructivist zeitgeist that had dominated thinking about social perception for 30 odd years. They took the ecological approach, which was originally developed to study object perception (Gibson, 1979), and applied it to social perception. This theory emphasized the information in the stimulus, which was in sharp contrast to the social cognitive emphasis on the categories, prototypes, schemas, and assorted cognitive structures existing in the perceiver's mind.

Next, Swann (1984) presented a broad and sweeping review of research on accuracy. Perhaps most influential was his discussion of "circumscribed" accuracy. Swann (1984) argued that perceivers often have no interest in predicting the behavior of targets across all situations and for all time. Thus, it is inappropriate to hold them to this standard. Instead, he suggested that people are usually content to understand and predict others' behavior only when they interact with those others. In terms of this more circum-

scribed notion of accuracy, Swann (1984) speculated that people might be considerably more accurate than had previously been recognized.

Next, Funder (1987) presented a conceptual and empirical assault on what he believed was social psychology's misplaced emphasis on error and bias. He made two main points: 1) Social psychology's knowledge base regarding error and bias stemmed almost exclusively from studies that were originally designed to assess social perceptual processes, and 2) the studies assessing process did not address the accuracy of outcomes produced by such processes. Funder drew a parallel between the laboratory social cognitive work on illusions, biases, and errors, and the laboratory vision research on illusions. Researchers used controlled visual illusions to probe the dynamics of visual information processing; they never assumed that these illusions reflected deficiencies likely to occur in vision under natural conditions. This, Funder argued, was also the most appropriate interpretation for research on human judgment. For Funder, accuracy was an issue of content, not of process. He also presented data documenting people's moderate to strong accuracy in perceiving others' personalities.

At about the same time, Kenny was publishing numerous articles describing his social relations model (See Kenny, 1994 for a review.). In one of the most influential of these early articles, Kenny and Albright (1987) 1) explained how the social relations model could be used to isolate error and accuracy in social perception; 2) pointed out the similarity between the accuracy components assessed by the social relations model and Cronbach's components of accuracy; and 3) showed how, when applied to social interaction, the model empirically documented considerable accuracy in social perception.

By 1990, the accuracy djinni was most of the way out of the bottle. One more paper popped the cork completely. A main bastion of the scholarly emphasis on error and bias was the expectancy effects literature, especially the literature on social stereotypes [see, e.g., the strong emphasis on the power of expectations to create reality in reviews by Jones (1990), Miller & Turnbull (1986), and Snyder (1984)]. In contrast, Jussim (1991) argued on both theoretical and empirical grounds that this emphasis was misplaced. He presented a model showing how people's beliefs could be in touch with reality most of the time, and yet still sometimes produce biases in person perception leading to self-fulfilling prophecies. This model was then used to interpret previous research on the effects of interpersonal expectancies. Jussim concluded that 1) interpersonal expectancies can lead to biases and self-fufilling prophecies, but these effects tend to be quite small; 2) perceivers' predictions of targets' future behavior and their impressions of targets' past behavior tend to be reasonably accurate; and 3) the evidence on the accuracy of social stereotypes is quite mixed (some accuracy, some

inaccuracy). Jussim also showed that much of what looked like expectancy-induced bias in experimental laboratory studies could actually enhance person perception accuracy under some naturally occurring conditions.

The revival in interest in accuracy has, however, with a few exceptions, occurred in parallel with continued interest in error and bias. Many researchers still study and emphasize error, bias, or self-fulfilling prophecy (e.g., Gilbert & Malone, 1995; Snyder, 1992; Stangor, 1995); others focus primarily on accuracy (e.g., Ambady & Rosenthal, 1992; Borkenau & Liebler, 1992; Funder & Colvin, 1988; Levesque & Kenny, 1993). Although several researchers have attempted to integrate accuracy and bias (e.g., Brewer, 1988; Fiske & Neuberg, 1990; Higgins & Bargh, 1987; Kunda, 1990), most have relied primarily on experimental laboratory studies (see Kenny, 1994, for a review of partial exceptions—nonexperimental laboratory studies of accuracy and bias). These attempts, therefore, suffer from the conceptual problem first identified by Funder (1987): Because they focus on process instead of content, and because their relevance to naturalistic situations is unclear, they provide little empirical evidence on accuracy, error, bias, and self-fulfilling prophecy in daily life. (Even some hardcore experimental social psychologists have expressed sympathy with the view that the relevance of laboratory studies to daily life is an unanswered, open, empirical question.) (See Gilbert & Malone, 1995, p. 35.)

Although some researchers have argued that one can only generate logical arguments to show why results from laboratory studies would be applicable in many real-life situations (e.g., Fiske & Neuberg, 1990), we respectfully disagree with this pessimistic sentiment. We do agree that lab experiments alone can never reveal how much the discovered processes actually occur under naturalistic conditions. We would be left to speculate.

Although claims about ecological or external validity are not necessarily a crucial component of all studies (e.g., Mook, 1983), they are essential for generalizing one's findings to the natural world. Moreover, many social psychologists actually suggest, either explicitly or implicitly, that the phenomena under study do indeed occur with considerable frequency outside the social psychological laboratory (see, e.g., Jussim, 1991 for a review of such claims). One of the best examples of such claims is the 1991 American Psychological Association (APA) brief to the U.S. Supreme Court in the case of Hopkins versus Price-Waterhouse. Drawing primarily (although not exclusively) on the results of experimental laboratory studies, the APA brief argued that we know quite a lot about stereotypes, including both the processes by which they lead to bias and discrimination, and the conditions that either facilitate or undermine their tendency to lead to bias and discrimination.

The APA brief was true to the spirit and content of many, if not most, perspectives on social stereotypes current at the time of the case

(e.g., Brewer, 1979, 1988; Fiske & Taylor, 1984; Hamilton, Sherman, & Ruvolo, 1990; Jones, 1986; Snyder, 1984; and subsequently, e.g., Greenwald & Banaji, 1995; Stangor, 1995; Von Hippel, Sekaquaptewa, & Vargas, 1995), all of which assume that laboratory studies of bias, error, and self-fulfilling prophecy have good enough ecological validity to justify generalizing the findings to real-world settings. Whether this assumption is warranted, however, is an empirical question. Moreover, if we plan to make policy recommendations and expect our findings to be relevant to legal decisions, then it behooves us to find out what people are actually like in their natural habitats despite the difficulties involved in doing naturalistic research.

Understanding the nature and extent of accuracy, error, bias, and self-fulfilling prophecy under naturalistic conditions is the focus of our own studies on the relations between teacher expectations and student achievement. Because these studies examine teachers and students in sixth-grade public school math classes, they suffer none of the ecological validity problems of laboratory experiments. We believe that these studies provide a wealth of evidence regarding naturally occurring social perception and interaction, evidence that bears directly on long-standing issues in social psychology regarding the prevalence of accuracy, bias, and self-fulfilling prophecy in daily life. We begin by identifying three separate ways in which targets may confirm perceivers' expectancies.

B. THREE SOURCES OF EXPECTANCY CONFIRMATION

Perceivers' expectations may be confirmed for any of at least three reasons—two that involve influences of expectations on behavior or perceptions and one that does not. First, perceivers' expectations sometimes produce *self-fulfilling prophecies:* Their initially erroneous expectations may cause targets to act in ways consistent with those expectations (Cooper, 1979; Darley & Fazio, 1980; Jussim, 1986; Rosenthal & Jacobson, 1968). Second, expectations may lead to *perceptual biases:* perceivers may interpret, remember, and/or explain targets' behavior in ways consistent with their expectations. This type of expectancy confirmation exists in the mind of the perceiver rather than in the behavior of the target (Darley & Fazio, 1980; Eccles & Jacobs, 1986; Jussim, 1991; Miller & Turnbull, 1986). Self-fulfilling prophecies and perceptual biases both represent perceiver expectations creating (or "constructing") social reality, either creating an objective social reality (when self-fulfilling prophecies change targets' actual behavior) or a subjective social reality (when perceptual biases influence perceivers' evaluations of target behavior). Finally, in contrast, expectations

also may accurately reflect or predict social reality without infuencing either objective target behavior or even subjective perceptions of that behavior (Brophy, 1983; Jussim, 1991).

Although these three expectancy phenomena are conceptually distinct, they are not mutually exclusive. Any combination of the three (or none at all) can characterize relations between perceiver expectations and student achievement (Jussim, 1989, 1991; Jussim & Eccles, 1992). Consider a teacher who believes a student is especially bright. The teacher may be (largely) accurate—this student may indeed have a stronger academic background than most others. Furthermore, highly positive interactions with the teacher may lead this student to achieve even more highly—thus, demonstrating a self-fulfilling prophecy. Finally, perceptual biases may lead the teacher to evaluate the student even more favorably than is warranted by the student's objective performance.

Although expectations may lead to many combinations of self-fulfilling prophecy, perceptual bias, and accuracy, they may also lead to none; expectations can be both inaccurate and noninfluential. For example, a teacher may expect a student to be a low achiever. Nevertheless, this student may successfully complete most homework assignments in a timely and thorough manner and go on to perform above average on a highly credible standardized achievement test and receive mostly "As" on in-class tests. The teacher may simply acknowledge the error (i.e., the original expectation was erroneous, but there is no perceptual bias), and the student may continue to perform highly (no self-fulfilling prophecy).

Although this article focuses exclusively on relations between teacher expectations and student achievement, expectancy effects undoubtedly occur in many other relationships: employer–employee, therapist–client, and parent–child. Consequently, as we analyze ways to distinguish among self-fulfilling prophecies, perceptual biases, and accuracy, and examine processes underlying expectancy-related phenomena, our discoveries may have some relevance and applicability to many other relationships beyond teachers and students (see also Eccles et al., 1993; Eccles & Hoffman, 1984; Jacobs & Eccles, 1992; Jussim, 1990, 1991; Jussim & Eccles, 1995; Jussim & Fleming, in press).

III. Teacher Expectations

There are few contexts more important for investigating self-fulfilling prophecies than teachers' expectations for their students. Ever since Rosenthal and Jacobson's (1968) seminal and controversial (e.g., Elashoff & Snow,

1971) Pygmalion study, writers in both scholarly journals and the popular press have implicated teacher expectations as a major perpetrator of injustices and inequalities based on ethnicity,[1] social class, and gender (see Wineburg, 1987, for a review). In this article, we present evidence suggesting that such claims present an oversimplified and exaggerated picture of the role of teacher expectations in perpetuating social inequalities. This evidence will convey two main points. First, we briefly review our own and others' research documenting that naturally occurring teacher expectations generally lead to only small self-fulfilling prophecies and perceptual biases. This research also shows that teacher expectations predict student achievement primarily because they are accurate.

Even though teacher expectation effects are generally small, under some conditions or among particular groups, such effects may be considerably larger than usual. Second, therefore, we report the results from some of our efforts to discover instances of more powerful self-fulfilling prophecy effects.

A. ACCURACY MORE THAN SELF-FULFILLING PROPHECY

Through the 1980s and early 1990s, social psychology abounded with testimonies to the power of expectancies to create social reality (e.g., Fiske & Taylor, 1984; Hamilton et al., 1990; Jones, 1986, 1990; Snyder, 1984; see Jussim, 1991, for a review). In contrast, most educational and developmental psychologists argued that expectancy effects were generally minimal (e.g., Brophy, 1983; Brophy & Good, 1974; Cooper, 1979; Eccles & Blumenfeld, 1985; Eccles & Wigfield, 1985; West & Anderson, 1976). Evidence from naturalistic studies consistently failed to support the strong claims of social psychologists, and instead confirmed the perspective of the educational and developmental psychologists, rarely uncovering expectancy effects larger than .1 to .2 in terms of standardized regression coefficients (see Jussim, 1991; Jussim & Eccles, 1995, for reviews). Furthermore, research in educational settings has repeatedly shown that teacher expectations predict student achievement mainly because they are accurate (see Brophy, 1983; Jussim, 1991; Jussim & Eccles, 1995, for reviews). Because two of our studies provided some of the clearest evidence of teacher accu-

---

[1] Throughout this chapter, the term "ethnicity" primarily refers to African-Americans and Whites. Although the term "race" is used far more widely in reference to these groups, it has little clear scientific meaning (e.g., Marger, 1994; Yee, Fairchild, Weizmann, & Wyatt, 1993). Ethnicity is also a fuzzy concept, although it generally includes physical appearance, similar geographical roots, a unique culture, sense of community, and ascribed membership (Marger, 1994). The term "ethnicity" has a considerably less controversial history and we therefore consider it preferable.

racy to date (Jussim, 1989; Jussim & Eccles, 1992), we describe them as follows in some detail.

B. THE JUSSIM (1989) AND JUSSIM AND ECCLES (1992) STUDIES

1. The Data

All studies described in this chapter are based on the Michigan Study of Adolescent Life Transitions (MSALT), which assessed a variety of social, psychological, demographic, and achievement-related variables in a sample that included more than 200 teachers and 3000 students in the sixth and seventh grades (see Eccles et al., 1989; Midgley, Feldlaufer, & Eccles, 1989; Wigfield, Eccles, MacIver, Reuman, & Midgley, 1991, for more details about this project). A total of about 100 teachers and 1700 students in sixth-grade math classes were the focus of the two studies we summarize here. Both studies tested the hypotheses that 1) teacher expectations early in the year are based on students' previous achievement and motivation and that 2) teacher expectations, student motivation, and students' previous achievement influence students' subsequent achievement (for detailed descriptions of the models and analyses, see Jussim, 1989; Jussim & Eccles, 1992).

Three sixth-grade teacher expectation variables were assessed in early October: perceptions of students' performance, talent, and effort at math. We assumed that teachers inferred students' effort and talent, in part, from their own perceptions of students' performance. Measures included student motivation self-concept of math ability, intrinsic and extrinsic value of math, and self-reports of effort and time spent on math homework. Fall and spring assessments of these motivational variables were included in Jussim (1989); only fall assessments were included in Jussim and Eccles (1992).

There were two measures of previous achievement: final marks in fifth-grade math classes and scores on standardized achievement tests taken in late fifth or early sixth grade. There were two outcome measures of achievement: Final grades in sixth-grade math classes and scores on the math section of the Michigan Educational Assessment Program (MEAP), a standardized test administered to students in Michigan early in seventh grade. All measures were reliable and valid (for more detail, see Eccles (Parsons), Adler, & Meece, 1984; Eccles-Parsons, Kaczala, & Meece, 1982; Jussim, 1987, 1989; Jussim & Eccles, 1992; Parsons, 1980).

2. Results

Because results reported here are from two studies, they are presented in pairs as follows. The first refers to Jussim (1989) and the second refers

to Jussim and Eccles (1992). Although the main analyses were performed using the LISREL VI program, all *betas* reported below are interpretable as standardized regression coefficients.

These two studies were the first to explicitly assess and compare self-fulfilling prophecy, perceptual bias, and accuracy. Both studies assessed models that were more complex versions of the model presented in Figure 1. In brief, we assessed whether teacher perceptions early in the school year predicted changes in achievement (by controlling for previous achievement) over and above changes accounted for by motivation (self-concept of ability, valued placed on math, effort, etc.). Table I summarizes the major results from both studies.

Consistent with the self-fulfilling prophecy hypothesis, teacher perceptions of students' math performance in October of the sixth grade significantly related to students' final grades in sixth-grade math (betas = .21, .34) and students' seventh-grade MEAP scores (betas = .10, .15). In Jussim's (1989) study 1) teacher perceptions of talent significantly related to both sixth-grade math grades (beta = .12) and seventh-grade MEAP scores (beta = .17); and 2) teacher perceptions of performance significantly predicted changes in students' self-concept of math ability across the sixth-grade school year (beta = .11).

Results consistent with the perceptual bias hypothesis showed that teacher perceptions of students' effort significantly predicted sixth-grade math grades (betas = .19, .19) to a larger extent than they predicted seventh-
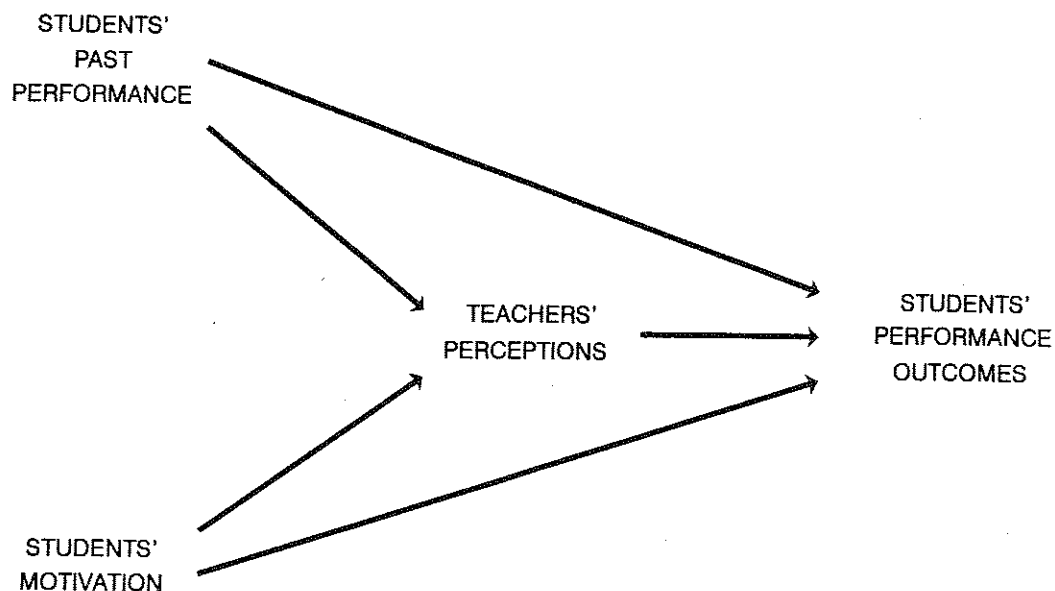

Fig. 1. Conceptual model of relationships between teacher perceptions and student achievement.

TABLE I
Construction and Reflection of Student Achievement

	Teachers' perceptions of students' performance	Teachers' perceptions of students' talent	Teachers' perceptions of students' effort	All three
Correlation with grades	.54, .71	.57, .64	.50, .63	.63, .73
Effect on grades	.21, .34	.12, .04	.19, .19	.21, .29
Correlation with MEAP scores	.46, .54	.57, .49	.34, .36	.57, .55
Effect on MEAP scores	.10, .15	.17, −.02	.00, −.07	.13, .12

Effects refer to standardized total effects (see Jussim, 1989, and Jussim & Eccles, 1992, for more detail about the models). The first entry within each column is for Jussim (1989, $N = 429$), the second entry is for Jussim & Eccles (1992, $N = 1288$). All coefficients greater than .03 are significant at $p < .05$. The difference between correlations and path coefficients is an index of predictive accuracy (see text for explanation). The column titled "All three" reports the multiple correlation of all three teacher perception variables with grades and MEAP scores in the "Correlation" rows and reports the multiple semipartial correlation (controlling for the student background variables) with grades and MEAP scores in the "Effect" rows. (Reprinted with permission from *Journal of Personality.*)

grade MEAP scores (betas = 0, −.07). Teachers assigned higher grades to students whom they believed had exerted more effort. This hypothetically could have represented accuracy—if teachers were rewarding students who actually were working harder. Instead, however, as the results suggest, teachers simply assumed that higher achievers were working harder, whereas we found no evidence that the students who received the higher grades actually worked any harder than their peers. In fact, the students who received low grades reported spending more time on homework than the other students (Jussim, 1989; Jussim & Eccles, 1992). Because effort is difficult to observe directly, we speculated that teachers, perhaps influenced by a belief in a just world (Lerner, 1980) or by the protestant work ethic (Schuman, Walsh, Olson, & Etheridge, 1985; Weber, 1930), simply assumed that "hard work pays off." Therefore, high achievers "must" have been working harder. A consequence, however, is that the academically "rich" (the high achievers) get richer (teachers assign them grades that are even higher than they deserve).

There was both accuracy and inaccuracy in teacher perceptions. Teacher perceptions were largely accurate because they were most strongly linked to appropriate factors: previous grades, standardized test scores, teacher

perceptions of in-class performance, and student motivation (the multiple correlation of these factors with teacher expectation variables ranged from about .6 to .8). The results also provide evidence of a small but consistent pattern of gender bias in teacher perceptions, but we discuss this issue in detail later in the chapter.

Results from both studies also provided considerable evidence of predictive accuracy. The zero-order correlation between teacher expectations early in the year and student achievement late in the year equals expectancy effects (influences of teacher expectations on student achievement) plus predictive accuracy (teachers basing their expectations on factors that influence student achievement). Therefore, one index of predictive accuracy is the difference between the zero-order correlation and the size of the expectancy effects (see Jussim, 1989, 1991, 1993; Jussim & Eccles, 1992 for more detailed explanations).

The zero-order correlations of teacher perceptions with seventh-grade MEAP scores ranged from .34 to .57, and the path coefficients ranged from −.07 to .17. The path coefficients relating teacher perceptions to MEAP scores accounted for about 20–30% of the zero-order correlations between initial teacher perceptions and subsequent MEAP scores; the remaining 70–80% represented predictive validity without influence (i.e., accuracy). There was a similar pattern for final grades in sixth-grade math. Zero-order correlations of initial teacher perceptions with year-end grades ranged from .50 to .71. Path coefficients ranged from .04 to .34. The path coefficients relating teacher perceptions to grades accounted for about 30–40% of the zero-order correlations between initial teacher perceptions and subsequent grades; the remaining 60–70% represented accuracy.

Are these results anomalous? Not at all: Research consistently shows that the zero-order correlations between teacher expectations and student achievement generally range from about .4 to .8, and that path coefficients representing effects of teacher expectations on student achievement are generally about .1 to .2 (see Brophy, 1983; Brophy & Good, 1974; Jussim, 1991; Jussim & Eccles, 1995, for reviews).

3. Limitations

The correlational nature of this research leaves open some alternative explanations for the relation between teacher expectancies and student achievement. However, because this study used longitudinal data, reverse causal influences are not possible. Students' achievement at the end of the sixth grade could not have caused teacher expectations at the beginning of the sixth grade.

The main limitation involves omitted variables. Path coefficients only reflect causal effects when all relevant causes of student achievement have been included in the model. If teacher expectations and student achievement are both caused by a third variable that has been omitted from the model, then the model may yield inflated path coefficients relating teacher expectations to student achievement (a "spurious" relation). Unfortunately, no matter how many potential sources of spuriousness are assessed, it is impossible to know if all such sources have been included.

Although this problem can never be completely overcome, it can be minimized with the inclusion of extensive control variables. Few naturalistic studies have included more controls (previous achievement test scores *and* grades, self-concept, and several motivational variables) than we have (Jussim, 1989; Jussim & Eccles, 1992). Thus, these findings provide some of the clearest evidence to date that teacher expectations influence the achievement of some students. Such a conclusion will warrant revision when future research demonstrates empirically that there are important sources of accuracy in teacher perceptions other than those assessed in this study.

It is also important to understand the nature of this limitation. Suppose we omitted important variables that cause both student achievement and teacher expectations. This has a very specific implication—that teachers are even more accurate than we have suggested [i.e., that teacher perceptions predict student achievement less because of their causal influence than because both teacher perceptions and student achievement are based on a third (set of) variables(s)]. Conceptually, of course, this "critique" strengthens our conclusion that teacher expectations predict student achievement more because they are accurate than because they create self-fulfilling prophecies. Nevertheless, it is important to keep these limitations in mind throughout the rest of this chapter. The potential for an omitted variable problem is always present in naturalistic research.

IV. In Search of the Powerful Self-Fulfilling Prophecy

We found our results showing small expectancy effects and high accuracy very surprising. Not only did the social psychological zeitgeist of the 1970s and 1980s emphasize error and bias in judgment and perception, but also accuracy was an all but taboo subject (see, e.g., Funder, 1987; Jussim, 1991; Kenny, 1994, for reviews). Similarly, the social psychological literature had for so long emphasized the power of self-fulfilling prophecies (Darley & Fazio, 1980; Jones, 1986; Merton, 1948; Miller & Turnbull,

1986; Rosenthal & Jacobson, 1968; Snyder, 1984) that we expected to find effects considerably larger than .1 and .2. Although one of us has taken the lead in arguing that such effects are small (Jussim, 1989, 1990, 1991, 1993; Jussim & Eccles, 1992), this position was based on the empirical data base, not on any preconceived notions that such effects always are, or must be, small. In fact, all the authors of this chapter became interested in self-fulfilling prophecies because of their potential to further understanding social injustice and the construction of social reality. We have, however, challenged researchers to empirically identify naturally occurring conditions under which self-fulfilling prophecy effects are large (Jussim, 1989, 1991, 1993; Jussim & Eccles, 1992). Nonetheless, although expectancy effects may be generally small, we strongly suspected that there were conditions under which expectancy effects were substantially larger. Thus, we embarked on a quest to identify more powerful self-fulfilling prophecies. As the next section shows, we have had some success.

A. STUDENT DEMOGRAPHICS AND SUSCEPTIBILITY TO SELF-FULFILLING PROPHECIES

Understanding the role of demographics in educational and occupational attainment has been a major goal of many of the social sciences for a long time. Research on sources of demographic differences is almost always controversial and highly politicized—regardless of whether it is research arguing for genetic explanations of group difference (e.g., Herrnstein & Murray, 1994) or research contending that schools oppress girls [American Association of University Women (AAUW), 1992)]. Still, clarion calls and pleas for social psychological research directly addressing issues of race, class, or gender periodically appear in the literature (e.g., Carlson, 1984; Fine & Gordon, 1989; Graham, 1992). The two phenomena (politicization and lack of research) are probably related: The potential for controversy and outright vilification is so strong that it may deter many scholars from vigorously pursuing research programs into these issues. Yet laboratory researchers often imply or explicitly insist that bias detected by highly artificial procedures or under highly unusual conditions provides insights into biases against African-Americans, women, and people from lower SES backgrounds in naturally occurring social interactions (e.g., Devine, 1989; Greenwald & Banaji, 1995; Von Hippel et al., 1995).

At the minimum there seems to be a broad consensus that issues of race, class, and gender are extremely important. Research on the role of student demographics in moderating teacher expectation effects appears to be particularly important. Such research goes to the heart of the question of

whether, how, and how much teacher expectations contribute to social problems and inequality. This alone would be sufficient to justify the exploration of the extent to which student demographics moderate teacher expectation effects.

In addition, however, there were several theoretical arguments leading us to suspect that students from stigmatized groups would be more susceptible to self-fulfilling prophecies than students in general. Abundant evidence suggests that school is often an unfriendly place for many African-American and lower SES students (e.g., Lareau, 1987; Steele, 1992). Although school can be difficult places for both boys and girls, though usually in different ways (e.g., Bye, 1994; Jussim & Eccles, 1995), math and science classes are often less supportive places for high achieving girls than for high achieving boys (Eccles & Blumfeld, 1985; Eccles & Hoffman, 1984; Eccles-Parsons et al., 1982; AAUW, 1992). When school is consistently a difficult place, students may often "disidentify" with achievement by devaluing the importance of school or the particular subjects in which they feel devalued (e.g., Eccles (Parsons), 1984; Eccles (Parsons) et al., 1983; Eccles & Harold, 1992; Jussim, 1986; Meece, Eccles-Parsons, Kaczala, Goff, & Futterman, 1982; Steele, 1992). Such responses may render them more readily influenced by teacher expectations in several ways.

When students with a history of negative school experiences find themselves faced with a supportive, encouraging teacher who also insists on high performance, they may feel as if they have caught a breath of fresh air. Such a teacher may inspire previously low achievers to new heights. This perspective may not be as unrealistic as it sounds. In his influential article on Black disidentification with school, Steele (1992) described academic programs in which previously low-performing students [e.g., some with Scholastic Aptitude Test (SAT) in the 300s] took on difficult honors-level work and came to outperform their White and Asian classmates. Steele's (1992) description of these programs implies that teachers often engage in behaviors much like those that lead to beneficial self-fulfilling prophecies in the classroom and workplace: They are challenging and supportive (e.g., Brophy & Good, 1974; Cooper, 1979; Eccles & Wigfield, 1985; Eden, 1984, 1986; Harris & Rosenthal, 1985; Jussim, 1986; Rosenthal, 1989).

However, these same underprivileged students may also be more susceptible to harmful self-fulfilling prophecies. Steele (1992) has articulately argued that students who feel undervalued and "marked" by stigma have fewer defenses against failure. Therefore, even if students do not fail more frequently than students in general, they are more likely to be psychologically devastated by such failures, leading them to "disidentify" with school and achievement. Although Steele's (1992) analysis focused primarily on the plight of African-American students, he speculated that his observations

might also be applicable to girls and students from lower-class social backgrounds.

Although negative teacher expectations are not identical to failure, we speculated that such expectations could readily produce effects analogous to those associated with failure. That is, if students must bear the brunt of inappropriately low teacher expectations, and if they belong to a stigmatized group, their enhanced vulnerability to negative school events may render them more susceptible to self-fulfilling prophecies. Social class and sex (at least in math and science classes) may be at least somewhat similarly stigmatizing. Poor students are frequently seen as inferior to their middle-class peers (Dusek & Joseph, 1983; Rist, 1970), and girls are often viewed as less skilled at math than are boys (Eccles & Hoffman, 1984; Eccles & Jacobs, 1986; Jacobs & Eccles, 1992; Meece et al., 1982; Yee & Eccles, 1988). It is widely believed in our culture that females have less ability in math than males (see Eccles, et al., 1983; Jacobs & Eccles, 1985, 1992). To the extent that females themselves have incorporated this stereotype into their own view of mathematics, they may be especially vulnerable to any behavioral indicators from others that are consistent with the stereotype. For example, Jacobs and Eccles (1985) found that mothers were more likely than fathers to lower their view of their daughters' math ability after being exposed to a media campaign reporting innate gender differences in math ability. These results suggested that females (in this case mothers) are more personally influenced than males by messages consistent with gender–role stereotypes.

There is another very different reason to suspect that students from stigmatized groups may be more strongly affected by teacher expectations. Social psychology has a long history of research suggesting that stereotypes of stigmatized groups are often inaccurate (e.g., G. Allport, 1954; Hamilton et al., 1990; Jones, 1986, 1990; Miller & Turnbull, 1986). By definition, the more inaccurate an expectation, the greater its potential to create self-fulfilling prophecies. Therefore, because students from stigmatized groups are perhaps more likely to be viewed inaccurately, they may be more strongly influenced by teachers' expectations.

B. CONCEPTUAL MODEL

Figure 2 presents the conceptual model underlying the following research. The model assumes that student backgrounds (previous grades and test scores, motivation, self-concept, etc.) influence both teacher perceptions and students' future performance outcomes. The model further assumes that teacher perceptions may also influence student performance outcomes;
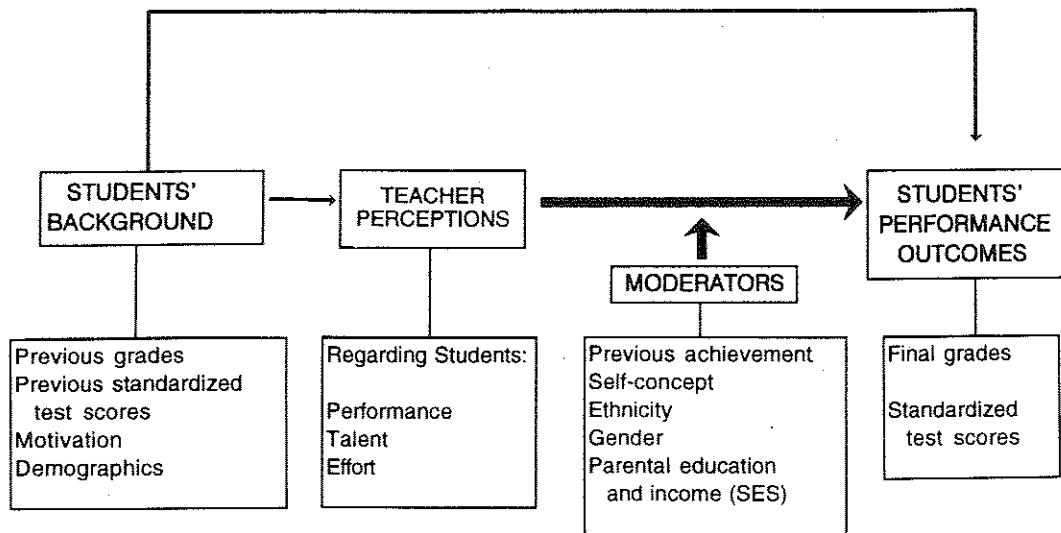
Fig. 2. Conceptual model of relationships between teacher perceptions and student achievement.

this is captured by the thick, horizontal arrow. Conceptually, this arrow represents self-fulfilling prophecies. The thick, vertical arrow represents the idea that various proposed moderators may increase or decrease the self-fulfilling influence of teacher expectations on student achievement.

The short thin arrow represents the possible influence of various aspects of student backgrounds on teacher perceptions. The long thin arrow represents the controls we have included in assessing relationships between teacher perceptions and students' future performance. The relationships represented by these thin arrows are not discussed in this section (although those relationships are precisely the focus of another series of studies reported later).

C. DATA ANALYTIC STRATEGY

Three separate sets of models were estimated: One set focused on student sex, a second on student social class, and a third on student ethnicity. Our analyses first assessed a baseline model, which assumed that the control variables (students' fifth-grade final math grades, scores on standardized tests taken in fifth or early sixth grade, self-concept of math ability, effort spent on math, time spent on math homework, and intrinsic and extrinsic value placed on math) and the three teacher perception variables (performance, talent, and effort) predict sixth-grade final grades and seventh-grade

MEAP scores (see Eccles, 1988; Jussim, 1989; Jussim & Eccles, 1992, for more information about these variables). The results for these baseline models are similar to those summarized earlier from Jussim (1989) and Jussim and Eccles (1992) because these analyses are based on students in the same sample.

Next, we assessed the moderation hypotheses. Specifically, we estimated a new model that added three product terms to the original model, terms representing the product of the student demographic category with each of the three teacher perception variables (performance, talent, and effort). The hypothesis that a particular student demographic characteristic moderated teacher expectation effects could be confirmed if either the block of three product terms or any of the individual predictors significantly predicted achievement outcomes. However, since the three product terms were highly correlated with each other,[2] testing all three simultaneously could artificially reduce the size and significance of all the product terms (e.g., Gordon, 1968), thereby substantially underestimating the role of any one moderator.[3] Consequently, if at least one of the predictors or the block of three moderators significantly increased the R^2 at $p \leq .10$ level, we examined the individual moderators in three steps.

In step one, we examined a model that added only the product term that most strongly predicted the outcome to the base model. This product term significantly predicted the outcome in each of the analyses that we performed. In step two, we added the other two product terms to the model. If this yielded no significant results (neither the R^2 increment, nor the individual coefficients were significant), the final model included the base model plus the first significant product term. However, if step two yielded significant results (either the R^2 increment or one of the individual coefficients were significant), we included a third step that essentially repeated step one. In step three, we added the stronger of the remaining two product terms to the final model. In this case, the final model included only the base model plus the two significant product terms. No models ever produced three significant product terms. These procedures reduced underestimation of moderator effects due to collinearity among the product terms. Finally, in order to fully explicate the significant moderator effects, we plotted the predicted relations separately for the two different demographic groups in each analysis (see, e.g., Judd & McClelland, 1989).

[2] The three product terms were correlated with each other for two reasons: 1) each teacher perception variable was multiplied by the same potential moderator, and 2) the teacher perception variables themselves were moderately to highly intercorrelated (approximately .5 to .8).

[3] This procedure is "artificial" because it estimates each individual coefficient after controlling for *all* other variables in the model. This includes not only the real control variables but the other teaacher perception-moderator product terms as well—in essence, potentially controlling "out" much of the very moderational relationship we are attempting to assess.

Except where noted, the N's for these analyses were 1765 for sex, 1020 to 1060 for social class, and 1609 to 1663 for ethnicity. Variations in sample size reflected different patterns of missing data, primarily with regard to family income and parent education. All analyses reported below were multiple regressions in which the student was the unit of analyses. Because teachers rated all of the students in their classrooms, teacher perceptions are not independent of one another. However, all analyses included classrooms (coded as dummy variables) as predictors of final grades in sixth-grade math and MEAP scores, thereby rendering all other relationships independent of teachers.

D. STUDENT SEX

Were girls more susceptible to self-fulfilling prophecies than were boys? Although student sex did not significantly interact with teacher perceptions to predict MEAP scores, the interaction did significantly predict final sixth-grade marks. The block of three interaction terms predicted final marks within our significance criterion ($F_{(3,1637)} = 2.32$, $p<.08$). The only statistically significant product term was Talent*Sex (Talent refers to teacher perceptions of talent). We then reestimated the model using only the one Talent*Sex product term, which significantly ($p < .05$) predicted grades. The results from this final analysis for predicting grades are presented in Table II.

The regression analysis yielded the following simplified prediction equation:

$$\text{grades} = 10.58 + .23(\text{Talent}) + .39(\text{Sex}) - .11(\text{Talent*Sex})$$

We refer to this as a "simplified" prediction equation because it contains *only* those coefficients and variables directly relevant to understanding how student sex moderates teacher expectation effects.[4]

[4] In creating this simplified prediction equation, we have assumed that all other variables (i.e., other than those in the simplified prediction equation) are at their mean. With this assumption, all other variables yield a constant. The constant in the simplified prediction equation equals the constant from the full regression equation plus the product of each variable's coefficient and its mean (each variable *except* for Talent, Sex, and the Talent*Sex terms, which are shown explicitly in the simplified prediction equation). Consider the following oversimplified example:

If grades = $1 + .1$(Standardized test scores) + .5(Previous grades) + 2(Talent) + 2(Sex) + .5(Talent*Sex), if the mean standardized test score is 50 (50th percentile), and if the mean grade is 10 (translating letter grades to a numerical scale), then the simplified prediction equation becomes:

$$\text{grades} = 11 + 2(\text{Talent}) + 2(\text{Sex}) + .5(\text{Talent*Sex})$$

The new constant of 11 = (the original constant) + .1(standardized test score mean) + .5(previous grades mean) = $1 + .1(50) + .5(10) = 11$.

TABLE II

Descriptive Statistics and Regression Results from Student Sex Moderation Analysis: Final Model Predicting Grades

Dependent variable	Mean	SD	Total R^2	F(1)	df(1)	Model R^2	F(2)	df(2)
Final marks (sixth grade)[1]	11.47	2.60	.75	40.00****	125,1639	.49	269.96****	12,1639

Predictors	Mean	SD			b		t	beta
Talent*Sex	—	—			-.11		2.30*	-.14
Teacher perceptions of								
Performance[2]	3.42	1.13			.50		8.08****	.22
Talent[3]	4.86	1.33			.23		2.68****	.12
Effort[3]	5.09	1.39			.28		7.29****	.15
Student background								
Sex[4]	1.47	.50			.39		1.57	.08
Standardized test scores[5]	61.06	25.67			.02		9.42****	.19
Fifth-grade final grades[1]	11.43	2.49			.29		15.07****	.28
Self-concept of ability[6]	10.15	2.46			.08		4.43****	.07
Effort at math[3]	5.73	1.31			.06		2.13*	.03
Time spent on homework[7]	2.46	.87			-.06		1.54	-.02
Intrinsic value of math[6]	9.47	3.44			.03		2.14*	.03
Extrinsic value of math[8]	18.36	3.12			.00		.06	.00

$*p < .05.$
$****p < .0001.$

b = unstandardized regression coefficients. *beta* = standardized regression coefficient. Total R^2, F(1), and df(1) refer to the full analysis including all dummy variables controlling for individual classrooms. Model R^2, F(2) and df(2) refer to the R^2 increment when the predictor variables shown in this table are added to a model that included only the classroom dummy variables.

[1] Grades were coded such that 16 represented an A+, 15 represented an A, 14 represented an A−, etc., down to a 3, which represented an F. A mean of 11.47, therefore, is between a B− and a B. [2] 1 to 5 scale. [3] 1 to 7 scale. [4] Girls = 1, boys = 2. [5] Percentile ranks based on national norms. [6] 2 to 14 scale. [7] 1 to 4 scale. [8] 3 to 21 scale.

In our data, girls were coded as "1" and boys as "2." Therefore, the simplified prediction equation showed that the equations relating teacher perceptions of talent to grades for girls and boys were:

$$\text{girls: Grades} = 10.97 + .12(\text{Talent})$$

$$\text{boys: Grades} = 11.36 + .01(\text{Talent})$$

These equations were obtained by simply entering the values for student sex into the simplified prediction equations (e.g., Judd & McClelland, 1989). For example, for girls:

$$10.58 + .39(\text{sex}) + .23(\text{Talent}) - .11(\text{Talent}*\text{Sex})$$
$$= 10.58 + .39(1) + .23(\text{Talent}) - .11(\text{Talent}*1)$$
$$= 10.58 + .39 + .12(\text{Talent})$$
$$= 10.97 + .12(\text{Talent}).$$

This equation shows that the slope for girls (.12) is steeper than the slope for boys (.01) (and the test of the product term has already shown that this difference is statistically significant). Figure 3 displays the relationship between teacher perceptions and grades separately for boys and girls. It clearly shows that boys' grades are virtually unaffected by teacher perceptions of talent, whereas girls' grades are affected. Even for girls, though, Figure 3 shows the effect to be quite small. The whole range of teacher
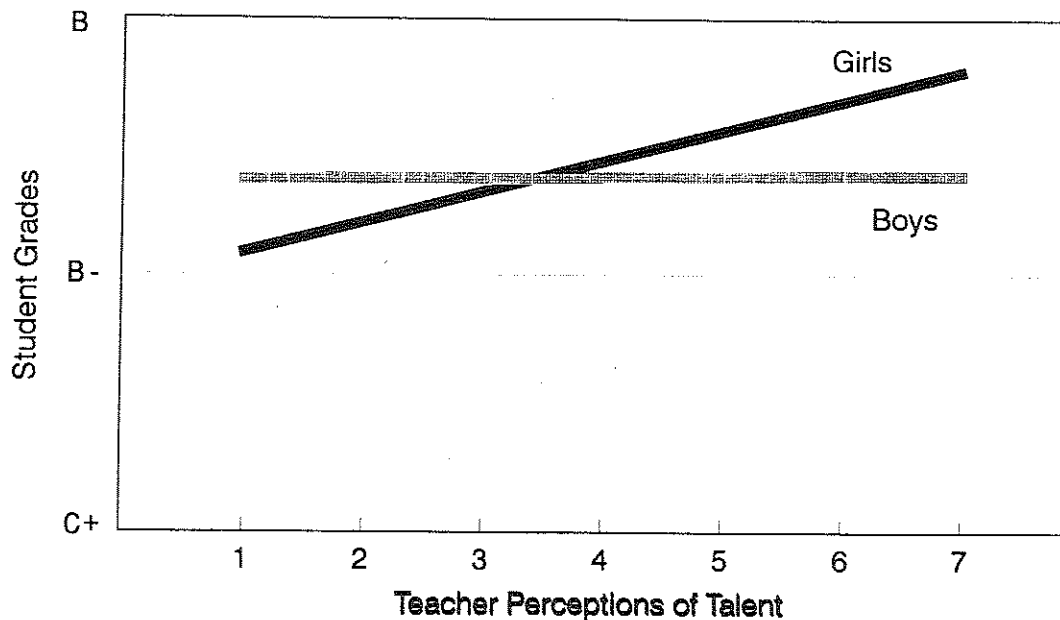

Fig. 3. Teacher expectations influence girls' grades more strongly than they influence boys' grades.

perceptions is barely enough to make a difference of one unit in student grades (e.g., B to B+).

Results reported thus far have been unstandardized. Although unstandardized coefficients are often preferable to the standardized coefficients (see Judd & McClelland, 1989; Pedhazur, 1982), standardized coefficients have one major advantage: They render results from different studies comparable. For example, as discussed previously, effect sizes (in terms of correlation coefficients or standardized regression coefficients) of teacher expectations on student achievement are typically .1 to .3. How do the separate coefficients for boys and girls compare to this general pattern? The standardized coefficient relating talent to grades for girls was .06, and for boys was .01. Thus, even for girls, for whom the effect is strongest, it is still quite small.

E. SOCIAL CLASS

Were students from lower SES backgrounds more vulnerable to self-fulfilling prophecies? Because there were two measures of social class (family income and education[5]), we made one modification to our data analytic approach. Instead of three product terms, six product terms were added to the equations predicting MEAP scores and final grades. Three product terms were created by multiplying each of the three teacher perception variables by mothers' education; three more terms were created by multiplying each of the three teacher perception variables by family income.

Although adding these six terms led to a significant R^2 increment ($F_{(6,1020)} = 3.57$, $p < .01$) in the prediction of MEAP scores, none of the coefficients relating the individual product terms to MEAP scores were statistically significant. This reflects collinearity among the product terms. Consequently, we used procedures like those described earlier to identify the most parsimonious subset of the significant predictive product terms: In this case, the product terms for teacher perceptions of effort*income and teacher perceptions of performance*education, each predicted MEAP scores (p's $< .05$). The results from this analysis are displayed in Table III. As with our student sex analyses, we obtained a simplified prediction equation by setting all variables that were not involved in the significant product terms at their mean. This yielded the following simplified prediction equation:

---

[5] Income was coded: 1 = less than $10,000/yr; 2 = $10,000–20,000/yr; 3 = $20,000–30,000/yr; 4 = $30,000–40,000/yr; 5 = more than $40,000/yr. Education ranged from 1 (never attended high school) through 9 (doctorate—MD, PhD, etc.). For 98% of the students, this was their mothers' education.

TABLE III

DESCRIPTIVE STATISTICS AND REGRESSION RESULTS FROM STUDENT SOCIAL CLASS MODERATION ANALYSIS: FINAL MODEL PREDICTING MEAP SCORES

Dependent variable	Mean	SD	Total R^2	$F(1)$	$df(1)$	Model R^2	$F(2)$	$df(2)$
MEAP scores[1]	23.45	4.39	.65	13.40****	126, 894	.38	70.48****	14, 1020

Predictors	Mean	SD	b	t	beta
Effort*income	—	—	-.13	2.16*	-.25
Performance*education	—	—	-.22	3.46***	-.39
Teacher perceptions of					
Performance[2]	3.55	1.10	1.43	4.70****	.36
Talent[3]	5.02	1.29	.19	1.40	.06
Effort[3]	5.22	1.36	.30	1.23	.09
Student background					
Education[4]	3.99	1.47	.91	3.51***	.31
Income[4]	3.66	1.23	.72	2.30*	.20
Standardized test scores[5]	64.58	24.93	.08	13.72****	.47
Fifth-grade final grades[1]	11.76	2.36	.26	4.70****	.14
Self-concept of ability[6]	10.33	2.38	.06	1.15	.03
Effort at math[3]	5.76	1.28	.01	.09	.00
Time spent on homework[7]	2.42	.86	-.15	1.30	-.03
Intrinsic value of math[6]	9.56	3.33	.04	1.16	.03
Extrinsic value of math[8]	18.47	3.07	.01	.26	.01

*$p < .05$.
***$p < .001$.
****$p < .0001$.

b = unstandardized regression coefficients. *beta* = standardized regression coefficient. Total R^2, $F(1)$, and $df(1)$ refer to the full analysis including all dummy variables controlling for individual classrooms. Model R^2, $F(2)$ and $df(2)$ refer to the R^2 increment when the predictor variables shown in this table are added to a model that included only the classroom dummy variables.

[1]MEAP scores ranged from 3 to 28 (a score of 0 is theoretically possible). [2]1 to 5 scale. [3]1 to 7 scale. [4]Education was on a 1 to 9 scale; income was on a 1 to 5 scale (see text for more details). [5]Percentile ranks based on national norms. [6]2 to 14 scale. [7]1 to 4 scale. [8]3 to 21 scale.

$$\text{MEAP} = 16.14 + .30(\text{Effort}) + 1.43(\text{Performance})$$
$$+ .91(\text{education}) + .72(\text{income})$$
$$- .13(\text{Effort*income}) - .22(\text{Performance*education}),$$

where Effort refers to teacher perceptions of students' effort and Performance refers to teacher percepetions of students' performance.

Among students whose parents had a lower education (some high school, coded as 2), the unstandardized relationship of teacher perceptions of performance to MEAP scores was .99 (.25, standardized). Among students whose parents had a higher education (having completed college, coded as 6), the unstandardized relation was .11 (.03, standardized). Among students from lower income families (family income of $10,000–$20,000/yr, coded as 2), the unstandaridzed relationship of teacher perceptions of effort to MEAP scores was .04 (.01, standardized). Among students from higher income families (greater than $40,000/yr, coded as 5), the unstandardized relationship of teacher perceptions of effort to MEAP scores was actually −.48 (−.15 standardized).

Figure 4 depicts the relationships of teacher perceptions of performance and effort to MEAP scores separately for students from lower and higher
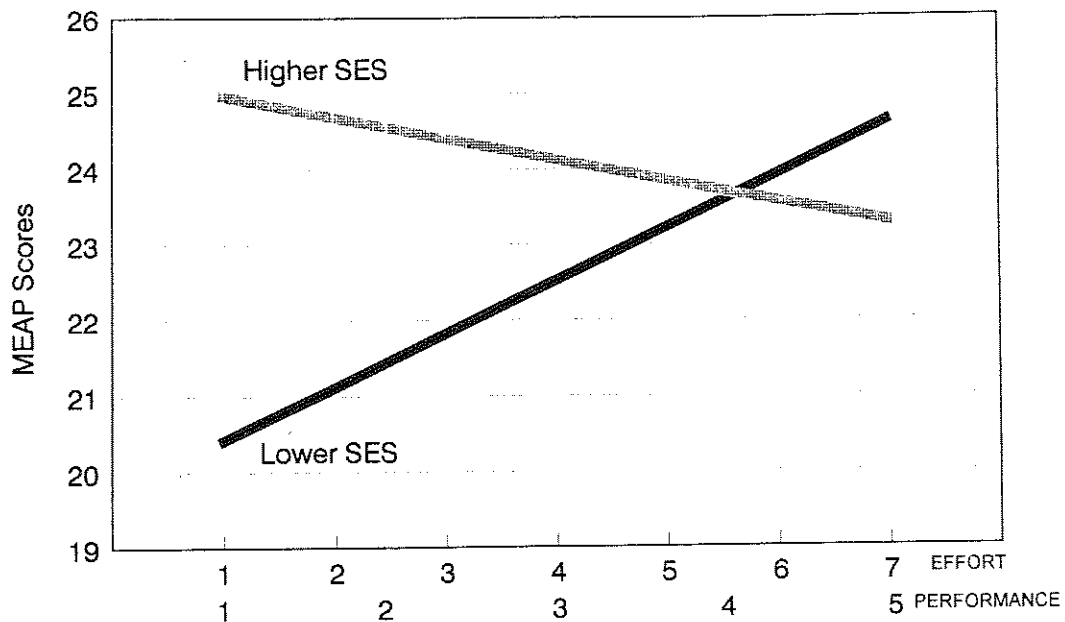

Fig. 4. Teacher expectations influence the standardized test scores of students from lower SES backgrounds more strongly than they influence the standardized test scores of students from higher SES backgrounds. Effort refers to teacher perceptions of effort. Performance refers to teacher perceptions of performance. Because product terms for teacher perceptions of performance with education and teacher perceptions of effort with income both predicted MEAP scores, both are shown here. Lower SES refers to an income of $10,000–20,000 and education of some high school. Higher SES refers to an income of greater than $40,000/yr. and having completed a college B.A.

SES backgrounds. Students from lower social class backgrounds were dramatically more vulnerable to self-fulfilling prophecies than were their more well-off classmates. As shown in Figure 4, the entire range of teacher perceptions makes about a 4-point difference on the MEAP. In terms of sample percentiles, going from 18 to 22 on the MEAP means going from the 17th percentile to the 38th; going from 21 to 25 means going from the 31st percentile to the 66th; and going from 24 to 28 means going from the 55th percentile to the 99th.

The negative relationship of teacher perceptions of effort to MEAP scores for upper class students completely accounts for the declining slope in Figure 4. Although unusual, attribution theory does provide one possible reason for why this negative relationship emerged among high SES students. Performance is often assumed to be influenced in a compensatory way by both effort and ability. If one has high ability, then less effort is needed to achieve the same performance level as that for someone with less ability (Covington & Omelich, 1979). This compensatory relationship may suggest the reason why attributing one's child's or one's students' successes to diligent effort might lead to both lowered ability self-concepts in the child or student and lowered perceptions of one's child's or students' abilities (Yee & Eccles, 1988). To the extent that a teacher is rating a high SES student as working very hard in his or her class, the teacher may also be conveying the indirect message that the student has lower ability. This message could then undermine the student's motivation or increase the student's anxiety such that the student performs more poorly in a standardized testing situation.

Alternatively, the teacher's view that the student is working hard may really mean that student does work harder than other students to compensate for lower ability. If so, then this lower ability level could explain why the student does not do as well as his or her peers in a timed standardized testing situation, in which there is insufficient time for greater effort to compensate for lower ability in determining final performance level.

A similar pattern was obtained for final grades. Although the R^2 increment associated with adding all six product terms approached significance ($F_{(6,926)} = 1.71$, $p = .12$) only when entered alone, the teacher perceptions of performance by income product term did significantly predict grades ($p < .01$). The final model, then, included only this one product term. These results are summarized in Table IV. This analysis yielded the following simplified prediction equation:

$$\text{Grades} = 8.54 + .88(\text{Performance}) + .33(\text{income}) - .09\ (\text{Performance*income}),$$

where Performance refers to teacher perceptions of performance.

TABLE IV

Descriptive Statistics and Regression Results from Student Social Class Moderation Analysis: Final Model Predicting Grades

Dependent variable	Mean	SD	Total R^2	F(1)	df(1)	Model R^2	F(2)	df(2)
Sixth grade Final grades[1]	11.72	2.53	.77	24.88****	125, 931	.49	153.77****	13, 931

Predictors	Mean	SD	b	t	*beta*
Performance*income	—		-.09	2.74**	-.14
Teacher perceptions of					
Performance[2]	3.52	1.11	.88	6.29****	.22
Talent[3]	4.99	1.31	.05	.75	.03
Effort[3]	5.20	1.37	.27	5.36****	.15
Student background					
Education[4]	3.96	1.47	-.02	.52	-.01
Income[4]	3.65	1.23	.33	2.90**	.16
Standardized test scores[5]	63.85	25.39	.02	6.97****	.19
Fifth-grade final grades[1]	11.69	2.41	.30	11.76****	.29
Self-concept of ability[6]	10.30	2.38	.06	2.83**	.06
Effort at math[3]	5.76	1.28	.02	.61	.01
Time spent on homework[7]	2.42	.86	-.03	.48	-.01
Intrinsic value of math[6]	9.55	3.34	.04	2.34*	.05
Extrinsic value of math[8]	18.45	3.06	.00	.25	.00

*p < .05.
**p < .01.
****p < .0001.

b = unstandardized regression coefficients. *beta* = standardized regression coefficient. Total R^2, F(1), and df(1) refer to the full analysis including all dummy variables controlling for individual classrooms. Model R^2, F(2) and df(2) refer to the R^2 increment when the predictor variables shown in this table are added to a model that included only the classroom dummy variables.

[1] Grades were coded such that 16 represented an A+, 15 represented an A, 14 represented an A−, etc., down to a 3, which represented an F. A mean of 11.47, therefore, is between a B− and a B. [2]1 to 5 scale. [3]1 to 7 scale. [4]Education was on a 1 to 9 scale; income was on a 1 to 5 scale (see text for more details). [5]Percentile ranks based on national norms. [6]2 to 14 scale. [7]1 to 4 scale. [8]3 to 21 scale.
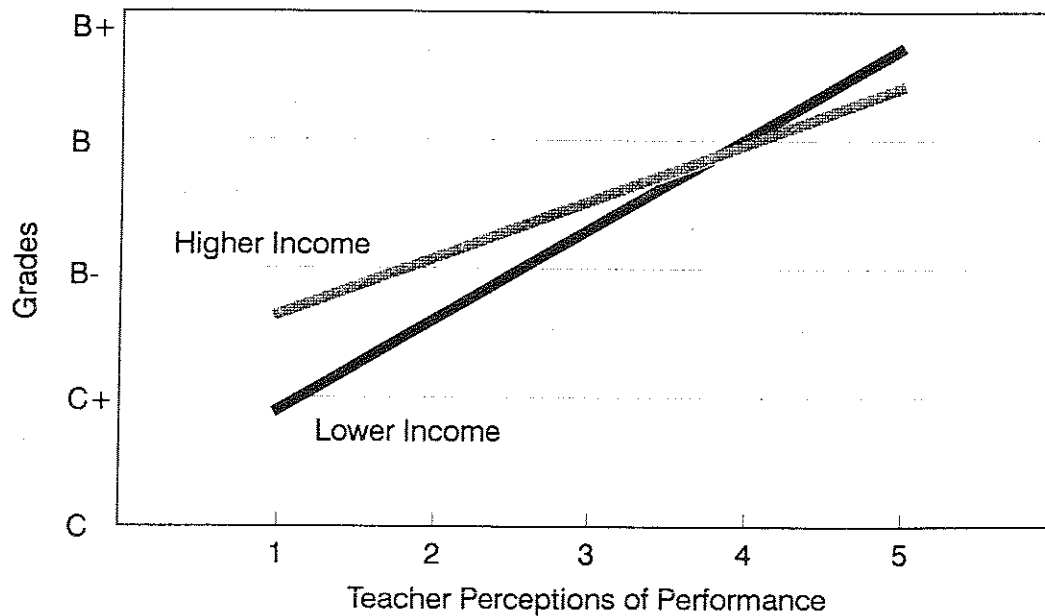
Fig. 5.   Teacher expectations influence the grades of students' from lower SES backgrounds more strongly than they influence the grades of students from higher SES backgrounds. Lower income refers to an income of $10,000–20,000. Higher income refers to an income of greater than $40,000/yr.

The relationship of teacher perceptions of performance to grades for students from low income backgrounds (family income of $10,000–$20,000/ yr) was .88 (unstandardized) and .31 (standardized). The relationship of teacher perceptions of performance to grades for students from higher income backgrounds was .43 (unstandardized) and .19 (standardized). These results, displayed in Figure 5, clearly show that teacher perceptions influenced the grades of lower income students more strongly than they influenced the grades of upper income students. Comparing these results to those displayed in Figure 4, however, shows that the difference is less dramatic for grades than for MEAP scores. Across the entire range of grades, the teacher expectation effect made a difference of about 2 grade levels[6] for upper income students (e.g., C+ to B) and three grade levels for lower income students (e.g., C+ to B+).

F. ETHNICITY

Were African-American students more susceptible to self-fulfilling prophecies than White students? To answer this question, three ethnic-

---

[6] Grades were coded on a scale going from 3 (F) to 16 (A+), with each unit representing the next grade level: grade 4 was F+; grade 5 was D−; grade 6 was D; grade 7 was D+, and so on.

ity by teacher perception product terms were added to the equations predicting MEAP scores and final grades. The analysis on MEAP scores was based on the 1536 White students and 72 African-American students who had valid data on all variables. Results are summarized in Table V.

The final analysis yielded a significant ethnicity by teacher perception of performance product term ($\beta = -.14$; $t = 2.79$, $p < .01$).[7] The simplified prediction equation was:

$$\text{MEAP} = 24.18 - 3.24(\text{ethnicity}) - .35(\text{Performance}) + .94(\text{Performance}*\text{ethnicity})$$

Ethnicity was coded as 1 for White students and 2 for African-American students. The relationship of teacher perceptions of performance to MEAP scores was .59 (unstandardized) and .14 (standardized) for White students and 1.53 (unstandardized) and .37 (standardized) for African-American students. Figure 6 displays these relations and clearly shows the dramatically greater expectancy effects for African-American students. For White students, the entire range of teacher perceptions of performance makes about a 2.5-point difference score on the MEAP, whereas for African-American students, that range makes about a 6-point difference.

To make this finding more concrete, consider a two- or three-point difference in MEAP scores (the magnitude of the largest changes associated with teacher perceptions among White students). Scores going from 19 to 22 would mean going from the 21st to the 38th percentile, and going from 24 to 26 would mean going from the 55th to the 78th percentile. Now consider a six-point difference (the magnitude of the largest possible changes associated with teacher perceptions among African-American students). Going from 17 to 23 would mean going from the 14th to the 45th percentile, and going from 21 to 27 would mean going from the 31st to the 89th percentile.

Analyses examining predictors of final grades, which included 76 African-American students and 1587 White students, showed a similar pattern. Adding the three product terms significantly increased the R^2 increment ($F(3,1535) = 6.91$, $p < .0001$). In the final model, however, only the ethnic-

[7] If added to this model, the ethnicity by teacher perceptions of effort product term was also statistically significant ($p < .05$). However, 1) its coefficient was negative; 2) the unstandardized coefficient for the ethnicity by performance interaction doubled, going from about .9 in the analysis reported in the main text to 1.8; and 3) the coefficient for ethnicity by effort was $-.9$. Therefore, the net effect of teacher perceptions on achievement was still highly positive, and almost identical to the effect reported in the main text. Because the analysis with only one product term is simpler to present and interpret, and provides essentially the same information as the analysis with two product terms, the main text reports the analysis with the one product term.

TABLE V

Descriptive Statistics and Regression Results from Student Ethnicity Moderation Analysis: Final Model Predicting MEAP

Dependent variable	Mean	SD	Total R^2	$F(1)$	$df(1)$	Model R^2	$F(2)$	$df(2)$
MEAP scores[1]	23.00	4.63	.77	17.63****	125, 1483	.36	112.10****	12, 1483

Predictors	Mean	SD	b	t	beta
Performance*ethnicity	—		.94	2.79**	−.14
Teacher perceptions of					
Performance[2]	3.48	1.11	−.35	.90	−.08
Talent[3]	4.91	1.31	.00	.04	.00
Effort[3]	5.13	1.36	−.04	.40	−.01
Student background					
Ethnicity[4]	3.96	.21	−3.24	2.31*	−.14
Standardized test scores[5]	1.04	25.23	.09	18.03****	.49
Fifth-grade final grades[1]	11.53	2.44	.23	5.07****	.12
Self-concept of ability[6]	10.20	2.44	−.01	.19	.00
Effort at math[3]	5.73	1.30	.00	.06	.00
Time spent on homework[7]	2.44	.87	−.04	.41	−.01
Intrinsic value of math[6]	9.50	3.43	.06	2.14*	.05
Extrinsic value of math[8]	18.44	3.04	.04	1.36	.03

*$p < .05$.
**$p < .01$.
****$p < .0001$.

b = unstandardized regression coefficient. Total R^2, $F(1)$, and $df(1)$ refer to the full analysis including all dummy variables controlling for individual classrooms. Model R^2, $F(2)$ and $df(2)$ refer to the R^2 increment when the predictor variables shown in this table are added to a model that included only the classroom dummy variables.

[1]MEAP scores ranged from 3 to 28 (a score of 0 is theoretically possible). [2]1 to 5 scale. [3]1 to 7 scale. [4]Whites = 1, African-Americans = 2. [5]Percentile ranks based on national norms. [6]2 to 14 scale. [7]1 to 4 scale. [8]3 to 21 scale.
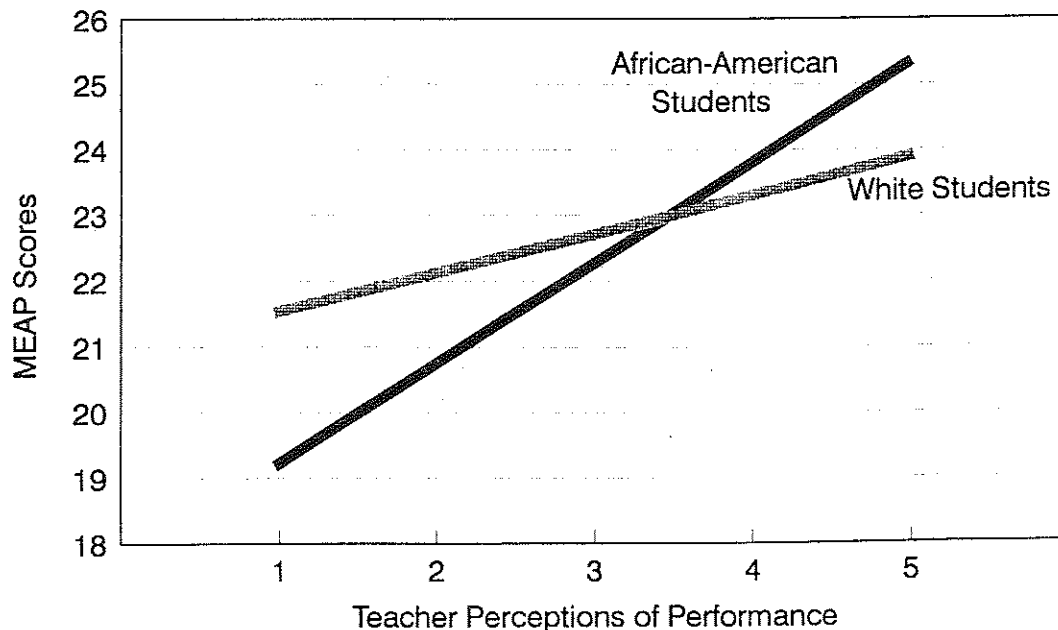
Fig. 6.   Teacher expectations influence the standardized test scores of African-American students more strongly than they influence the standardized test scores of White students.

ity by teacher perceptions of performance product term was significant (adding the other two terms, individually or together, yielded no additional significant individual coefficients and no significant R^2 increments). These results are summarized in Table VI. The simplified prediction equation was:

$$\text{Grades} = 12.05 - .18(\text{Performance}) - 2.11(\text{ethnicity}) + .64(\text{Performance*ethnicity}).$$

The results (displayed in Figure 7), are consistent with the prediction that teacher expectation effects are stronger among African-American students than among White students. For White students, the relationship of teacher perceptions of performance to final grades was .46 (unstandardized) and .20 (standardized), and for African-American students, the relationship as 1.28 (unstandardized) and .56 (standardized). Figure 7 shows that, all other variables being held equal, going from the lowest to the highest teacher perceptions predicted a 4-unit change in grades (e.g., going from C to B+) among African-American students, but only a 2-unit change in grades (e.g., going from C+ to B) among White students.

We were concerned about two limitations to our ethnicity study. First, the sample of African-American students was quite small. Because attrition came mainly from students moving in and out of the district during the 3-year span covered by the analyses (fifth through seventh grades), we were able to increase the sample size by omitting some presixth-grade data. We

TABLE VI
DESCRIPTIVE STATISTICS AND REGRESSION RESULTS FROM STUDENT ETHNICITY MODERATION ANALYSIS: FINAL MODEL PREDICTING GRADES

Dependent variable	Mean	SD	Total R^2	F(1)	df(1)	F(2)	df(2)
Final grades[1]	11.52	2.57	.87	37.06****	125, 1537	251.88****	12, 1537

Predictors	Mean	SD	Model R^2	t	b	beta
Performance*ethnicity	—	—	.49	4.51**	.64	.34
Teacher perceptions of						
Performance[2]	3.45	1.12		1.12	−.18	−.08
Talent[3]	4.89	1.33		.80	.04	.02
Effort[3]	5.12	1.37		7.65****	.30	.16
Student background						
Ethnicity[4]	1.05	.21		3.51***	−2.11	−.17
Standardized test scores[5]	61.51	25.55		9.23****	.02	.20
Fifth-grade final grades[1]	11.47	2.48		15.18****	.30	.29
Self-concept of ability[6]	10.17	2.45		4.49****	.08	.08
Effort at math[3]	5.73	1.30		1.26	.04	.02
Time spent on homework[7]	2.45	.87		.94	−.04	−.01
Intrinsic value of math[6]	9.48	3.44		2.18*	.03	.04
Extrinsic value of math[8]	18.41	3.04		.32	.00	.00

$*p < .05.$
$**p < .01.$
$***p < .001.$
$****p < .0001.$

b = unstandardized regression coefficients. *beta* = standardized regression coefficient. Total R^2, F(1), and df(1) refer to the full analysis including all dummy variables controlling for individual classrooms. Model R^2, F(2) and df(2) refer to the R^2 increment when the predictor variables shown in this table are added to a model that included only the classroom dummy variables.

[1] Grades were coded such that 16 represented an A+, 15 represented an A, 14 represented an A−, etc., down to a 3, which represented an F. A mean of 11.47, therefore, is between a B− and a B. [2] 1 to 5 scale. [3] 1 to 7 scale. [4] Whites = 1, African-Americans = 2. [5] Percentile ranks based on national norms. [6] 2 to 14 scale. [7] 1 to 4 scale. [8] 3 to 21 scale.
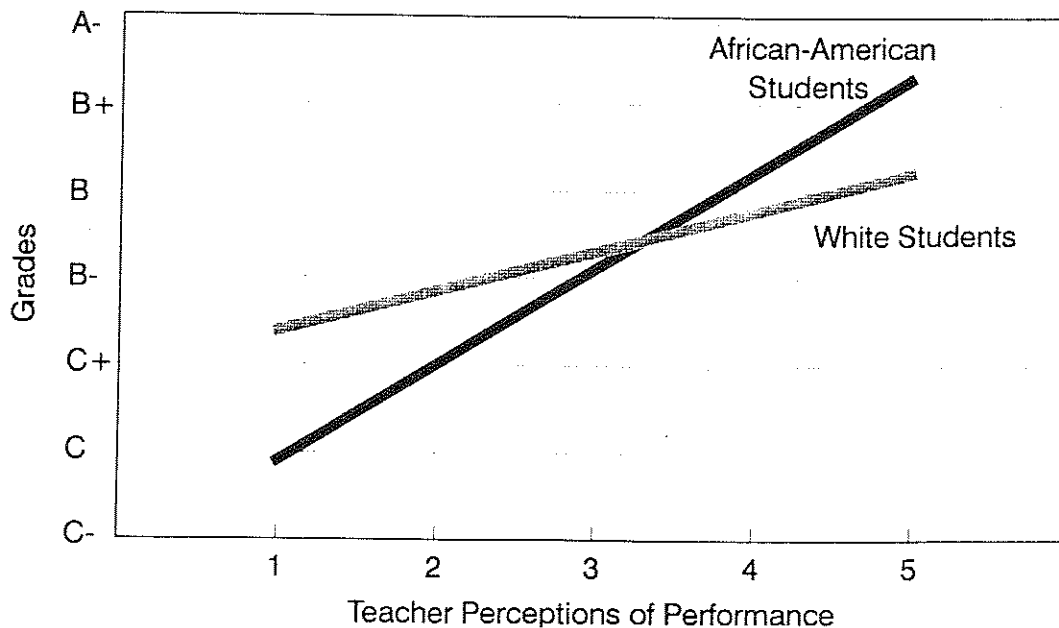
Fig. 7. Teacher expectations influence the grades of African-American students more strongly than they influence the standardized test scores of White students.

ran one set of additional models that did not require fifth-grade final grades, and another set that did not require previous standardized tests. For the analyses predicting seventh-grade MEAP scores and sixth-grade final grades, these analyses increased the African-American Ns from 72 and 76, to 90 and 121, respectively. The results were virtually identical: Stronger expectancy effects emerged among the African-American students than among the White students.

Second, we were concerned about ethnic heterogeneity in our school districts. Specifically, in one district 97% of the students were African-American; another district had a more even distribution (63% White, 34% African-American, and 3% other). The other 10 districts were predominantly White (95% on the average). We tested for district differences in the pattern of expectancy effects in several ways. Because the students in the ethnically mixed district did not take a standardized test in fifth or early sixth grade, they were excluded from the main analysis, but were included in the analysis that did not use standardized tests as a predictor. These analyses yielded results similar to those already reported.

We also ran the main set of models a second time (i.e., the models depicted in Tables II through VI) excluding the students in the predominantly African-American district. Again, the results were similar to those already reported. Last, we ran a set of models in which we coded the type of district (predominantly White or African-American) as a dummy

variable, created product terms representing the interaction of the district with teacher perceptions, and then ran the main models again. None of these interaction terms significantly predicted either student grades or MEAP scores. Therefore, we found no evidence that the patterns we observed varied greatly in districts with considerably different ethnic mixes.

G. UNCONFOUNDING THE EFFECTS OF ETHNICITY AND SOCIAL CLASS

A common problem in research on ethnicity and social class is assessing their separate roles in social phenomena. This is a problem because, on the average, Whites are from higher SES backgrounds than many minorities. However, this was not a major problem in the current study. Ethnicity was uncorrelated with parental education ($r = -.01$, ns), and only weakly related to income ($r = -.18, p < .001$). Nonetheless, several additional analyses further probed this issue. First, we reran the main social class analyses (see Tables III & IV) excluding all African-American students. The results were virtually identical to those reported in Tables III and IV, indicating that results regarding social class do not derive primarily from the African-American students.

Second, despite some important limitations, we ran an additional set of analyses adding both the ethnicity and SES product terms (together) to the models predicting grades and MEAP scores. One limitation is that, because family income and education information was available for only slightly more than half of the sample, the number of African-American students was reduced even further—to 27. Nonetheless, the results for the analysis predicting MEAP scores were striking: All three two-way product terms (involving education, income, and ethnicity) significantly ($p < .05$) predicted MEAP scores, and the regression coefficients were nearly identical to those reported in Tables III and IV.

Results for the model predicting grades were similar but weaker. Here, colinearity was a nuisance. Adding the performance by income and performance by ethnicity terms to the base model significantly increased the R^2 increment ($F(2, 897) = 4.35, p < .02$). However, neither product term, individually, reached statistical significance. The performance by income coefficient was similar to the term shown in Table IV ($-.09$ in the table versus $-.07$ in the current analysis, $p < .06$), but the ethnicity by performance coefficient was lower (.64 in the table versus .32 in the current analysis, $p < .12$).

These results, despite their limitations, were similar to those obtained in the analyses that addressed ethnicity and social class separately. It seems,

therefore, that the higher expectancy effects among lower SES students are not due to the greater proportion of African-American students from lower SES backgrounds. It is also unlikely that the greater expectancy effects among African-American students are due to their lower SES backgrounds.

H. MULTIPLE VULNERABILITIES

These analyses, however, raise a more general question: Are students who are members of more than one vulnerable group more susceptible to expectancy effects than other students? For example, are lower SES African-American students most susceptible? Are lower SES girls also doubly vulnerable? Are African-American girls and lower SES, African-American girls also especially vulnerable? The results just reported suggest that the greater vulnerability of lower SES students is largely independent of the greater vulnerability of African-American students. Put differently, they suggest that these effects are additive, which implies that, because SES and ethnicity are both powerful moderators, lower class African-American students would be the most vulnerable of all to teacher expectation effects.

Hypothetically, these questions could be answered by adding the requisite three-way product terms. Unfortunately, however, because African-American students with parental SES information are relatively few in number, the results from a model including three-way product terms combining ethnicity, SES, and teacher perceptions would not have been meaningful. Because we are presently unable to include these three-way product terms, direct assessment of this question must await future research.

We did, however, assess whether lower SES girls and African-American girls were particularly susceptible by assessing models including three-way product terms (between student sex, teacher perceptions, and either SES or ethnicity). Because these models are very complex (base model, plus two-way product terms for all combinations of sex by the three teacher perception variables by either the two SES variables or by ethnicity, plus three-way product terms) we can only summarize our analyses here.

There were four main analyses (Two outcomes by two separate tests of three-way interactions: teacher perceptions by student sex by student SES predicting grades and MEAP; and teacher perceptions by student sex by student ethnicity predicting grades and MEAP). Each analysis was performed in two steps. The first step included the base model, plus all lower-order two-way product terms. Previously obtained significant two-way product terms remained significant in all of these models. In the second step, we added the three-way product terms. For three of the four models, none of the three-way product terms significantly predicted student outcomes.

Even the fourth did not qualify our previous conclusions: It showed that, for MEAP scores, lower SES girls were slightly more susceptible to expectancy effects than were other groups.

These results mean that lower SES girls and African-American girls are more susceptible to expectancy effects than are students who belong to only one vulnerable group (i.e., the two interaction effects were generally additive). These results are consistent with our general conclusions that girls are slightly more susceptible than boys, that lower SES students are considerably more susceptible than upper SES students, and that being a lower SES or African-American girl is a double vulnerability.

V. Stereotypes and Self-Fulfilling Prophecies

Teacher stereotypes would appear to be a likely explanation for why expectancy effects are more powerful among students from stigmatized or disadvantaged social groups. Both the evidence regarding the self-fulfilling effects of social stereotypes and the limitations to that evidence are discussed next.

Many social psychological perspectives and reviews claim or assume that stereotypes are often inaccurate and could likely lead to self-fulfilling prophecies (Hamilton et al. 1990; Jacobs & Eccles, 1992; Miller & Turnbull, 1986; Snyder, 1984; von Hippel et al., 1995). The role of stereotypes in creating self-fulfilling prophecies that contribute to inequalities between ethnic and socio-economic groups and between the sexes may seem "obvious." The dominant group (White men) hold negative stereotypes about other groups. White men treat members of these groups less favorably than they treat other White men, so that members of these other groups receive lower quality education (and lower paying jobs, too).

Undoubtedly, this sequence sometimes occurs. Although the claim that social stereotypes are self-fulfilling appears straightforward, it is considerably more complex than it seems for several empirical and conceptual reasons. Although any stereotype, hypothetically, may be self-fulfilling, most research to date has focused on four particular stereotypes: ethnicity, social class, gender, and physical attractiveness. This research literature is reviewed next, after which we address basic theoretical issues involved in understanding the extent to which stereotypes create social injustices through self-fulfilling prophecies.

A. ETHNIC STEREOTYPES

We are aware of only one study that comes close to documenting self-fulfilling prophecies produced by an ethnic stereotype. In the first classic

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 conducted a shorter 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 (as were 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 examine 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 indeed prejudiced against African-Americans. Further, a recent study (Jussim, Nelson, Manis, & Soffin, 1995) found that prejudice toward (disliking of) 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 anxiety. 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 (Dusek & Joseph, 1983; Jussim, Coleman, & 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 three groups: students supposedly smart, average, or dumb. Each group sat at its own table (Tables A, B, and C, respectively). However, the main difference between the students 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 suppored 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 Tables 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 biasing 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 tear-jerker," 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 make-up and accessories); and, if they believed that he was attractive, they expressed more traditional gender-role attitudes on questionnaires and actually performed worse on an 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 (Skyrpnek & Snyder, 1982).

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; Palardy, 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, 1992). 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. Andersen and Bem (1981) had androgenous or sex-typed men and women perceivers interact with men and women targets. In contrast to the Snyder et al. (1977) study, Andersen 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, androgenous female perceivers created a "boomerang" effect: Unattractive targets interacting with them were actually rated more favorably than were the attractive targets! These two experiments (Andersen & 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., Goldman & Lewis, 1977; see a meta-analysis by Feingold, 1992). It seems likely that more socially skilled MBAs would deserve and actually receive higher salaries 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 Jussim & Eccles, 1995, 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, but 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 Jussim & Eccles, 1995; Jussim & 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 achievment 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 so 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, 1984; Stangor, 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 rarely 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 researchers 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 (Deparle, 1994; Kimball, 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, at least if 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 validity or invalidity of many beliefs about groups. Moreover, the little empirical evidence that does exist provides a decidedly mixed picture (e.g., Brigham, 1971; Judd & Park, 1993; Jussim, 1990; Jussim, 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., Judd & Park, 1993; McCauley, 1995). However, the accuracy of people's beliefs about the distribution of social groups 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, as broad sweeping generalizations, claims that stereotypes are both inaccurate and self-fulfilling are mutually exclusive.

Of course, stereotypes may be inaccurate originally, 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 undermining the motivation, skills, and performance of other groups. Oblivious to the social bases 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 perceivers themselves would be a very specious 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 at 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 sex-differentiated socialization 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 as 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., Devine, 1989; Greenwald & Banaji, 1995; Jones, 1986; Miller & Turnbull, 1986). Consequently, stereotypes are frequently accused of being the cognitive culprits in prejudice and discrimination (e.g., Fiske & Taylor, 1990; Hamilton et al., 1990; Stangor, 1995). 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 biases and discrimination is actually sparse, weak, and equivocal (see reviews by Jussim, 1990, 1991; Jussim et al., 1995; McCauley, 1995; McCauley, Stitt, & Segal, 1980; Oakes, Haslam, & Turner, 1994). Even the link between stereotypes and prejudice itself is often quite weak (Eagly & Mladinic, 1989; Haddock, Zanna, & Esses, 1993; McCauley & 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 stereo-types 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 & Gross, 1983; Fiske & Neuberg, 1990; Krueger & Rothbart, 1988; Linville, 1982; Locksley, Borgida, Brekke, & Hepburn, 1980). 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 intentionallly 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 may accompany identification of real differences. In contrast, naturalistic research 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 experimental studies of stereotypes and person perception. Such studies do show that perceivers sometimes see differences between individuals and differing groups when none exist. However, they are completely mute 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 education, 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., Beckett & Park, 1995; Darley & Gross, 1983; 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 in 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 (these 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 and accuracy 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 really 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 parlance, *individuating 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, 1991; Jussim et al., 1995; Mackie, 1973; Ottati & Lee, 1995; Wineburg, 1987) 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 *processes* involved in stereotyping. However, even if processes are high in experimental 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; Judd & Park, 1993; McCauley, Jussim, & Lee, 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 expectation effects stronger among girls, African-American students, and students from lower SES backgrounds? Classic social psychology suggests that stereotypes, 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 criterion 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 teacher–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, than 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 individuating 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, than their 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 criterion 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).[8]

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

---

[8] These analyses did not control for classroom for several reasons. The criteria, student grades, standardized test scores, and motivation, 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 need 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, motivation, 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 ($r = .02$).

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 sex difference in 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 exerting 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 achievement. 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 F's < 2.5, 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
HOMOGENEOUS DISTRICTS: WERE TEACHER PERCEPTIONS OF NO DIFFERENCES JUSTIFIED?

	White students in the White districts	African-American students in the White districts	African-American students in the African-American district
Standardized test scores (percentile ranks)	60 ($n = 2064$)	52 ($n = 39$)	38 ($n = 95$)
Fifth-grade final grades	B−/B ($n = 2040$)	C+/B− ($n = 27$)	C+ ($n = 61$)

In both analyses, the difference between African-American students in the predominantly African-American district and White students is statistically significant ($p < .001$).

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

(From Jussim, L., McCauley, C. R., & Lee Y. T. (1995). Why study stereotype accuracy and inaccuracy? In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences*. (pp. 3–27). Copyright © 1995 by the American Psychological Association. Adapted with permission.)

partially justified. In the predominantly White districts, neither the standardized test score differences nor the grade differences were statistically significant (all t'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 t'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 ($F_{(2,2383)} = 4.68$, $p < .01$) and self-concept of math ability ($F_{(2,2383)} = 4.58$, $p < .02$), and marginal significance for self-perceptions of effort ($F_{(2,2388)} = 2.81$, $p < .07$). 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 had in the predominantly White district ($t = 2.11$, $p < .05$). Furthermore, all of the *eta*s were below .07, indicating that although statistically significant, the differences were minor.

6. The Ethnically Mixed District

Did teachers perceive the 22 African-American students differently than they perceived the 40 White students in the ethnically mixed district? They

did. Teachers perceived White students as performing better ($r = -.27$, $p < .05$), as more talented ($r = -.26, p < .05$) and as trying harder, although this last difference did not reach statistical significance ($r = -.20, p = .12$).

Were these perceptions justified? The African-American students did have lower grades than White students in this district (C/C+ vs. B−, $r = -.21$). However, this difference was based on only 32 White students and 14 African-American students (not all students attended this district in fifth grade), and it was not statistically significant ($p = .16$). The difference was, however, of about the same magnitude as the differences teachers perceived. Unfortunately, no standardized test was given to fifth-graders in this district.

Did teacher perceptions of effort differences correspond to ethnic differences in the motivation variables? African-American and White students in this district all claimed to be exerting about the same amount of effort, spending about the same amount of time on homework; all reported similar self-concepts of ability (all r's $< .07$, all p's $> .6$). Thus, the nonsignificant ethnic difference that teachers perceived in effort corresponded reasonably well to the lack of difference in the student effort and motivation variables.

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. Iif they held no stereotype, 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; Dusek & Joseph, 1983; 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; Sedikedes & 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, 1974; 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 those 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 STEREOTYPES?

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 & Gross, 1983; Duncan, 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; Dar-
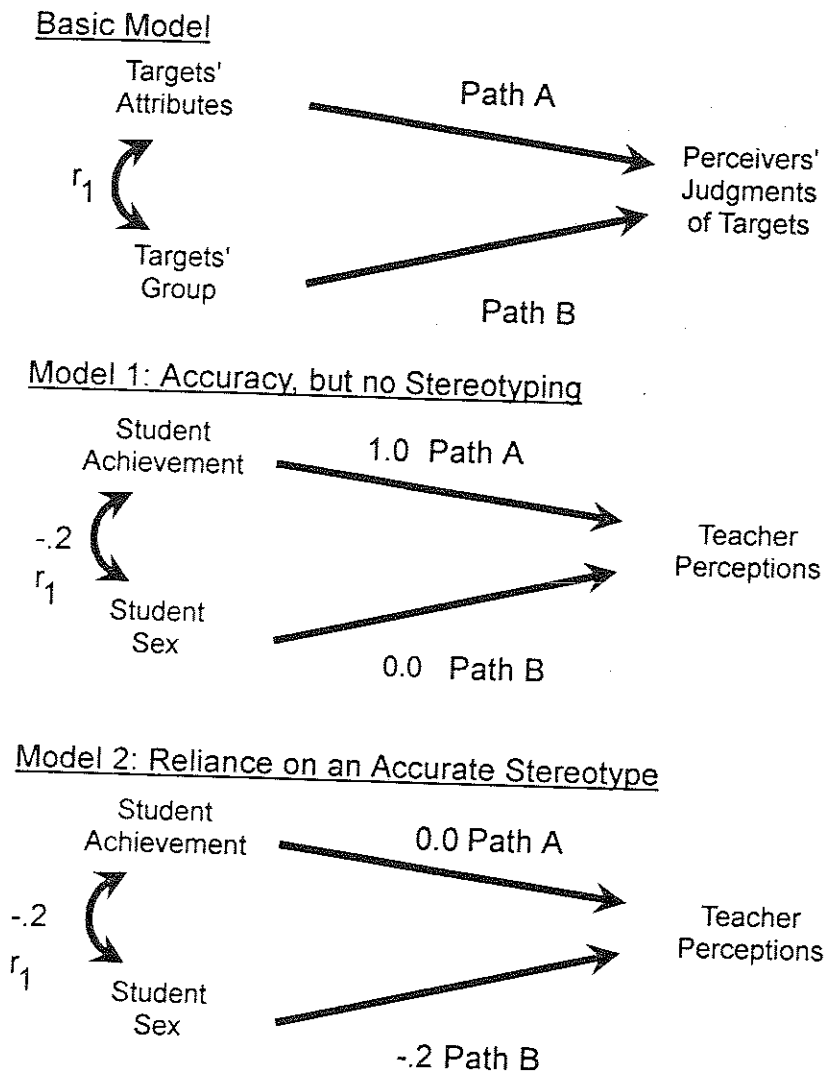
Basic Model

Targets'
Attributes

r_1

Targets'
Group

Path A

Path B

Perceivers'
Judgments
of Targets

Model 1: Accuracy, but no Stereotyping

Student
Achievement

-.2

r_1

Student
Sex

1.0 Path A

0.0 Path B

Teacher
Perceptions

Model 2: Reliance on an Accurate Stereotype

Student
Achievement

-.2

r_1

Student
Sex

0.0 Path A

-.2 Path B

Teacher
Perceptions

Fig. 8. Hypothetical models of accuracy in teacher perceptions of differences between boys and girls. Teacher perceptions of sex differences are accurate when the size of the perceived difference between boys and girls equals the actual difference between boys and girls. The perceived difference (r2, not shown) is the correlation between teacher perceptions and student sex, which equals Path B + (−.2 × Path A). The actual sex differences equals −.2 (the correlation between student sex and student achievement).

ley & Gross, 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 (individuating 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 (r_1); 2) the influence of targets' attributes (i.e., individuating information) on perceivers' judgments (Path A); 3) and 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 r_2 (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 $r_2 = r_1$.

In all experimental laboratory research on stereotypes and person perception of which we are aware, r_1 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 (Beckett & Park, 1995; Krueger & Rothbart, 1988). In such 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., $r_2 > 0$, even though $r_1 = 0$).

This model goes beyond the experimental laboratory research because it shows that the comparison of r_2 to r_1 is a more general index of bias. Even if there is a real difference between groups (i.e., $r_1 \neq 0$), the model shows that bias occurs whenever $r_2 > r_1$. 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; Jussim, 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

$$r_2 = \text{Path B} + r_1 \text{ (Path A)}.$$

"Judging solely on their attributes" means that Path B (the stereotype

effect) is zero, and that Path A is high. However, because r_1 (the real difference) is zero, r_2 (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., $r_1 > 0$). Even if there is no stereotype effect (Path B is zero), r_2 (the perceived difference) approaches r_1 (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 $= r_1$ (i.e., perceivers stereotype the groups as differing to the same extent that they actually differ), then $r_2 = r_1$ (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 difference 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 numerous 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 $r_2 =$ Path B $- .2$(Path A), where r_2 is the correlation of teacher perceptions with student sex.

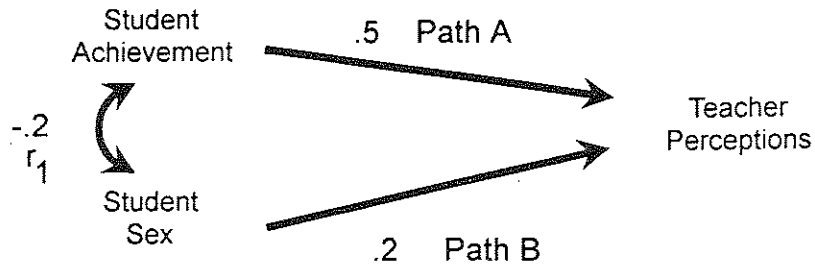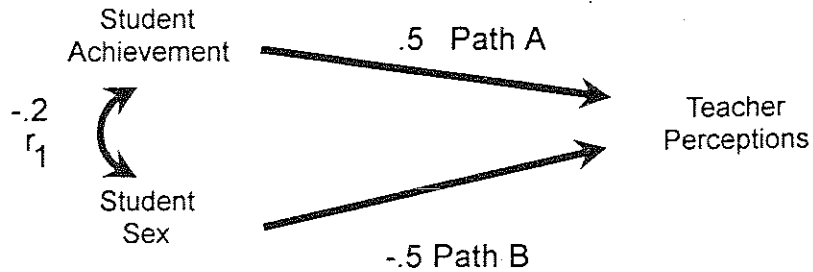## Model 3: Inaccuracy -- The Stereotype is in the Wrong Direction

Student
Achievement   .5 Path A

    Teacher
    Perceptions

-.2
r_1

Student
Sex   .2   Path B

Model 4: Inaccuracy -- The Stereotype Exaggerates Differences

Student
Achievement   .5 Path A

    Teacher
    Perceptions

-.2
r_1

Student
Sex   -.5 Path B

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

Student
Achievement   .5 Path A

    Teacher
    Perceptions

.0
r_1

Student
Sex   -.2 Path B
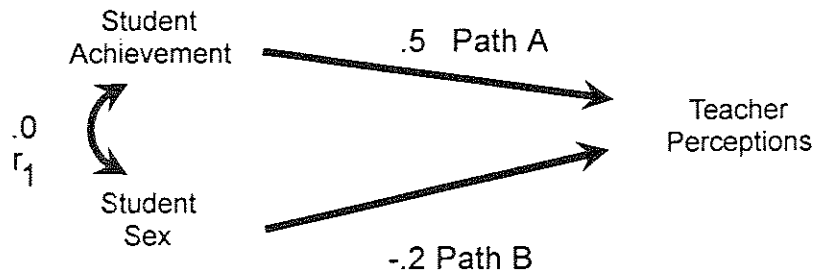
Fig. 9. Hypothetical models of inaccuracy in teacher perceptions of differences between boys and girls. Teacher perceptions of sex differences are inaccurate when the size of the perceived difference between boys and girls does not equal the actual difference between boys and girls. The perceived difference (r_2, not shown) equals the correlation between teacher perceptions and student sex, which equals Path B + ([r{student sex, student achievement}] × 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 B = 0 (indicating no stereotyping at all) and Path A = 1.0 (indicating complete reliance on achievement information), so that

r_2 (the correlation of teacher perceptions with student sex)
$$= 0 - .2 \ (1.0) = -.2.$$

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(the correlation of teacher perceptions with student sex)
$$= -.2 - .2(0) = -.2.$$

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_2(the correlation of teacher perceptions with student sex)
$$= .2 + (-2 \times .5) = .1.$$

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

$$r_2 = -.5 + (-.2 \times .5) = -.6.$$

The perceived difference between boys and girls ($r_2 = -.6$) 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_1, is zero):

$$r_2 = -.2 + 0(.5) = -.2$$

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., Beckett & Park, 1995; Bodenhausen, 1988; Darley & Gross, 1983; Krueger & Rothbart, 1988) showing that the influence of stereotypes on perceivers' judgments regarding targets is sufficient to demonstrate bias for or against a group. The models displayed in Figures 8 and 9, however, show that this is necessary but not sufficient to demonstrate bias. An influence of stereotypes on judgments of individual targets is systematic group bias when that influence leads perceivers to over- or underestimate the differences between individuals belonging to different groups (or when it leads perceivers to over- or underestimate how much individuals from differing groups vary, but this is an issue beyond the scope of the current paper).

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 be 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., Beckett & Park, 1995; Krueger & Rothbart, 1988) have rendered aggressiveness orthogonal to target sex. Of course, on the average, men really are more aggressive than women (Eagly 1995; Swim, 1994). In terms of the models shown in Figures 8 and 9, in the experimental studies, r_1 is artificially rendered zero. However, with regard to sex and aggressiveness under naturalistic conditions, $r_1 \neq 0$.

Consider the implication when this finding is combined with another frequent finding in the stereotype and person perception area. Virtually every study that manipulates targets' group memberships and individuating information finds that individuating information influences judgments more powerfully than does group membership (see, e.g., Jussim, 1990, 1991; Jussim, Madon, & Chatman, 1994, for reviews). Consider the following hypothetical situation: 1) there is a real sex difference (e.g., $r_1 = .2$); 2) targets' behavior influences judgments (e.g., Path A = .5); and 3) sex directly influences judgments to a small extent (e.g., Path B = .1). In this example, equation 1 is:

$$r_2 = .1 + .2(.5) = .2.$$

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 small basis of judgment? The answer to both

questions, in this 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 claiming 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 impressions 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 clearly errors. Of course, perceivers also may see some boys as less aggressive than they really are, and some girls as more aggressive than they really 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 overestimate than to underestimate 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 individuating 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 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 reassess the accuracy of these perceptions—accuracy is an issue of content, not process, and has already been demonstrated. Drawing on the models in Figures 8 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 (individuating information). Because the groups really did differ on some of these variables, teacher perceptions of the groups would also differ if teachers used the individuating 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 −.09, 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

Did Teachers Rely on Students' Sex, Independent of Students'
Achievement and Motivation?

	Betas relating to teacher perceptions of:		
Predictors	Performance $R^2 = .47*$	Talent $R^2 = .47*$	Effort $R^2 = .32*$
Student sex	−.09*	.02	−.16*
Fifth-grade final marks	.23*	.21*	.22*
Standardized test scores	.36*	.42*	.25*
Self-concept of math ability	.22*	.18*	.15*
Effort self-perceptions	.05*	.00	.11*
Time spent on homework	−.06*	−.05*	−.02

$*p < .01.$
$N = 1789$ (942 girls and 847 boys).
Betas are standardized regression coefficients.

(From Jussim, L., McCauley, C. R., & Lee, Y. T. (1995). Why study stereotype accuracy
and inaccuracy? In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward
appreciating group differences*. (pp. 3–27). Copyright © 1995 by the American Psychological
Association. Reprinted with permission.)

teacher perceptions. This means that teachers apparently stereotyped girls
as performing slightly higher than boys, independent of the actual slight
sex differences in performance. However, the extent to which teachers did
so corresponded reasonably well with the small sex difference in perfor-
mance. In other words, teachers perceptions of differences between boys
and girls were accurate because teachers relied on an accurate stereotype.

Results for teacher perceptions of talent provided no evidence of teachers
relying on a stereotype. The *beta* relating student sex to teacher perceptions
of talent was .02 (*ns*), corresponding closely to the 0.0 correlation of student
sex with standardized test scores.[9]

[9] The lack of a perceived difference for talent may seem to conflict with our previous
research (Jussim, 1989; Jussim & Eccles, 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 Jussim and Eccles (1992), the effect of
gender on teacher perceptions of talent was .08; the effect of gender on teacher perceptions
of performance was −.10; 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 Jussim and Eccles (1992), therefore, was .08 + (−.10*.64) = .02, which corresponds exactly
to the result reported here.

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 0.0, 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 .002 and was never statistically significant (all F's < 2.3, all p's $> .1$). Of the six possible individ-

TABLE IX

Did Teachers Rely on Students' Social Class, Independent of Students' Achievement and Motivation?

Predictors	Betas relating to teacher perceptions of:		
	Performance $R^2 = .48**$	Talent $R^2 = .48**$	Effort $R^2 = .30**$
Parental income	$-.02$	.00	.00
Parental education	.02	.05*	.02
Fifth-grade final marks	.26**	.21**	.27**
Standardized test scores	.35**	.42**	.21**
Self-concept of math ability	.20**	.16**	.11**
Effort Self-perceptions	.09**	.03	.13**
Time spent on homework	$-.06*$	$-.05*$	$-.03$

*$p < .05$.
**$p < .01$.
$N = 1066$.

Betas are standardized regression coefficients.

(From Jussim, L., McCauley, C. R., & Lee, Y. T. (1995). Why study stereotype accuracy and inaccuracy? In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences.* (pp. 3–27). Copyright © 1995 by the American Psychological Association. Reprinted with permission.)

ual relations between income and education and the three teacher perception variables, only one was statistically significant (education predicted teacher perceptions of talent, $p < .05$) but the *beta* was trivially small (.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 1588 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

TABLE X

Did Teachers Rely on Students' Ethnicity, Independent of Students' Achievement and Motivation?

Predictors	Betas relating to teacher perceptions of:		
	Performance $R^2 = .46****$	Talent $R^2 = .48****$	Effort $R^2 = .29*****$
Student ethnicity	.05**	.11****	.10***
Fifth-grade final marks	.24****	.21****	.24****
Standardized test scores	.32****	.44****	.26****
Self-concept of math ability	.20****	.17****	.10****
Effort self-perceptions	.03	.02	.10****
Time spent on homework	−.06**	−.05**	−.03

$*p < .05$.
$**p < .01$.
$***p < .001$.
$****p < .0001$.
$N = 1664$ (1588 White students and 76 African-American students).
Betas are standardized regression coefficients.

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 *betas* 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 *betas* 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 process 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

favorably biased ratings that teachers gave to the African-American students in the predominantly African-American district. Similarly, student sex had no influence on teacher perceptions of talent, after controls were used for students' actual achievement and motivation (although student sex did influence teacher perceptions of performance and effort—which we discuss later). These results clearly show that teachers did not rely on stereotypes to arrive at these judgments of students. Teachers were either oblivious to sex, class, and ethnic stereotypes, or they did not apply their stereotypes when evaluating their students.

Teachers probably were not oblivious to three of the major stereotypes in American culture. The cumulative wisdom of years of social psychological research on stereotypes instead points to the second explanation—that teachers did not apply their stereotypes in their evaluations of students. Thus, our results are consistent with abundant laboratory and field research showing that perceivers evaluate targets far more on the basis of targets' personal characteristics, than on targets' membership in social groups (e.g., Krueger & Rothbart, 1988; Linville, 1982; Locksley et al., 1980; see Fiske & Neuberg, 1990; Jussim, 1990, 1991, 1993, for reviews). In general, the more individuating information perceivers have, the less they rely on stereotypes (Eagly et al., 1991; Krueger & Rothbart, 1988; Locksley et al., 1980). Of course, teachers interacting with students over the first month of the school year generally have considerably more (and probably more objective) individuating information about students than do subjects in even the most ecologically valid laboratory experiment. Therefore, perhaps it should come as no surprise that, in general, these teachers did not rely much on their stereotypes when evaluating students.

There were a few exceptions to this pattern. In the case of student sex, teachers did seem to rely on their stereotypes regarding performance. They apparently evaluated students' performance based on their sex, independent of their actual achievement. However, the extent to which they did so yielded a relationship between student sex and teachers' perceptions that corresponded well with actual prior sex differences in achievement. It is important to highlight just what this means. Because even a valid stereotype does not apply equally well to all members of the stereotyped group, teachers probably misperceived some boys and girls. However, it also means that there was no tendency to systematically over- or underestimate the performance of girls as compared to boys.

In contrast, however, teachers seemed to be relying on an inaccurate stereotype in evaluating boys' and girls' effort. Teachers' more favorable impressions of girls' effort probably occurred because, on the average, girls are more cooperative and pleasant than boys, and because teachers prefer more cooperative and pleasant students (e.g., Brophy & Good, 1974; Bye,

1994; Wentzel, 1989). This is consistent with a growing body of literature showing that school is often a hostile place for low-achieving boys. For example, at least some teachers believe that boys suffer from inferior verbal skills, and this belief may become self-fulfilling (Palardy, 1969). Similarly, boys are referred for psychological evaluations far more often than girls, even when the teachers themselves do not rate boys as any more aggressive or in need of psychological services than girls (Bye, 1994). Similarly, one usually finds far more boys than girls in "special education" classes (Bye, 1994). Moreover, boys often receive lower grades than girls, even when their performance on standardized achievement tests are similar (Kimball, 1989).

In fact, this discussion highlights the possibility that affect, rather than or in addition to stereotypes, was driving the effort bias in favor of girls. Recent research on stereotypes and expectancies has suggested a more important role for affect (liking or disliking groups or individuals) in the occurrence of biased judgments and self-fulfilling prophecies (Esses, Haddock, & Zanna, 1993; Jussim et al., 1995; Rosenthal, 1989). Thus, if, on average, girls are more pleasant and cooperative than boys, teachers may come to like girls more than boys (on the average), and this may at least partially contribute to teachers' favorable views of girls' effort.

Teachers' reliance on an inaccurate sex stereotype regarding effort may also reflect attributional biases. Adults often are more likely to attribute females' math achievement to their effort than to their high ability (Yee & Eccles, 1988). Because teachers rated girls' performance slightly higher than that of the boys, but rated their talent the same, teachers may have needed to see girls as trying harder than boys to explain girls' higher performance level.

VIII. If the Cause Was Not Stereotype Bias, Then Why Were Expectancy Effects More Powerful Among Lower-SES and African-American Students, and Girls?

The previous section had two purposes. Our broader purpose was to provide some empirical evidence on the extent to which stereotypes bias person perception. A narrower purpose, which we hope did not get lost in the broader one, was to examine teacher stereotypes as a possible source of the greater expectancy effects among girls, lower SES students, and African-American students. However, we found so little evidence of stereotype-based biases that inaccurate stereotypes did not seem to be a particularly viable explanation for the pattern of differential expectancy effects. Thus, the question remained: Why are expectancy effects more powerful among some stigmatized demographic groups?

The next sections address this issue. First, we examine (and rule out) the possibility that teachers develop less accurate perceptions of students from stigmatized social groups. Second, we discuss another study showing that students with low self-concepts of ability or histories of low achievement in math, much the same as students from stigmatized groups, are considerably more vulnerable to self-fulfilling prophecies—a finding broadly consistent with Steele's (1992) perspective on African-Americans' disidentification with school and with research on students' vulnerability to school transition effects (Midgley et al., 1989).

A. WERE TEACHERS' PERCEPTIONS OF STUDENTS FROM STIGMATIZED GROUPS LESS ACCURATE?

One possibility is that, even if teachers were not particularly biased against these groups, they could still be less accurate in perceiving them. Thus, they may not systematically over- or underestimate the ability and performance of students from differing backgrounds. However, their errors, both positive and negative, might be larger for girls, low SES students, and African-American students. This may explain the pattern of differential expectancy effects because more inaccurate expectations have the potential to create larger self-fulfilling prophecies than more accurate expectations.

We performed another series of analyses to test this possibility. Specifically, we used the variables in the Base Model (except the teacher perceptions), plus one of the demographic characteristics, to predict the three teacher perception variables. We then examined the absolute value of the residuals produced by such an analysis. The residuals indicate whether teachers overestimate (positive residuals) or underestimate (negative residuals) particular students' performance, talent, and effort. Of course, because the demographic variable is controlled, its correlation with the raw residuals will be zero. Nonetheless, there still may be group differences in the absolute values of the residuals. For example, residuals of +8 and −8 for two girls, and +4 and −4 for two boys, will be uncorrelated with student sex. Obviously, however, in this hypothetical example, teachers are more accurate in perceiving boys than in perceiving girls.

For student sex, however, these analyses (Base Model, plus student sex predicting teacher perceptions of performance, talent, and effort) yielded no evidence that teachers were less accurate in perceiving girls. The correlations of student sex with the absolute value of the residuals were .03, −.01, and .09 for performance, talent, and effort, respectively (p's $= ns, ns,$ and .001), indicating that teachers were slightly less accurate in perceiving boys (girls

and boys were coded as 1 and 2, respectively). This, therefore, cannot possibly account for the larger teacher exepctancy effect on grades among girls.

A similar pattern emerged for social class. The multiple correlations of income and education with the absolute value of the residuals from the models predicting teacher perceptions (Base Model plus income and education) were .05, .02, .07, for teacher perceptions of performance, talent, and effort, respectively, (all p's $= > .05$). Thus, there was no evidence that teachers held more erroneous perceptions regarding lower SES students.

The results for ethnicity showed that teachers did hold slightly more inaccurate perceptions of African-American students than of White students. The correlations of ethnicity with the absolute values of the residuals from the models predicting teacher perceptions were .06, ($p < .05$), .07 ($p < .05$), and $-.02$ (ns) for performance, talent, and effort, respectively. Although greater error may have contributed to the stronger expectancy effects among African-American students, these differences are so small that they probably represent only a small or minor contribution.

B. MORE ON WHY: SELF AND PREVIOUS ACHIEVEMENT AS MODERATORS

The aforementioned findings indicate that teachers are about as accurate in perceiving girls as boys, lower SES as upper SES students, and African-American as White students. Therefore, greater inaccuracy cannot explain much, if any, of the greater expectancy effects among these students.

Then what does explain these greater expectancy effects? Perhaps something about these students (rather than something about their teachers) renders them more susceptible to expectancy effects. Perhaps students from stigmatized groups have fewer social and psychological resources for resisting teacher expectations. Their families may be less involved in their education (see, e.g., Lareau, 1987, regarding social class), rendering them more susceptible to the influence of other adult figures (such as teachers). The stresses associated with poverty and low income (single parent households, neighborhood crime and drug abuse, etc.) may reduce psychological resistance to teachers' influence. Students who face a relentless barrage of negative teacher expectations may "disidentify" with school (Steele, 1992) and may even take a certain pleasure in confirming teachers' negative expectations (Jussim, 1986). Perhaps a supportive teacher who holds students to higher standards may be seen as such a breadth of fresh air that many students are inspired to achieve more highly.

Although direct measures of students' social and psychological resources were not available, we did test these ideas indirectly. If lower SES and

African-American students (and to a lesser extent, girls) were more suscep-
tible to expectancy effects because they had fewer resources to resist such
expectations, then other students with fewer resources should also be more
susceptible to expectancy effects. Who might such students be? Those who
lack confidence and have histories of low achievement.

1. Self-Concept

Working with the same MSALT data and using essentially the same
Base Model and procedures much like those described earlier for assessing
moderation, we examined whether students' self-concepts moderated ex-
pectancy effects (Madon, Jussim, & Eccles, 1995). Using procedures much
the same as those reported here, we found that the self-fulfilling effects of
teacher perceptions were considerably stronger among students with lower
self-concepts of math ability than among student with higher self-concepts
of math ability. For example, for students whose self-concept was one
standard deviation below the sample mean, the standardized coefficient
relating teacher perceptions of performance to MEAP scores was .24,
whereas it was only .10 for students whose self-concept was one standard
deviation above the sample mean.

2. Previous Achievement

In much the same way, students with a history of low achievement might
also be more susceptible to expectancy effects. For example, Midgley et
al. (1989) examined the self-concepts and self-expectations for both low-
and high-achieving adolescents as they made the transition from elementary
school-based sixth grades to junior high school-based seventh grades. About
40% of the students moved from sixth-grade teachers with a high sense of
efficacy for their own teaching ability to seventh-grade teachers who had
doubts about their ability to teach low-skill students. Another 20% moved
in the opposite direction—from sixth-grade teachers who doubted their
ability to teach low-skill students to seventh-grade teachers who were con-
fident in their ability to teach students of all ability levels. The pattern of
change in self-perceptions of the high-achieving students was not affected
by which type of teacher transition they experienced. In contrast, the pattern
of change in the low-achieving students' self-perceptions were significantly
linked to the type of change they experienced in their teachers' expectations.
If they moved to a high-expectancy teacher, their own self-perceptions
increased in the seventh grade. In contrast, if they moved to a low-expec-
tancy teacher, their own self-perceptions decreased.

Why might low-achieving students be more susceptible to teacher expectancy effects? One possibility is that because low-achieving students feel less positive about their academic competence and are less certain of their future success than high-achieving students, these students may be more extrinsically motivated (Harter & Connell, 1984). Students who are more extrinsically motivated are likely to rely more on the teacher for motivation and for interpreting evaluative feedback, which makes these students more susceptible to teacher expectancy effects.

The work by Steele (1992) provides another possible explanation. When students "disidentify" with school, a history of low academic achievement may not be strongly reflected in their global self-esteem (Steele, 1992). Students seem most likely to disidentify with school when school becomes a painful place (either because of failure or cultural devaluation; see Steele, 1992). Disidentification means, in part, investing less energy in school work, which consequently leads to lower academic performance. It also means devaluing the importance of school achievement to one's self-worth. Thus, such students can maintain high self-esteem in the face of difficulties in school. This, in part, may help to explain why African-American students, despite lower levels of academic achievement, do not score lower on self-esteem measures than do White students (Crocker & Major, 1989).

Students with a history of low achievement may respond much the same as students with low self-esteem. Their motivation may be readily undermined by failure (or low teacher expectations), but be dramatically enhanced by a supportive and demanding teacher. We (Madon et al., 1995) have confirmed the hypothesis that self-fulfilling prophecies are stronger among low achievers than among high achievers. In this study, low achievement was operationalized as scores one standard deviation below the sample mean on standardized tests or previous grades and high achievement was operationalized as scores one standard deviation above the sample mean on standardized tests or previous grades. The standardized regression coefficient relating teacher perceptions of performance to MEAP scores was .28 for low achievers and .04 for high achievers. Similarly, the standardized regression coefficient relating teacher perceptions of performance to sixth-grade final marks was .24 for low achievers and .16 for high achievers.

3. Multiple Vulnerabilities

We also performed a series of follow-up analyses to determine whether the self-concept and achievement moderation effects we observed were independent of one another, and independent of the demographic moderation we described previously in this chapter. Because these analyses were quite complex, we only summarize our main findings here.

First, we assessed the independence of moderation by self-concept and by previous achievement. A set of analyses including both sets of moderators (self-concept by teacher perceptions, and previous achievement by teacher perceptions) showed that only the achievement moderators significantly predicted MEAP scores and sixth-grade final grades. The self-concept by teacher perception product terms did not significantly predict either students' future grades or MEAP scores in models that also included the previous achievement by teacher perception product terms.

Overall, therefore, these results indicate that achievement rather than self-concept is an active moderator of expectancy effects. Nonetheless, these results do not undermine the conclusion that students with lower self-concepts are more vulnerable to expectancy effects. They do help to explain why. Self-concept of ability is substantially correlated with actual performance (.41 with previous standardized test scores and .45 with previous grades). Students with records of lower previous achievement, who are the most vulnerable, are more likely also to have lower self-concepts of ability.

We also considered whether the moderating effects of achievement were similar for some of the differing demographic groups of students. However, because of the small number of African-American students, models with three-way product terms for ethnicity by previous achievement by teacher perceptions would not have yielded meaningful results. We did, however, examine whether the overall patterns of achievement moderation were similar for girls and boys, and for students from different SES backgrounds.

First, we created three-way product terms combining achievement, sex, and teacher perceptions and added these to the Base Model plus all lower-order two-way product terms. Neither the block of three-way terms nor any of the individual three-way terms significantly predicted either final grades or MEAP scores. These results indicated that the pattern of achievement moderation was similar for boys and girls.

Next, we examined whether the pattern of achievement moderation was similar for groups of students from different socioeconomic backgrounds. Again, we created three-way product terms combining achievement, SES, and teacher perceptions and added them to the Base Model plus all lower-order two-way product terms. None of the three-way terms (individually or as a block) significantly predicted grades, indicating that the pattern of achievement moderation was similar for groups of students from differing SES backgrounds.

However, the three-way product term combining parental education, students' previous standardized test scores, and teacher perceptions of performance did significantly ($p < .01$) predict MEAP scores. Examination of the regression coefficients showed that the relation of teacher perceptions of performance to MEAP scores was much higher (2.46 unstandardized, .62 stan-

dardized) for students with a history of lower achievement and with parents who did not complete high school than for any other combination of SES and previous achievement (coefficients ranging from .50 to 1.08, unstandardized, and .13 to .27 standardized).[10] The .62 standardized coefficient for low-achieving students from lower SES backgrounds, like the coefficients we observed relating teacher perceptions to achievement among African-American students, is one of the most powerful expectancy effects yet observed.

Overall, therefore, these results showed that low math achievers (who also tend to have lower self-concepts of math ability), much the same as low SES and African-American students, were more susceptible to self-fulfilling prophecies. Moreover, our results also showed that students with multiple vulnerabilities are more susceptible to self-fulfilling prophecies than are students with only one vulnerability.

IX. Other Moderators

Our quest for identifying conditions under which expectancy effects are large has only just begun. Undoubtedly, researchers will discover many conditions other than student demographics, self-concept, and previous achievement. Next, therefore, we discuss three classes of factors that may influence expectancy effect sizes: 1) characteristic of the perceiver; 2) characteristics of the target; and 3) situational factors.

A. PERCEIVER CHARACTERISTICS

1. Goals

Perceivers' goals may moderate the influence of their expectations on targets (Hilton & Darley, 1991). Self-fulfilling prophecies are more likely to occur when perceivers desire to arrive at a stable and predictable impression of a target (Snyder, 1992), when perceivers are more confident in the

---

[10] These coefficients were derived in the same way that the coefficients were derived in the earlier section on demographic moderation (see footnote 4). To obtain the coefficients reported here, we operationalized low parental education as "some high school," and high education as having completed a college BA. Low achievement was operationalized as performing at the 10th percentile of our sample, and high achievement was operationalized as performing at the 90th percentile of our sample. The coefficients for each of the four combinations of SES and previous achievement reported in the text were based on the results of the full model, which included the Base Model, the significant three-way product term, and all lower-order two-way product terms.

validity of their expectations (Jussim, 1986; Swann & Ely, 1984), and when they have an incentive for confirming their beliefs (Cooper & Hazelrigg, 1988). Self-fulfilling prophecies and perceptual biases are less likely when perceivers are motivated to develop an accurate impression of a target (Neuberg, 1989), when perceivers' outcomes depend on the target (Neuberg, 1994), and when perceivers' main goal is to get along in a friendly manner with targets (Snyder, 1992). Perceptual biases are more likely when perceivers strive to rapidly reach a particular conclusion (Kunda, 1990; Neuberg, 1994; Pyszczynski & Greenberg, 1987). These findings raise the following question: When are perceivers likely to be motivated by accuracy or a desire to get along in a friendly manner, and when are they likely to be overconfident in their beliefs or motivated by desires to reach a particular conclusion?

2. Prejudice, Cognitive Rigidity, and Belief Certainty

Prejudiced individuals seem especially unlikely to be motivated by either accuracy concerns or the desire to get along with members of the group they dislike. Instead, they seem likely to desire to reach the particular conclusion that members of the stigmatized group have negative, enduring attributes (Pettigrew, 1979). People high in cognitive rigidity or belief certainty also may not be motivated to consider different viewpoints. Cognitive rigidity, which is usually construed as an individual difference factor (e.g., Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950; Allport, 1954; Harris, 1989), and belief certainty, which is usually construed as a situational factor (Jussim, 1986; Swann & Ely, 1984), are both similar in that they describe people who may be unlikely to alter their beliefs when confronted with discomfirming evidence. Whether the source is prejudice, cognitive rigidity, or belief certainty (which may tend to co-occur with individuals; see Adorno et al., 1950), people overly confident in their expectations may be most likely to maintain biased perceptions of individuals and to create self-fulfilling prophecies (Babad, Inbar, & Rosenthal, 1982; Harris, 1989; Swann & Ely, 1984).

3. Other Individual Differences

Experienced perceivers may be less likely to create self-fulfilling prophecies. We use the term "experienced" here in two different but related senses. One aspect of experience refers to time on the job or in one's role. Thus, for example, more experienced teachers, therapists, doctors, and so forth have probably developed considerably more competence and expertise at appraising people such as students, clients, and patients. If so, then

their impressions may be more accurate. The second sense in which we use "experience" involves perceivers' experience with targets. When perceivers have greater information and more opportunities to interact with targets, they have greater opportunity to develop accurate beliefs. Thus, perceivers who have had more information about or experience with particular targets are also likely to be more accurate. Of course, accuracy reduces the potential for self-fulfilling prophecies.

Another such moderator may be professional efficacy. In general, efficacy refers to beliefs concerning one's ability to engage in the behaviors necessary for accomplishing a particular goal (Bandura, 1977). Professional efficacy, therefore, refers to beliefs regarding one's ability to engage in the behaviors necessary for accomplishing the essential work of one's profession. For example, teaching efficacy would refer to beliefs regarding one's ability to teach. When teachers are less confident in their teaching ability (low teaching efficacy), they may be more likely to create expectancy effects. Teachers low in teaching efficacy may feel less able to improve the skills of low-expectancy students; consequently, they may spend less time and effort with such students than do teachers high in teaching efficacy (Midgley et al., 1989). By virtue of spending less time with low-expectancy students (and perhaps more time with high-expectancy students), teachers low in teaching efficacy may exacerbate differences between high- and low-expectancy students to a greater extent than do teachers high in teaching efficacy (Midgley et al., 1989). A similar analysis could be readily applied to other professions (e.g., clinicians, managers, etc.).

A need to control others may also moderate expectancy effects. For example, the more that teachers strive to control students, the more likely it may be that their expectancies will be self-fulfilling and biasing. A high emphasis on control may include a particularly strong preference for having one's expectations confirmed. Control implies predictability, so unpredictable situations (or students) may be perceived as implying a lack of control. When students disconfirm expectations, therefore, teachers who emphasize control may feel threatened. These teachers may be most motivated to "ensure" that students confirm their expectations. This analysis is consistent with a less well-known finding of the original Rosenthal and Jacobson (1968) study: Some teachers responded especially negatively to the successes of students not specifically designated as late-bloomers.

B. TARGET CHARACTERISTICS

1. Goals

Targets may become more or less susceptible to self-fulfilling prophecies, depending on their goals. When perceivers' have something targets want

(such as a job), and when targets are aware of the perceiver's beliefs, they often confirm those beliefs in order to create a favorable impression (von Baeyer et al., 1981; Zanna & Pack, 1975). Similarly, when targets desire to facilitate smooth social interactions, they are also more likely to confirm perceivers' expectations (Snyder, 1992). In contrast, when targets believe that perceivers hold a negative belief about them, they often act to discomfirm that belief (Hilton & Darley, 1985). Similarly, when their main goal is to defend a threatened identity, or to express their personal attributes, targets are also likely to disconfirm perceivers' inaccurate expectations (Snyder, 1992).

2. Age

Self-fulfilling prophecies were strongest among the youngest students in the original Rosenthal and Jacobson (1968) study, suggesting that younger children may be more malleable than older children and adults. However, a meta-analysis has shown that the strongest teacher expectation effects occurred in the first, second, and *seventh* grades (Raudenbush, 1984). Further, the largest self-fulfilling prophecy effects yet reported were obtained in a study of adult Israeli military trainees (Eden & Shani, 1982). Although these findings do not deny a moderating role for age, they do suggest that situational factors may also influence targets' susceptibility to self-fulfilling prophecies.

C. SITUATIONAL FACTORS

1. New Situations

People may be more susceptible to confirming others' expectations when they enter new situations. Whenever people engage in major life transitions, such as entering a new school or starting a new job, they may be less clear and confident in their self-perceptions. Unclear self-perceptions render targets more susceptible to confirming perceivers' expectations.

This analysis may help to explain the seemingly inconsistent findings regarding age. Students in the first, second, and seventh grades, and new military inductees, are all in relatively unfamiliar situations. Therefore, all may be more susceptible to self-fulfilling prophecies.

2. Class Size and Resources

Expectancy effects may be more likely in classrooms with large numbers of students than they are in smaller classrooms. People are more susceptible

to biases when more of their "cognitive capacity" is being used—when they are trying to do several things at once (Gilbert & Osborne, 1989). The more students in a class, the more "cognitively busy" the teacher is likely to be, and, therefore, the more susceptible to biases and expectancy effects.

A related moderator may be class and school resources. Not only do resources (access to books, computers, laboratories, indoor and outdoor athletic facilities, fine arts, etc.) create a more generally pleasant learning environment; they probably make it easier for teachers to manage the students in their classes. Consequently, they, too, may be less likely to be cognitively overloaded, and, therefore, less suceptible to self-fulfilling prophecies.

At least one study (Finn, 1972) found results consistent with this perspective on class size and resources. Finn (1972) found that teacher expectations influenced the grades they assigned, but only in urban schools (not in suburban schools). Although urban and suburban schools differ in many dimensions, two differences often are class size (suburban schools often have smaller class sizes) and resources (suburban schools are often wealthier).

3. Tracking

School tracking refers to the policy of segregating students into different classes according to their ability. For example, smart students may be assigned to one class, average students to another, and slow students to a third. Tracking may be intended as a prosocial intervention. By putting students with similar capacities together, teachers have the opportunity to tailor their lessons in a way that maximizes those students' learning and achievement.

However, tracking may also moderate expectancy effects. Tracking represents institutional justification for believing that some students are smart and others are not. Due to our cultural beliefs regarding the meaning of low ability, particularly in math and science, tracking essentially provides students and teachers with an explanation for the students' low skill level that absolves both the student and the teacher of responsibility for continued learning. Thus, it may lead to the type of rigid teacher expectations that are most likely to evoke self-fulfilling prophecies and perceptual biases (Eccles & Wigfield, 1985; Jussim, 1986, 1990).

In addition, poor quality instruction may occur in at least some low-tracked courses. In part, this is a consequence of student characteristics. These classes are harder to manage, and traditional teaching techniques are not likely to be successful. However, teachers' expectations can also

exacerbate the poor environment. If teachers think that low-skill children cannot learn, or do not want to learn, they may reduce their teaching efforts (Allington, 1980; Evertson, 1982)—exactly the behavior that often leads to self-fulfilling prophecies (Harris & Rosenthal, 1985).

This situation is indeed unfortunate, considering the somewhat arbitrary nature of student placement in tracks, particularly for students of color and students from lower social class backgrounds (Dornbusch, 1994). In addition, when low-skill students were moved up in their track placement, both teacher expectations and students' actual performance on standardized tests improved (Tuckman & Bierman, 1971). In addition, the teachers in this study recommended that most of the students remain in the higher track the following year. These results suggest that long-term differences in the performance level of students in different tracks may reflect expectancy effects as well as ability differences. More field-based studies are needed to test this hypothesis.

X. Accumulation

Even if expectancy effects are small within a single school year, if such effects accumulate over several years, they may produce dramatic differences among students. Consider two students starting the sixth grade with identical IQs of 100. Nevertheless, the sixth-grade teacher believes that one student is bright and the other is dull. Assume that teacher expectations have an effect (in terms of standardized regression coefficients) of only .2 on student achievement. Further assume that the student believed to be bright by her sixth-grade teacher is believed to be bright by teachers in subsequent years, and that the student believed to be dull by her sixth-grade teacher is believed to be dull by her subsequent teachers.

An effect of .2 is equivalent to 1/5 of a standard deviation, and the standard deviation of IQ tests is 15. If a self-fulfilling prophecy increases the IQ of the high-expectancy student by only three points per year, and decreases the IQ of the low-expectancy student by only three points per year, by the end of high school, the "bright" student will have an IQ of 115, the "dull" student an IQ of 85. This is the power of "small" effects that accumulate!

The assumption that small effects accumulate lies at the heart of many strong claims regarding the power of expectations to create social reality. Such claims are usually based on experimental laboratory studies (see reviews by Jones, 1986; Snyder, 1984), even though they involve a single, brief interaction among strangers (e.g., Snyder et al., 1977; see Jussim, 1991,

for a review of these claims). The accumulation issue is particularly relevant to social stereotypes. Widely shared social stereotypes may lead many different perceivers to hold similar expectations for targets who are "marked" by some sort of stigma (race, handicap, institutional labels). Consequently, even if the self-fulfilling effects of perceivers' expectations are small within a single interaction, such effects may accumulate over many years and become a major source of individual differences.

However, do expectancy effects actually accumulate? Instead, perhaps they dissipate over time. Even if a teacher does create a 6-point IQ difference between two students, perhaps the next year that difference will tend to lessen or disappear completely. We know of only four studies that have empirically assessed the accumulation of expectancy effects. These are discussed next.

A. ROSENTHAL AND JACOBSON (1968)

Rosenthal and Jacobson (1968) manipulated teachers' expectations in the first year by randomly selecting students and designating them as "late bloomers." However, in the second year, teachers developed expectations without direct intervention by the experimenters. The accumulation hypothesis predicts that there would be greater differences between "late bloomers" and controls in the second year than in the first year. In fact, the opposite was found: The differences between these students significantly declined after two years. On the average, "late bloomers" had a 3.80 IQ point advantage over controls at the end of the first year, but only a 2.67 IQ point advantage at the end of the second year.

B. RIST (1970)

Rist (1970, described previously) followed a class of kindergarten students through second grade. Unfortunately, he provided no quantitative information regarding students' learning, IQ scores, or achievement in first or second grade. Thus, it is impossible to determine whether expectancy effects accumulated. Although Rist (1970) concluded that he had observed a rigid cast-like system based on social class, which suggests large and powerful accumulation effects, his own observations actually suggest dissipation instead. As did the kindergarten teacher, the first-grade teacher assigned students to three tables (apparently according to her beliefs about the smart, average, and dumb students). All of the Table 1 ("smart") students in kindergarten were assigned to Table 1 in first grade. However, students

at Tables 2 and 3 in kindergarten were all assigned to Table 2. Thus, if table assignment is the criterion, kindergarten differences between Tables 2 and 3 disappeared by first grade, although differences between those children and Table 1 students were maintained.

Rist (1970) reported further reduction of apparent differences in second grade. In the second-grade class, the students who had been assigned to Table 1 in first grade were all assigned to their own table (they were referred to as "tigers"). Students who had been assigned to Tables 2 and 3 in first grade, in the second-grade class were assigned to a second table (referred to as "cardinals"). None of the students from the first-grade glass Rist observed were assigned to the "slow" table (called "clowns"). In addition, Rist (1970) observed that in January, two of the tigers were moved to the cardinals' table, and two of the cardinals were moved to the tigers' table. Thus, although some of the differences among students in kindergarten were maintained through second grade, overall differences between the groups seem to have declined.

C. WEST AND ANDERSON (1976)

West and Anderson (1976) examined relationships between teacher expectations and student achievement in a period running from the freshman through the senior year of high school. The accumulation hypothesis predicts that the coefficients relating freshman-year teacher expectations to senior-year achievement will be larger than those relating freshman-year teacher expectations to sophomore-year achievement. However, their results showed dissipation: The coefficient relating freshman-year teacher expectations to senior-year achievement (.06) was smaller than the coefficient relating to sophomore-year achievement (.12).

D. FRIEZE ET AL. (1991)

Frieze et al. (1991) addressed the accumulation issue by comparing the extent to which the attractiveness of MBAs predicted starting salary versus salary in 1983 (several years later). The unstandardized coefficients relating attractiveness to 1983 salary (2.60 for men and 2.13 for women) were higher than those relating to starting salary (1.13 and 0.28, respectively). Whether these results indicate accumulation of self-fulfilling prophecy effects or accumulation of greater rewards to more socially skilled managers (as discussed previously, the more attractive tend to be more socially skilled), however, is unclear.

These are the only four studies (Frieze et al., 1991; Rist, 1970; Rosenthal & Jacobson, 1968; West & Anderson, 1976) to our knowledge that have directly assessed whether expectancy effects accumulate. These paint a decidedly mixed picture, and all have major conceptual or methodological limitations (see also Elashoff & Snow, 1971; Jussim & Eccles, 1995). Although expectancy effects may accumulate over time, there is currently no evidence clearly demonstrating that they actually do. Strong, empirical evidence on this issue is sorely needed.

E. CONCURRENT ACCUMULATION EFFECTS

The sparce empirical research on accumulation effects has all focused on accumulation over time. However, it is also possible that, within a single time frame (e.g., 1 school year), the effects on targets of multiple perceivers' expectations may accumulate. To distinguish such effects from the accumulation of expectancy effects over time, we refer to these as "concurrent accumulation effects."

The notion of concurrent accumulation effects is implicit in most perspectives that emphasize the potentially self-fulfilling nature of social stereotypes (e.g., Deaux & Major, 1987; Hamilton et al., 1990; Jones, 1990; Snyder, 1984). Because stereotypes are often shared, multiple perceivers will often develop similar expectations for individual members of the stereotyped group. Perceiver after perceiver will presumably heap self-fulfilling prophecy after self-fulfilling prophecy upon stereotyped targets.

Such a perspective appears to imply that the self-fulfilling prophecy effects observed in most individual studies probably underestimate the true extent to which individual targets' are influenced by others' expectancies, because all previous research has focused on the potentially self-fulfilling effects of only one perceiver on each target. If multiple perceivers influence targets, then one might expect that, in the course of daily life, people would be more influenced by self-fulfilling prophecies than is implied by existing research.

Figure 10 presents a simplified general model of concurrent accumulation effects. The model includes two self-fulfilling prophecy paths: Path A (linking one perceiver's expectations to targets) and Path B (linking other perceivers' expectations to targets). r_1 is the correlation between perceivers' expectations. The displayed models are simplified in three ways. First, if there are many "other perceivers," there really could be many more paths and correlations. Second, none of the control variables necessary to actually assess expectancy effects are displayed. Third, we assume all paths are standardized.
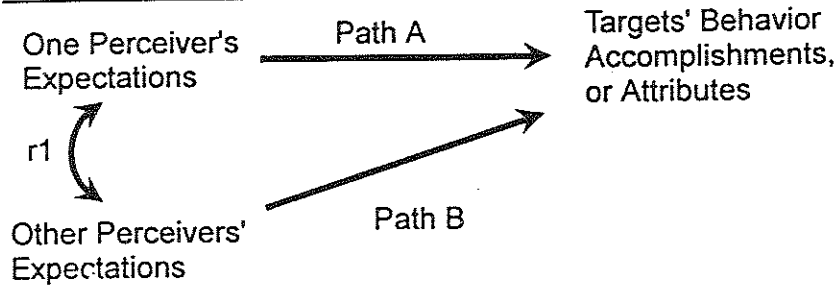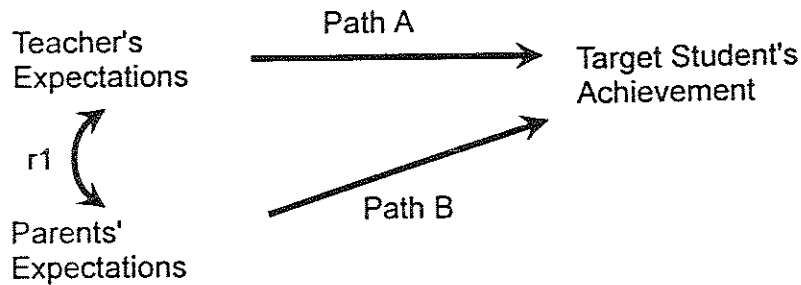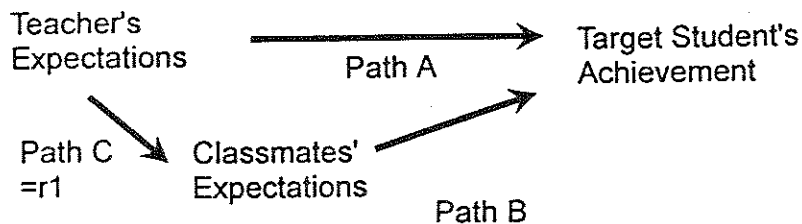
Model 1: General Model

One Perceiver's   Path A   Targets' Behavior
Expectations   ──────────▶   Accomplishments,
                                     or Attributes
r1

Other Perceivers'   Path B
Expectations

Model 2: Teachers and Parents Hold Similar Expectations

Teacher's   Path A
Expectations   ──────────▶   Target Student's
                                   Achievement
r1

Parents'   Path B
Expectations

Model 3: Teacher Expectations Cause Classmate Expectations

Teacher's   ──────────▶   Target Student's
Expectations   Path A   Achievement

Path C   ↘ Classmates'
=r1   Expectations
                      Path B

Fig. 10. Models of concurrent accumulation of expectancy effects. For simplicity of presentation, none of the models displayed here include the control variables that would be necessary to actually assess expectancy effects.

Nonetheless, simply making this model explicit leads to some surprising insights. Model 2 presents a concrete hypothetical example. The model assumes that both teachers' expectations and parents' expectations have self-fulfilling effects on students (Paths A and B, respectively). The similarity between parents' and teachers' expectations is represented by r_1. One might be tempted to conclude that studies focusing only on effects of teacher expectations, for example, would underestimate total expectancy effects because they do not assess effects of parents' expectations.

Figure 10 shows that such a conclusion is not warranted. This analysis is a variant on the omitted variable problem (see our earlier discussion of

limitations to naturalistic research on expectancies). In the hypothetical teacher expectation study shown in Model 2, parental expectations correlate with teacher expectations and also cause student outcomes. If effects of parental expectations are not explicitly included in the model, the effects of teacher expectations will be overestimated. In fact, the estimated path coefficient linking teacher expectations to student achievement will equal

$$\text{Path A} + (r_1 * \text{Path B}).$$

Conceptually, the teacher expectation–student achievement coefficient will be biased upward precisely to the extent that parent expectations overlap with teacher expectations and parent expectations themselves cause student outcomes. In other words, the coefficient linking teacher expectations to student achievement will also include the self-fulfilling effect of parental expectations, to the extent that parent and teacher expectations overlap. If they do not overlap, there is no potential for concurrent accumulation, even if parent and teacher expectations are both self-fulfilling.

Model 3 presents another variation on this idea. In this example, teacher expectations cause both classmates' expectations and target students' achievement. However, if classmates' expectations are not assessed, the estimated path coefficient linking teacher expectations to student achievement will equal

$$\text{Path A} + (\text{Path C} * \text{Path B}).$$

In this case, failure to assess the classmates' mediating paths does not "overestimate" teacher expectation effects at all. Model 3 is a classic example of a direct and indirect effects model (e.g., Alwin & Hauser, 1975; see Jussim, 1991, for several examples applied to social perception and expectancies). The total effect of teacher expectations on student achievement equals the sum of its direct effect (Path A) and its indirect effect (Path C * Path B). In other words, if there are important mediators, even if they are not assessed, the total effect of teacher expectations on student achievement simply equals the path coefficient linking them. In this situation, there is no underestimation of concurrent accumulation.

This analysis leads to several surprising conclusions. Studies that assess effects of only a single perceiver's expectations on each target are not likely to be underestimating concurrent expectancy effects. If there is any bias, it is likely to be in overestimating the effects of the expectations of the perceivers who are included in the study. However, as indicated in the

models displayed in Figure 10, the estimated effects on targets of individual perceivers who are included in the study should approximate the total self-fulfilling effect of all perceivers (even those excluded from the study) whose expectations overlap with those of the included perceivers. Thus, studies of individual perceiver–target relationships probably do not underestimate the accumulation of concurrent self-fulfilling prophecies.

One caveat is in order. Concurrent accumulation requires different perceivers to hold similar expectations for the target. This is captured by r_1 in the models in Figure 10. Concurrent accumulation generally will underestimate the total extent to which targets are influenced by self-fulfilling prophecies, because perceivers will rarely hold identical expectations for those targets. To the extent that perceivers hold different expectations for the target, even if their expectations are self-fulfilling, there will be little net accumulation. For example, consider Fred, who is neither introverted nor extraverted. Let us assume, furthermore, that two of Fred's friends believe him to be extraverted and two other friends believe him to be introverted. If all of their expectations are approximately equally self-fulfilling, overall, there will be no accumulation—he will remain neither particularly introverted nor very extraverted.

XI. Conclusion

A. ARE SELF-FULFILLING PROPHECIES OFTEN POWERFUL AND PERVASIVE?

This article has described our own and others' research documenting three main phenomena. First, claims about the power of expectancy effects notwithstanding, current evidence from both naturalistic and experimental studies indicates that, in general, self-fulfilling prophecies are not very powerful (see reviews by Brophy, 1983; Eccles & Wigfield, 1985; Jussim, 1991; Jussim & Eccles, 1995; Wineburg, 1987; see also meta-analyses by Raudenbush, 1984; Rosenthal & Rubin, 1978). To date, only naturalistic studies have attempted to compare the extent to which perceivers' expectations predict targets' behavior because those expectations are accurate versus self-fulfilling. These studies consistently show that teacher perceptions predict student achievement more because those perceptions are accurate than because they lead to self-fulfilling prophecies. The little research that has addressed naturally occurring expectancy effects outside of the classroom generally yields similar findings (see Jussim & Eccles, 1995, for a review).

Despite this repeatedly documented pattern of high accuracy and low self-fulfilling prophecy (see reviews by Brophy, 1983; Brophy & Good, 1974; Jussim, 1990, 1991, 1993; Jussim & Eccles, 1995; Jussim et al., 1994), many social psychological perspectives focusing primarily on experimental research often assume or conclude that self-fulfilling prophecies are common and even powerful (e.g., Fiske & Taylor, 1991; Hamilton et al., 1990; Jones, 1986, 1990; von Hippel et al., 1995). To the extent that the criterion for arriving at such conclusions is the evidence regarding what happens under naturalistic conditions, we would argue that it is time for social psychology to discard its belief that expectancy effects are generally powerful and pervasive.

B. WHEN ARE TEACHER EXPECTATION EFFECTS MORE POWERFUL?

This, of course, does not mean that expectancy effects are never powerful. Since we first discovered this pattern of high-accuracy and low-expectancy effects in our own initial studies (Jussim, 1989; Jussim & Eccles, 1992), we have been on a quest to identify conditions under which expectancy effects are more powerful. We have actually uncovered quite a few (these are the second major phenomena we have documented in this article). Expectancy effects are considerably stronger among students from stigmatized groups (African-Americans, lower SES, and, to a smaller extent, girls), and among students with low self-concepts and records of poor previous achievement. It is likely that different processes partially account for each of these groups' greater susceptibility to expectancy effects. However, we have speculated that reduced social and psychological resources for combating erroneous teacher expectations may at least partially underlie the greater susceptibility to expectancy effects that characterizes each of these groups.

C. THE ROLE OF STEREOTYPES IN TEACHERS' PERCEPTIONS OF STUDENTS

The third major contribution of this chapter has been to provide some of the first evidence regarding the role of stereotypes in naturally occurring person perception. Although the role of stereotypes in person perception has been a hot topic (e.g., Beckett & Park, 1995; Bodenhausen, 1988; Darley & Gross, 1983; Krueger & Rothbart, 1988; Locksley et al., 1980, 1982; Nelson, Biernat, & Manis, 1990), there has been little naturalistic research addressing the question (see Jacobs & Eccles, 1992, for an excep-

tion). Thus, another contribution of the research described in this article is to provide some of the first empirical evidence regarding the extent to which stereotypes bias person perception among real people making real decisions in real situations. We think the time is ripe for a flood of naturalistic social psychological studies addressing this issue.

Our results show that, in general, teacher perceptions of sex, social class, and ethnic differences and similarities were highly accurate. Such results would seem to contrast with much emphasis on stereotypes biasing social perception (e.g., Fiske & Neuberg, 1990; Fiske & Taylor, 1991; Hamilton et al., 1990; Jones, 1986, 1990; Stangor, 1995). Although we did find some evidence of bias, for the most part, teachers' perceptions of the groups closely corresponded to the group members' grades, achievement, and motivation. Such findings are actually consistent with a number of perspectives all arguing that stereotypes may be either accurate or inaccurate, and that, in general, issues of stereotype accuracy and inaccuracy are considerably more complex than once thought (see, e.g., Ashmore & Longo, 1995; Brigham, 1971; Eagly, 1995; Fox, 1991; Judd & Park, 1993; Jussim, 1990; Jussim et al., 1995; Mackie, 1973; McCauley et al., 1980; Oakes, Haslam, & Turner, 1994; Ottati & Lee, 1995; Ryan, 1995).

We also presented a simple theoretical model for addressing issues of both content and process in stereotyping. The model shown in Figures 8 and 9 may be used to identify whether perceivers' judgments of the differences between individual members of different groups actually corresponds to the existing group differences, if there are any (see also Beckett & Park, 1995; Jussim, 1991). This model is also useful for determining the extent to which judgments were based on individuating information versus social category membership, and shows that people's use of categorical information does not necessarily lead people to unfairly favor one group over another. When individuating information is less than perfectly diagnostic, and when there are real differences between groups, perceivers who base their judgments of individual targets on those targets' social category will arrive at more valid perceptions of group differences than perceivers who do not base their judgments on those targets' social category (see also Funder, in press; Jussim, 1991; Kahneman & Tversky, 1973).

Of course, we are not claiming that either our model or our empirical results show that bias, prejudice, and discrimination do not exist or are unimportant. Obviously, they do exist, and they are terribly important. However, it is also possible that person perception biases produced by stereotypes exist to a smaller extent than one might assume on the basis of the experimental laboratory research. To challenge the tentative hypothesis that biases produced by stereotypes outside of the laboratory may not be that powerful, social psychologists will have to move their research pro-

grams out of their laboratories and investigate stereotype-induced biases in naturally occurring situations.

D. BIAS AND DISCRIMINATION WITHOUT EXPECTANCY EFFECTS

It is also possible that bias and discrimination may be manifest in ways very different from those assessed in the current studies, or that they must be assessed in a manner different from that of the typical social psychology laboratory experiment. Barriers may exist to equal employment opportunities, even in the complete absence of employer bias. For example, different social networks may constitute one such barrier (e.g., Braddock & McPartland, 1987). We live in a (still) highly segregated society—Whites are more likely to associate with Whites; African-Americans are more likely to associate with African-Americans. Whites hold more managerial jobs, and job openings are often filled through informal networks. Because Whites are more likely to be "plugged in" to such networks, they will have greater job opportunities. This may occur even if White employers judge the applicants who come to their attention solely on their merits.

E. SELF-FULFILLING PROPHECIES WITHOUT BIASED OR INACCURATE PERCEIVERS

Similarly, stereotypes may create self-fulfilling prophecies that are in no way the fault of the individual perceiver. A series of studies by Zanna (von Baeyer, Sherk, & Zanna, 1981; Zanna & Pack, 1975) showed that when women believed they were to be interviewed by a traditional or sexist man, they often acted in such a way as to confirm sex stereotypes. One reason these studies are interesting is that many of the dependent variables (how much make-up and accessories the women wore, their performance on a test, etc.) were all assessed prior to the interview (which, in the case of Zanna & Pack, 1975, never took place). Similarly, when targets believed that perceivers viewed them as mentally ill, even if perceivers were blind to targets' mental health status, targets actually evoked more rejection from those perceivers (e.g., Farina, Allen, & Saul, 1968; Farina, Gliha, Boudreau, Allen, & Sherman, 1971). These are still self-fulfilling prophecies in the sense that stereotypes create their own reality. However, the self-fulfilling prophecy trigger in these studies is not the beliefs held by bigoted or error-prone perceivers—it is targets' beliefs about how perceivers view them.

Steele's (1992) analysis of disidentification as a source of African-American underachievement is also consistent with this perspective. Steele argued that, because the cultural milieu devalues African-Americans, they are wounded more deeply by scholastic difficulties than are other students. Note, however, that Steele's analysis predicts African-American underachievement, even if African-American students never take a class with a biased teacher (see also Fordham & Ogbu, 1986). Cultural stereotypes may have a life of their own, and may create self-fulfilling prophecies even when individual perceivers do not.

F. BEYOND THE DYAD: SELF-FULFILLING PROPHECIES AT ORGANIZATIONAL, INSTITUTIONAL, AND SOCIETAL LEVELS OF ANALYSIS

In addition to culturally based self-fulfilling prophecies, institutional policies may also create self-fulfilling prophecies. For example, Merton (1948) documented how, in the early part of this century, most labor unions barred African-Americans from membership. Union members often claimed that African-Americans were strikebreakers and could not be trusted. This severely limited the job opportunities of African-Americans. When faced with a strike, companies often offered jobs to all takers, and African-Americans often jumped at the chance for work. Thus, the union's beliefs about African-Americans were confirmed. It is important to note, however, that if an individual union member, acting alone, held this stereotype of African-Americans, it would have had no effect at all on reducing the job opportunities of African-Americans.

In fact, Merton's (1948) original analysis of self-fulfilling prophecies focused primarily on broad-based sociological patterns and institutional practices. However, self-fulfilling prophecies are probably considerably more easily studied as a dyadic interaction level phenomenon (and, of course, it is important at the dyadic level, too). Elsewhere, however, we (Jussim & Fleming, in press) have attempted to update Merton's (1948) analysis by identifying ways in which modern institutional practices create self-fulfilling prophecies. We have suggested that school tracking may contribute to ethnic self-fulfilling prophecies, that funding schools through local property taxes may contribute to social-class self-fulfilling prophecies, and that the allocation of academic rewards (jobs, article acceptances, etc.) are characterized by self-fulfilling prophecies based on institutional prestige.

In the spirit of Merton's (1948) original essay, we (Jussim & Fleming, in press) have speculated that a sociological level analysis of self-fulfilling prophecies might contribute to understanding the 1992 Los Angeles riots. These riots, among the most destructive civil disturbances of this century,

are often considered to be a response to the perceived injustice of the "not guilty" verdicts returned in the case against the police officers who beat Rodney King, an African-American motorist. The riot surely resulted from the interplay of many social forces, and a three-step self-fulfilling prophecy analysis may contribute to understanding some sources of the riots.

The first step is expectations: Many Whites have historically held, and continue to hold, negative stereotypes about many minority groups (see reviews by Allport, 1954; Marger, 1991). These beliefs probably contributed to the second step: discrimination. In the last 20 years, Whites have seemingly become less sympathetic to social programs, such as school desegregation and affirmative action, that are designed to provide greater educational and occupational opportunities for minorities (Marger, 1991). Through blatant and subtle forms of discrimination, many Whites continue to limit and undermine the quality of life for many minority groups.

Discrimination may lead to the final step in this self-fulfilling prophecy— riots—in several ways. First, discrimination may create a deep resentment among many minority group members, a resentment that may be triggered by certain conditions into riotous behavior. Second, discrimination probably reduces support for the general social structure. For example, many African-American teenagers may not vigorously pursue high educational achievement because 1) high achievement may be seen as "acting White" and as rejecting one's own ethnic group (e.g., Fordham & Ogbu, 1986; Steele, 1992); or 2) as a result of later job discrimination, education is seen as producing little or no economic payoff. People who have not greatly invested in the social system are probably more likely to take whatever they can get away with when a golden opportunity, such as a riot, appears. Thus, even when the rioters were inspired more by self-interest than by abstract political agendas, discrimination probably played an important role. This type of violent, antisocial behavior, of course, confirms for many Whites the validity of their negative beliefs about minorities.

Of course, we are not claiming that this type of self-fulfilling prophecy analysis completely accounts for such a large-scale and complex social phenomenon as the Los Angeles riots. Furthermore, empirical research that actually documents such sociological self-fulfilling prophecies is considerably more difficult to perform than research on dyadic self-fulfilling prophecies. However, we suspect that at least sometimes, such effects may be quite powerful.

G. WHENCE RESEARCH ON SELF-FULFILLING PROPHECIES?

Social scientists have learned much about self-fulfilling prophecies in the 50 years since Merton (1948) coined the term and in the 30 years since

Rosenthal & Jacobson (1968) triggered an explosion of interest in the area. We know that the phenomenon is indeed real (an issue that was hotly contested through the 1970s (see, e.g., Elashoff & Snow, 1971, or the commentaries on Rosenthal & Rubin's 1978 meta-analysis). We also know much about how they happen (see reviews by Brophy, 1983; Darley & Fazio, 1980; Eccles & Wigfield, 1985; Jussim, 1986; Rosenthal, 1974; see Harris & Rosenthal, 1985, for a meta-analysis) and something of the conditions under which they are more or less likely. Next, therefore, we offer some suggestions regarding potentially fruitful directions for future research on self-fulfilling prophecies.

1. Moderators

In the last 15 years, much research on self-fulfilling prophecies has focused on moderators (e.g., Brattesani, Weinstein, & Marshall, 1984; Neuberg, 1989; see reviews by Neuberg, 1994; Snyder, 1992; see meta-analyses by Cooper & Hazelrigg, 1988; Raudenbush, 1984). Social scientists are only beginning to understanding how the power of self-fulfilling prophecies depends on characteristics of perceivers, targets, and situations. Research on moderators, therefore, is likely to continue to contribute important insights into the role of expectancies in creating social reality and social problems.

2. Mediators

Research on mediators has consistently supported Rosenthal's (1974) four-factor theory (see Jussim, 1986, for a review; see Harris & Rosenthal, 1985, for a meta-analysis). The four-factor theory claims that perceivers act on their expectations in four broad classes of ways that can be described in these terms: climate, feedback, input, and output. Perceivers provide more socioemotional warmth (climate), clearer and more positive feedback (feedback), spend more time with and lavish more attention on (input), and provide more opportunities for high achievement to (output) high-expectancy targets. Perhaps because this pattern has been so well documented, there has been little research on mediators in the last 10 years.

However, other types of mediators have been underexplored. In performance situations, abundant research attests to the power of setting high goals for students, employees, and athletes, to name a few groups (Locke & Latham, 1990). However, whether high expectations often lead perceivers to explicitly set higher goals for targets is not known. However, even if perceivers do not set explicit goals for targets, perceivers may sometimes

explicitly convey high expectations, which may have an effect much the same as setting high goals. However, both the extent to which perceivers do this and its effect on targets, is currently unknown.

Two recent studies suggest that the role of affect in driving "expectancy" effects has been underexplored. The first found that children were less warm, friendly, and involved when playing with other children who were stigmatized (Harris, Milich, Corbitt, Hoover, & Brady, 1992). The second found that perceivers' liking or disliking of (prejudice toward) a target's group was a more potent source of biases in judgments of that target's sanity than were perceivers' beliefs (stereotypes) about that group (Jussim et al., 1995).

In addition, the results of several classic self-fulfilling prophecy studies may be readily interpreted as the result of perceivers' affect. For example, in the classic Snyder et al. (1977) study, college men were more pleasant to the supposedly attractive college women. The interpretation of Snyder et al. (1977) was that the men's behavior was triggered by the physical attractiveness stereotype. Perhaps, however, many of the college-age men liked the supposedly attractive women because of their beauty per se and gave little thought to their personal characteristics. Similarly, in Word et al.'s (1974) classic study of race-based self-fulfilling prophecies, many of the behavioral mediators (more speech errors, greater distance, shorter interview to African-American applicants) seemed to reflect anxiety or dislike more than beliefs. Also, much of what drives teacher-expectancy effects may be that teachers like high-expectancy students more than they like low-expectancy students (Rosenthal, 1989; see also Olson, Roese, & Zanna, in press, for a review of how expectancies influence affect).

Another underexplored mediator is targets' beliefs about perceivers' beliefs. A few experiments have shown that targets sometimes confirm the beliefs that they (erroneously) think perceivers hold (Farina et al., 1968, 1971; von Bayer et al., 1981; Zanna & Pack, 1975). The general question here is: How important is targets' awareness (accurate or not) of perceivers' expectations? We speculate that although awareness is not a necessary mediator of self-fulfilling prophecies (i.e., self-fulfilling prophecies may occur without target awareness of the perceivers' expectancies), awareness will often tend to enhance the power of self-fulfilling prophecies, especially among children and people in new situations. Of course, targets may sometimes intentionally resist confirming expectations when they believe that a perceiver holds inappropriate expectations (Hilton & Darley, 1985; Swann & Ely, 1984). Understanding the role of target awareness in self-fulfilling prophecies, then, poses an important question for both mediation and moderation studies.

3. Naturalistic Research Beyond Teachers and Students

To date, naturalistic studies of expectancies have focused almost exclusively on teachers and students (see Jussim & Eccles, 1995, for a review). There have only been a very few naturalistic studies of self-fulfilling prophecies in other areas (Berman, 1979; Frieze et al., 1991; Jacobs & Eccles, 1992). Although research on teacher expectations will remain important, naturalistic research on expectancy effects among parents and children, employers and employees, clinicians and patients, and so on is greatly needed in order to understand the extent and power of self-fulfilling prophecies in daily life.

4. Accumulation and Sociological Level Self-Fulfilling Prophecies

We believe that the accumulation issue is inherently linked to sociological level self-fulfilling prophecies and to self-fulfilling prophecies resulting from targets' beliefs about the beliefs of others. At the sociological level, many negative stereotypes are widely shared, so that targets will frequently confront others' unfavorable views of them. They may also sometimes face social policies designed to exclude them from full equality with other citizens (e.g., in the United States, neither the federal government nor most states have civil rights laws providing equal protection for gays and lesbians). When group membership is physically salient (gender, race/ethnicity, attractiveness, disability, etc.) the potential for dyadic-level bias, blatant or subtle (e.g., glass ceilings), is increased. Moreover, for many such groups, the societal-cultural discourse focuses on some alleged inferiority (e.g., the ongoing festering and inflammatory "debate" over whether Blacks are intellectually inferior to Whites genetically; the presumption of many people that women and minorities in positions of power and prestige got there through unfair and preferential selection procedures). It does not seem particularly far-fetched to suggest that members of such groups may eventually either internalize some of these beliefs (e.g., Heilman, Simon, & Repper, 1987), become more deeply wounded by the failures that accrue to almost everyone (Steele, 1992), or themselves develop an (at least sometimes) inaccurate but ultimately self-fulfilling expectation that others hold negative views of them. Empirical research on the accumulation of the effects of socially, institutionally, and organizationally shared beliefs and discourses could begin to fulfill the promise of Merton's (1948) original sociological level analysis of self-fulfilling prophecies.

Acknowledgments

References

Adorno, T., Frenkel-Brunswick, E., Levinson, D., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.

Allington, R. (1980). Teacher interruption behavior during primary grade oral reading. *Journal of Educational Psychology, 72,* 371–377.

Allport, F. H. (1955). *Theories of perception and the concept of structure*. New York, Wiley.

Allport. G. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Alwin, D., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40,* 37–47.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111,* 256–274.

American Association of University Women (AAUW). (1992). *How schools shortchange girls*. American Association of University Women Education Foundation.

American Psychological Association. (1991). In the Supreme Court of the United States: Price Waterhouse v. Ann B. Hopkins (Amicus curiae brief). *American Psychologist, 46,* 1061–1070.

Andersen, S. M., & Bem, S. L. (1981). Sex typing and androgeny in dyadic interaction: Individual differences in responsiveness to physical attractiveness. *Journal of Personality and Social Psychology, 41,* 74–86.

Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of Personality and Social Psychology, 35,* 443–449.

Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual appraoches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale, NJ: Erlbaum.

Ashmore, R. D., & Longo, L. C. (1995). The accuracy of stereotypes: Considerations suggested by research on physical attractiveness. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 63–86). Washington, DC: American Psychological Association.

Babad, E., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology, 74,* 459–474.

Bandura. A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Bulletin, 84,* 191–215.

Baron, R. M., Albright, L., & Malloy, T. E. (1995). The effects of behavioral and social class information on social judgment. *Personality and Social Psychology Bulletin, 21*, 308–315.

Beckett, N. E., & Park, B. (1995). Use of category versus individuating information: Making base rates salient. *Personality and Social Psychology Bulletin, 21*, 21–31.

Berman, J. S. (1979). Social bases of psychotherapy: Expectancy, attraction, and the outcome of treatment. (Doctoral disertation, Harvard University). *Dissertation Abstracts International, 40*, 5800B.

Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. *Journal of Personality and Social Psychology, 55*, 726–737.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645–657.

Braddock, J. H., & McPartland, J. M. (1987). How minorities continue to be excluded from equal employment opportunities: Research on labor market and institutional barriers. *Journal of Social Issues, 43*, 5–40.

Brattesani, K. A., Weinstein, R. S., & Marshall, H. H. (1984). Student perceptions of differential teacher treatment as moderators of teacher expectation effects. *Journal of Educational Psychology, 76*, 236–247.

Brewer, M. B. (1979). In-group bias in the minimal group situation: A cognitive motivational analysis. *Psychological Bulletin, 86*, 307–324.

Brewer, M. B. (1988). A dual-process model of impression formation. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Advances in social cognition* (pp. 1–36). Hillsdale, NJ: Erlbaum.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin, 76*, 15–38.

Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*, 631–661.

Brophy, J., & Good, T. (1974). *Teacher-student relationships: Causes and consequences.* New York: Holt, Rinehart and Winston.

Bruner, J. S. (1957). Going beyond the information given. In H. Gruber, K. R. Hammond, & R. Jessor (Eds.), *Contemporary approaches to cognition.* Cambridge, MA: Harvard University Press.

Bye, L. (1994). Referral of elementary age students for social skill training: Student and teacher characteristics. Doctoral dissertation, Rutgers University, New Brunswick, NJ.

Carlson, R. (1984). What's social about social psychology? Where's the person in personality research? *Journal of Personality and Social Psychology, 47*, 1304–1309.

Cooper, H. (1979). Pygmalion grows up: A model for teacher expectation, communication, and performance influence. *Review of Educational Research, 49*, 389–410.

Cooper, H., & Hazelrigg, P. (1988). Personality moderators of interpersonal expectancy effects: An integrative research review. *Journal of Personality and Social Psychology, 55*, 937–949.

Covington, M. V., & Omelich, C. L. (1979). Effort: The double-edged sword in school achievement. *Journal of Educational Psychology, 71*, 169–182.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*, 608–630.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52*, 177–193.

Darley, J., & Fazio, R. H. (1980). Expectancy-confirmation processes arising in the social interaction sequence. *American Psychologist, 35*, 867–881.

Darley, J., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44*, 20–33.

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review, 94*, 369–389.

DeParle, J. (1994, June 19). Welfare as we've known it. *The New York Times* (The Week in Review), p. 4.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Dornbusch, S. (1994, February). *Off the track.* Presidential address at the biennial meeting of the Society for Research on Adolescence, San Diego, CA.

Doyle, W. J., Hancock, G., & Kifer, E. (1972). Teachers' perceptions: Do they make a difference? *Journal of the Association for the Study of Perception, 7,* 21–30.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology, 34,* 590–598.

Dusek, J., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology, 75,* 327–346.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist, 50,* 145–158.

Eagly, A. H., Makhijani, M. G., Ashmore, R. D., & Longo, L. C. (1991). What is beautiful is good, but . . .: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109–128.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward men and women. *Personality and Social Psychology Bulletin, 15,* 543–558.

Eccles Parsons, J., et al. (1983). Expectancies, values and academic behaviors. In J. Spence (Ed.). *Achievement and achievement motivation* (pp. 75–146). San Francisco: W. H. Freeman.

Eccles, J. (1988). *Achievement beliefs and environment.* Final report to National Institute of Child Health and Development.

Eccles (Parsons), J. (1984). Sex differences in mathematics participation. In M. L. Maehr & M. W. Steinkamp (Eds.), *Women in science, Vol. 2, Advances in motivation and achievement* (pp. 93–137). Greenwich, CT: JAI Press.

Eccles (Parsons), J., Adler, T., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46,* 26–43.

Eccles, J., & Blumenfeld, P. (1985). Classroom experiences and student gender: Are there differences and do they matter? In L. C. Wilkinson & C. Marett (Ed.), *Gender influences in classroom interaction* (pp. 79–114). Hillsdale, NJ: Erlbaum.

Eccles, J. S., et al. (1993). School and family effects on the ontogeny of children's interests, self-perceptions, and activity choices. In R. Dienstbier & J. E. Jacobs (Eds.), *Developmental perspectives on motivation* (Vol. 40 of the Nebraska Symposium on Motivation) (pp. 145–208). Lincoln : University of Nebraska Press.

Eccles, J., & Harold, R. D. (1991). Gender differences in sports involvement: Applying the Eccles' expectancy-value model. *Journal of Applied Sports Psychology, 3,* 7–35.

Eccles, J., & Harold, R. D. (1992). Gender differences in educational and occupational patterns among the gifted. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development.* (pp. 3–29) Unionville, NY: Trillium Press.

Eccles, J., & Hoffman, L. W. (1984). Socialization and the maintenance of a sex-segregated labor market. In H. W. Stevenson & A. E. Siegel (Eds.), *Research in child development and social policy* (Vol. 1. pp. 367–420). Chicago: University of Chicago Press.

Eccles, J., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs, 11,* 367–380.

Eccles-Parsons, J., Kaczala, C. M., & Meece, J. L. (1982). Socialization of achievement attitudes and beliefs: Classroom influences. *Child Development, 53,* 322–339.

Eccles, J., & Wigfield, A. (1985). Teacher expectations and student motivation. In J. Dusek (Ed.), *Teacher expectancies* (pp. 185–226). Hillsdale, NJ: Erlbaum.

Eccles, J. S., Wigfield, A., Flanagan, C. A., Miller, C., Reuman, D., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality, 57,* 283–310.

Eden, D. (1984). Self-fulfilling prophecy as a management tool: Harnessing Pygmalion. *Academy of Management Review, 9,* 64–73.

Eden, D. (1986). OD and self-fulfilling prophecy: Boosting productivity by raising expectations. *Journal of Applied Behavioral Science, 22,* 1–13.

Eden, D., & Shani, A. B. (1982). Pygmalion goes to boot camp: Expectancy, leadership, and trainee performance. *Journal of Applied Psychology, 67,* 194–199.

Elashoff, J. D., & Snow, R. E. (1971). *Pygmalion reconsidered.* Worthington, OH: Charles A. Jones.

Esses, V. M., Haddock, G., & Zanna, M. P. (1993). Values, stereotypes, and emotions as determinants of intergroup attitudes. In D. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 137–166). San Diego: Academic Press.

Evertson, C. (1982). Differences in instructional activities in higher and lower achieving junior high English and math classes. *Elementary School Journal, 82,* 329–350.

Farina, A., Allen, J. G., & Saul, B. B. B. (1968). The role of the stigmatized person in affecting social relationships. *Journal of Personality, 36,* 169–182.

Farina, A., Gliha, D., Boudreau, L. A., Allen, J. G., & Sherman, M. (1971). Mental illness and the impact of believing others know about it. *Journal of Abnormal Psychology, 77,* 1–5.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304–341.

Fine, M., & Gordon, S. M. (1989). Feminist transformations of/despite psychology. In M. Crawford (Ed.), *Gender and thought* (pp. 146–174). New York: Springer-Verlag.

Finn, J. (1972). Expectations and the educational environment. *Review of Educational Research, 42,* 387–410.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). New York: Academic Press.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition.* Reading, MA: Addison-Wesley.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Fordham, S., & Ogbu, J. U. (1986). Black students' school success: Coping with the burden of "acting White." *The Urban Review, 18,* 176–206.

Fox, R. (1991). Prejudice and the unfinished mind: A new look at an old failing. *Psychological Inquiry, 3,* 137–152.

Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and income for men and women in management. *Journal of Applied Social Psychology, 21,* 1039–1057.

Funder, D. C. (1987). Errors and mistakes: Evaluatinig the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Funder, D. C. (1995). Stereotypes, base rates, and the fundamental evaluation mistake: A content-based approach to judgmental accuracy. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 141–156). Washington, DC: American Psychological Association.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgments. *Journal of Personality and Social Psychology, 55,* 149–158.

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62,* 411–422.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117,* 21–38.

Gilbert, D. T., & Osborne, R. E. (1989). Thinking backward: Some curable and incurable consequences of cognitive busyness. *Journal of Personality and Social Psychology, 57,* 940–949.

Goldman, W., & Lewis, P. (1977). Beautiful is good: Evidence that the physically attractive are more socially skillful. *Journal of Experimental Social Psychology, 13,* 125–130.

Gordon, R. A. (1968). Issues in multiple regression. *American Journal of Sociology, 73,* 592–616.

Graham, S. (1992). "Most of the subjects were White and middle class": Trends in published research on African Americans in selected APA journals, 1970–1989. *American Psychologist, 47,* 629–639.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology, 65,* 1105–1118.

Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-based expectancies: Effects on information processing and social behavior. *Journal of Social Issues, 46,* 35–60.

Harris, M. J. (1989). Personality moderators of expectancy effects: Replication of Harris and Rosenthal (1986). *Journal of Research in Personality, 23,* 381–387.

Harris, M. J., Milich, R., Corbitt, E. M., Hoover, D. W., & Brady, M. (1992). Self-fulfilling effects of stigmatizing information on children's social interactions. *Journal of Personality and Social Psychology, 63,* 41–50.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin, 97,* 363–386.

Harter, S., & Connell, J. P. (1984). A model of the relationship among children's academic achievement and their self-perceptions of competence, control, and motivational orientation. In J. G. Nicholls (Ed.), *Advances in motivation and achievement* (pp. 219–250). Greenwich, CT: JAI Press.

Heilman, M. E., Simon, M. C., & Repper, D. P. (1987). Intentionally favored, unintentionally harmed? The impact of gender-based preferential selection on self-perceptions and self-evaluations. *Journal of Applied Psychology, 72,* 62–68.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve.* New York: The Free Press.

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology, 38,* 369–425.

Hilton, J., & Darley, J. (1985). Constructing other persons: A limit on the effect. *Journal of Experimental Social Psychology, 21,* 1–18.

Hilton, J. L., & Darley, J. M. (1991). The effects of interaction goals on person perception. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 235–267). New York: Academic Press.

Jacobs, J. E., & Eccles, J. S. (1985). Gender differences in math ability: The impact of media reports on parents. *Educational Researcher, 14,* 20–25.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology, 63,* 932–944.

Jones, E. E. (1986). Interpreting interpersonal behavior: The effects of expectancies. *Science,* *234,* 41–46.

Jones, E. E. (1990). *Interpersonal perception.* New York: W. H. Freeman.

Jones, R. A. (1977). *Self-fulfilling prophecies: Social, psychological, and physiological effects of expectancies.* Hillsdale, NJ: Erlbaum.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* Orlando, FL: Harcourt Brace Jovanovich.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review, 100,* 109–128.

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review, 93,* 429–445.

Jussim, L. (1987). Interpersonal expectations in social interaction: Self-fulfilling prophecies, confirmatory biases, and accuracy. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International, 48,* 1845B.

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57,* 469–480.

Jussim, L. (1990). Social reality and social problems: The role of expectancies. *Journal of Social Issues, 46,* 9–34.

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review, 98,* 54–73.

Jussim, L. (1993). Accuracy in interpersonal expectations: A reflection-construction analysis of current and classic research. *Journal of Personality, 61,* 637–668.

Jussim, L., Coleman, L., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology, 52,* 536–546.

Jussim, L., & Eccles, J. (1992). Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology, 63,* 947–961.

Jussim, L., & Eccles, J. (1995). Naturalistic studies of interpersonal expectancies. *Review of Personality and Social Psychology, 15,* 74–108.

Jussim, L., & Eccles, J. (1995). Are teacher expectations biased by students' gender, social class, or ethnicity? In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 245–271). Washington, DC: American Psychological Association.

Jussim, L., & Fleming, C. (in press). Self-fulfilling prophecies and the maintenance of social stereotypes: The role of dyadic interactions and social forces. In N. Macrae, M. Hewstone, & C. Stangor (Eds.), *The foundations of stereotypes and stereotyping.* New York: Guilford.

Jussim, L., Madon, S., & Chatman, C. (1994). Teacher expectations and student achievement: Self-fulfilling prophecies, biases, and accuracy. In L. Heath, R. S. Tindale, J. Edwards, E. J. Posavac, F. B. Bryant, E. Henderson-King, Y. Suarez-Balcazar, J. Myers (Eds.), *Applications of heuristics and biases to social issues* (pp. 303–334). New York: Plenum.

Jussim, L., Nelson, T., Manis, M., & Soffin, S. (1995). Prejudice, stereotypes, and labeling effects: Sources of bias in person perception. *Journal of Personality and Social Psychology, 68,* 228–246.

Jussim, L., McCauley, C. R., & Lee, Y. T. (1995). Why study stereotype accuracy and inaccuracy? In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 3–27). Washington, DC: American Psychological Association.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology, 28*, 280–290.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102*, 390–402.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*, 198–214.

Krueger, J., & Rothbart, M. (1988). Use of categorical and individuating information in making inferences about personality. *Journal of Personality and Social Psychology, 55*, 187–195.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.

LaPiere, R. T. (1936). Type-rationalizations of group antipathy. *Social Forces, 15*, 232–237.

Lareau, A. (1987). Social-class differences in family-school relationships: The importance of cultural capital. *Sociology of Education, 60*, 73–85.

Lee, Y. T., & Ottati, V. (1993). Determinants of in-group and out-group perceptions of heterogeneity. *Journal of Cross-Cultural Psychology, 24*, 298–318.

Lerner, M. (1980). *The belief in a just world.* New York: Plenum.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology, 65*, 1178–1187.

Linville, P. (1982). The complexity-extremity effect and age-based stereotyping. *Journal of Personality and Social Psychology, 42*, 193–211.

Linville, P., & Jones, E. E. (1980). Polarized appraisal of out-group members. *Journal of Personality and Social Psychology, 38*, 689–703.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice-Hall.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology, 39*, 821–831.

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology, 18*, 23–42.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*, 215–238.

Mackie, M. (1973). Arriving at "truth" by definition: The case of stereotype inaccuracy. *Social Problems, 20*, 431–447.

Madon, S., Jussim, L., & Eccles, J. (1995). *In search of the powerful self-fulfilling prophecy.* Manuscript in preparation.

Marger, M. N. (1991). *Race and ethnic relations* (2nd ed.). Belmont, CA: Wadsworth.

Marger, M. N. (1994). *Race and ethnic relations* (3rd ed.). Belmont, CA: Wadsworth.

McCauley, C. (1995). Are stereotypes exaggerated? Looking for perceptual contrast and cognitive miser in beliefs about group differences. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 215–243). Washington, DC: American Psychological Association.

McCauley, C., Jussim, L., & Lee, Y. T. (1995). The time is now: Stereotype accuracy and intergroup relations. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 293–312). Washington, DC: American Psychological Association.

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology, 36*, 929–940.

McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin, 87*, 195–208.

McCauley, C. R., & Thangavelu, K. (1991). Individual differences in sex stereotyping of occupations and personality traits. *Social Psychology Quarterly, 54,* 267–279.

Meece, J. L., Eccles-Parsons, J., Kaczala, C. M., Goff, S. E., & Futterman, R. (1982). Sex differences in math achievement: Towards a model of academic choice. *Psychological Bulletin, 91,* 324–348.

Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review, 8,* 193–210.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81,* 247–258.

Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology* (Vol. 37, pp. 233–256). Stanford, CA: Annual Reviews.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38,* 379–387.

Nelson, T. E., Biernat, M. R., & Manis, M. (1990). Everybody base rates (sex stereotypes): Potent and resilient. *Journal of Personality and Social Psychology, 59,* 664–675.

Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology, 56,* 374–386.

Neuberg, S. L. (1994). Expectancy-confirmation processes in stereotype-tinged social encounters: The moderating role of social goals. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium* (Vol. 7). Hillsdale, NJ: Erlbaum.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality.* Cambridge, MA: Blackwell.

Olson, J. M., Roese, N. J., & Zanna, M. P. (in press). Expectancies. In E. T. Higgins & A. W. Kruglanski, (Eds.), *Social psychology: Handbook of basic principles.* New York: Guilford.

Ottati, V., & Lee, Y. T. (1995). Accuracy: A neglected component of stereotype research. In Y. T. Lee, L.Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 29–59). Washington, DC: American Psychological Association.

Palardy, J. M. (1969). What teachers believe—what children achieve. *Elementary School Journal, 69,* 370–374.

Parsons, J. E. (1980). *Final Report to the National Institute of Education,* Washington, DC (ERIC Document Reproduction Service No. ED 186 477).

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin, 5,* 461–476.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297–340). New York: Academic Press.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy inductions: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76,* 85–97.

Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review, 40,* 411–451.

Rosenthal, R. (1974). *On the social psychology of the self-fulfilling prophecy: Further evidence for Pygmalion effects and their mediating mechanisms.* New York: MSS Modular Publications.

Rosenthal, R. (1989, August). *Experimenter expectancy, covert comunication, & meta-analytic methods.* Invited address at the 97th annual convention of the American Psychological Association, New Orleans, LA.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectations and student intellectual development.* New York: Holt, Rinehart and Winston.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences, 3,* 377–386.

Ryan, C. (1995). Motivation and group membership: Consequences for stereotype accuracy. In L. Jussim, Y. T. Lee, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 189–214). Washington, DC: American Psychological Association.

Schuman, H., Walsh, E., Olson, C., & Etheridge, B. (1985). Effort and reward: The assumption that college grades are affected by quantity of study. *Social Forces, 63,* 945–966.

Sedikedes, C., & Skowronski, J. J. (1991). The law of cognitive structure activation. *Psychological Inquiry, 2,* 169–184.

Skrypnek, B. J., & Snyder, M. (1982). On the self-perpetuating nature of stereotypes about women and men. *Journal of Experimental Social Psychology, 18,* 277–291.

Snyder, M. (1984). When belief creates reality. In M. P. Zanna (Ed.), *Advances in experimental psychology,* (Vol. *18,* pp. 247–305). Orlando, FL: Academic Press.

Snyder, M. (1992). Motivational foundations of behavioral confirmation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. *25,* pp. 67–114). San Diego: Academic Press.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35,* 656–666.

Stangor, C. (1995). Content and application inaccuracy in social stereotyping. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 275–292). Washington, DC: American Psychological Association.

Steele, C. M. (1992). Race and the schooling of black Americans. *Atlantic Monthly,* April, 68–78.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues, 41,* 157–175.

Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91,* 457–477.

Swann, W. B., Jr., & Ely, R. J. (1984). A battle of wills: Self-verification versus behavioral confirmation. *Journal of Personality and Social Psychology, 46,* 1287–1302.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66,* 21–36.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52,* 1–23.

Tuckman, B. W., & Bierman, M. L. (1971). *Beyond Pygmalion: Galatea in the schools.* Paper presented at the meeting of the American Educational Research Association, New York City.

Vernon, P. E. (1933). Some characteristics of the good judge of personality. *Journal of Social Psychology, 4,* 42–58.

von Baeyer, C. L., Sherk, D. L., & Zanna, M. P. (1981). Impression management in the job interview: When the female applicant meets the male (chauvinist) interviewer. *Personality and Social Psychology Bulletin, 7,* 45–51.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. *27,* pp. 177–254). San Diego: Academic Press.

Weber, M. (1930). *The Protestant ethic and the spirit of capitalism.* New York: Scribner.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*, 3–25.

Wentzel, K. R. (1989). Adolescent classroom goals, standard for performance, and academic achievement: An interactionist perspective. *Journal of Educational Psychology, 81*, 131–142.

West, C., & Anderson, T. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research, 46*, 613–630.

Wigfield, A., Eccles, J. S., MacIver, D., Reuman, D., & Midgley, C. (1991). Transitions at early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*, 552–565.

Williams, T. (1976). Teacher prophecies and the inheritance of inequality. *Sociology of Education, 49*, 223–236.

Wineburg, S. S. (1987). The self-fulfillment of the self-fulfilling prophecy. *Educational Researcher, 16*, 28–37.

Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology, 10*, 109–120.

Yee, D., & Eccles, J. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles, 19*, 317–333.

Zanna, M. P., & Pack, S. J. (1975). On the self-fulfilling nature of apparent sex differences in behavior. *Journal of Experimental Social Psychology, 11*, 583–591.

Yee, Fairchild, Weizmann, & Wyatt, (1993). Addressing psychology's problems with race. *American Psychologist, 48*, 1132–1140.