DEVELOPMENTAL REVIEW 1, 187–206 (1981)

Sex Differences: Historical Perspectives and Methodological Implications

ANNE ANASTASI

Fordham University

Apart from its historical interest, early psychological research on sex differences provides methodological and interpretive insights and permits the comparison of sex differences under different societal conditions. Conclusions about sex differences from published studies should consider sample size, differential selection, overlapping of distributions, psychometric properties of the measuring instruments, and nature of the constructs employed as a basis for sex comparisons. In order to advance from a description of sex differences to an understanding of the operation of biological and cultural factors in their etiology, male and female differences in both cognitive and personality variables should be investigated under changing cultural conditions. Examples of etiological hypotheses that can be tested through such an approach are followed by some illustrative findings on changing sex differences over time. The need for well-designed cohort studies spanning critical periods of societal change is indicated.

According to Boring (1929, p. vii), it was Ebbinghaus who first remarked that psychology has a long past, but only a short history. To trace the origins and development of psychological research on any given topic, one does not need to go back very far in time. For psychological studies of sex differences, I have accordingly chosen to concentrate on the first 60 years of the 20th century, with occasional minor excursions into immediately preceding or following years. It was in the 1960s that, stimulated by the women's movement, research on sex differences assumed fresh vigor and began to explore some new directions. That decade thus provides a convenient marker to separate "then" from "now," to differentiate the historical background from the modern scene.

CONTRIBUTIONS OF EARLY STUDIES

Why should we re-examine this early research? One reason can certainly be found in the mounting interest in all aspects of the history of psychology. This interest is evidenced in recent writings not only by psychologists (e.g., Benjamin, 1977; Shields, 1975a, b; Wolf, 1973) but also by historians (e.g., Cravens, 1978; Sokal, 1973, 1980, 1981). The two articles by Shields (1975a) are especially relevant to the early research on

This article is an expanded version of a portion of a Master Lecture given at the 1979 Convention of the American Psychological Association, as part of the APA Continuing Education Program. Requests for reprints should be sent to Anne Anastasi, Psychology Department, Fordham University, Bronx, NY 10458.

0273-2297/81/030187-20$02.00/0
Copyright © 1981 by Academic Press, Inc.
All rights of reproduction in any form reserved.

sex differences. The book by Cravens (1978) provides a lucid and comprehensive picture of the social and scientific climate in which research on individual and group differences as a whole was conducted in America during the first 4 decades of the 20th century.

When dealing with sex differences, however (as when dealing with any other group or individual differences), we can identify more compelling reasons for studying the research of earlier decades. First, an analysis of earlier studies can sensitize us to the methodological and interpretive pitfalls that beset this field of research. We can learn both from the investigators who were misled by weak methodology and inadequate data, and from those research pioneers who designed ingenious ways of coping with methodological hazards.

Second, in research on group differences, if we want to go beyond description to an understanding of the causes of differences, we should not limit ourselves to one time period. We need cohort studies of psychological sex differences within a changing social context. We can think of this approach as the longitudinal study of populations (Anastasi, 1958a; pp. 209–211; 1982, Chap. 12). It is illustrated by comparisons of the test performance of cohorts of the same chronological age examined 10 or 20 or 30 years apart. A classic example is provided by the Scottish surveys (Scottish Council for Research in Education, 1949), in which the same intelligence test was administered to nearly all 11-year-old Scottish children in 1932 and to a similar sample of 11-year-olds in 1947. The findings of such a study can be used to test hypotheses about what is happening to a particular population over time; and they can be related to relevant societal changes occurring over the interval.

When observations on sex differences are limited to a relatively uniform and unchanging societal context, the findings are likely to remain at a descriptive level. We can tell only how the behavior of females and that of males compare under the existing conditions. For obvious ethical and practical reasons, research on human subjects cannot expose individuals to drastic and long-lasting variations in living conditions. In comparison, societal changes are likely to provide differences in experiential variables that are more extreme in degree, of longer duration, and more pervasive in their influence on psychological development than could be achieved by experimental manipulation. Hence cohort studies spanning one or more decades, during which significant social changes were under way, represent a valuable natural experiment—or at least a quasi-experiment. This type of study helps in analyzing the relation between cultural and behavioral variables. A comparison of findings obtained in the last quarter of this century with the published findings of earlier studies should contribute to an understanding of the origins of psychological sex differences.

The data obtained in the earlier studies are irreplaceable, insofar as the

societal conditions under which they were collected cannot be reproduced. A fundamental approach of differential psychology as a whole is to study a behavioral characteristic under differing cultural conditions. Such an approach contributes to an understanding of the operation of heredity and environment in the etiology of behavior in general. The method is essentially the same whether the source of variation is horizontal (across contemporary cultures or subcultures) or vertical (across time periods within a single culture). There have been scattered attempts to investigate sex differences in different cultures, from the early work of Mead (1930, 1935) in New Guinea and other studies in primitive societies (Berry, 1966; Schlegel, 1977) to the more recent surveys conducted in Israeli kibbutzim (Mednick, 1975, 1981; Spiro, 1979; Tiger & Shepher, 1975), in the Soviet Union (Lapidus, 1978), in the People's Republic of China (Tavris & Offir, 1977, Chap. 9; Wolf & Witke, 1975), in Scandinavian countries (Tavris & Offir, 1977, Chap. 9), and in Morocco (D. H. Dwyer, 1978). It is noteworthy that these cross-cultural studies provide meager objective, quantitative data on either aptitudes or personality characteristics. In general, cross-cultural studies present special methodological problems; and comparisons of results across widely different cultures are difficult. For these reasons, cohort studies across time would seem to be more productive.

To summarize, early psychological research on sex differences is of interest for at least three reasons: (1) it has historical significance; (2) it can provide methodological and interpretive insights; and (3) it permits comparisons between sex differences that develop under different societal conditions. The implications of the third point are twofold. In order to advance beyond mere description toward an understanding of the causes of observed sex differences, we should: (1) preserve and utilize the irreplaceable early data on sex differences; and (2) conduct systematic cohort studies of sex differences in the future, as the delayed effects of gradual societal changes in sex roles and status become manifest.

METHODOLOGICAL AND INTERPRETIVE INSIGHTS

We may now turn to some examples of methodological and interpretive insights that can be gleaned from the early literature on sex differences. Such insights pertain not only to the way research is conducted, but also to the way research literature is surveyed (see, e.g., Block, 1976a, b; Maccoby & Jacklin, 1974). How should the results of independent studies, with their many procedural differences, be integrated, in the effort to draw general conclusions from their findings? An examination of early studies on sex differences highlights several procedural variables to which the student of group differences should be alerted. The illustrations to be discussed have been grouped into three categories: sampling problems, the overlapping of distributions, and the nature of measuring instruments.

Sampling Problems

Statistical significance. With regard to sampling problems, we have all learned to look for statistically significant differences. That is because we want to be able to generalize from the sex difference found in the tested sample to the difference expected in the specified population that this sample represents. If the difference is *not* significant, it may indicate that there is no mean sex difference in the measured variable in the population (or the difference may be in the opposite direction). It may also indicate, however, that the sample was too small to demonstrate the difference. On the other hand, if the difference is statistically significant, it could mean that the sample is large enough to yield statistical significance even for a very small difference. The implication is that we should not draw conclusions by simply tallying the number of studies that yielded significant or insignificant differences. We need to look also at the magnitude of the differences—in terms of the scale of measurement, the size of the standard deviations, the practical meaning of a difference of such magnitude, or any other standard appropriate to the context. In comparing results from studies conducted at different time periods, or from studies using group tests with those using individual tests, large discrepancies in number of cases are likely to be found. It is particularly important under those conditions to take sample size into account in combining or comparing such studies.

Differential selection. Another methodological pitfall associated with sampling pertains to the differential operation of selective factors for males and females. Under these circumstances, the two sex samples are not equally representative of their respective populations and hence not comparable. Suppose that, because of some accident of availability, the men in a study come from the upper end of their distribution in the trait in question, while the women come from the lower end of theirs. The results would undoubtedly show a statistically significant mean difference in favor of the men; but this would not be a difference associated with sex, nor would it hold for the general population.

The effect of differential selective bias was illustrated by two early studies in which the same group intelligence test was administered to a sample of over 2500 elementary school pupils (Pressey, 1918) and to approximately 6000 high school students (Book & Meadows, 1928). In the elementary school sample, the girls excelled the boys in mean score at all ages; but among the high school seniors, the boys excelled. At the time these data were obtained, a favorite interpretation was that, since girls develop faster than boys physically, they are also accelerated in their mental development; and this acceleration would account for their temporary superiority during the elementary school years. By the time they reach the high school senior years, it was argued, boys have caught up

developmentally, and they accordingly excel girls in intelligence test performance.

An alternative explanation emerged, however, when the rate of high school dropouts was compared for the two sexes. The boys not only dropped out in much larger numbers than did the girls, but the dropouts also came largely from the low end of the distribution of academic achievement. Superficially, it would seem that samples of high school boys and girls attending the same classes should be comparable. The obtained score difference, however, resulted from a selective bias which acted differentially for the two sexes.

Another classical example of the unsuspected operation of differential sampling bias pertains to the hypothesis of sex differences in variability. Early writers on sex differences, including especially Ellis (1894/1934), Cattell (1903), and Thorndike (1914), maintained that, while men and women might have the same mean intelligence, the male sex showed wider variability. The evidence adduced in support of this hypothesis included the greater number of men in institutions for the mentally retarded, at one end of the distribution, and the greater incidence of eminent men in published directories, at the other end. In reference to the latter, critics of the hypothesis were quick to point out the conspicuous differences in opportunities available to men and women for attaining eminence (Hollingworth, 1914; Woolley, 1914).

The discrepancy at the low end of the scale was more difficult to explain and was widely attributed to a fundamental biological difference, possibly related to the greater frequency of sex-linked defects in the male. It now seems doubtful, however, that such defects could account for sizable discrepancies in the incidence of mental retardation in the two sexes. An alternative explanation can be found in the operation of selective bias in the institutionalization of mental retardates. The operation of this bias was first clearly demonstrated in a study by Hollingworth (1922). Essentially, this study showed that, while there were more mentally retarded males in institutions, there were more mentally retarded females outside. Among persons referred for mental examination, the females as a group were older and had lower IQs than the males. Both of these discrepancies were even greater when the cases actually committed were compared. A survey of the previous occupations and case histories of the subjects indicated that the probable explanation of these differences lay in the less demanding nature of the occupations and life patterns open to the mentally retarded females. Such a female could survive outside an institution by remaining at home with her family, or by turning to domestic service, prostitution, or marriage. Males, on the other hand, were more likely to be pushed into competitive industrial work at a relatively early age, where their mental retardation was soon revealed.

These two examples, one with high school seniors and one with mental retardates, point up the importance of investigating the possible operation of differential selection in any study of sex differences.

Overlapping Distributions

When Samuel Johnson was asked which is more intelligent, man or woman, he is said to have replied, "Which man, which woman?" This is a succinct way of expressing the wide individual differences found within each sex, with the resulting overlapping of their distributions. Since in any psychological trait women differ widely from one another, and men likewise vary widely among themselves, any relationship found between group means will not necessarily hold for individuals. Even when one group excels the other by a large and statistically significant amount, individuals can be found in the lower-scoring group who surpass certain individuals in the higher-scoring group. Because of the large extent of individual differences within any one group as contrasted to the relatively small difference between group means, a person's membership in a given group provides little or no information about his or her status in most traits.

Most published reports of sex differences focus on means or other group parameters. For a complete picture of the relative standing of the two groups, however, some index of the extent of overlapping should be included. The best procedure would be to reproduce the entire frequency distributions of the two groups. A simpler alternative, in the case of normally distributed samples, is to report the percentage of persons in one group who reach or exceed the median of the other. In terms of this index, complete overlapping is indicated when 50% of one group reach or exceed the median of the other and the ranges are the same. This is understandable when we remember that, within any one group, 50% of its own members reach or exceed its own median.

When more than 50% of Group A reach or exceed the median of Group B, then Group A is to that extent superior to Group B. When less than 50% of A reach or exceed the median of B, Group A is inferior to Group B. It should be borne in mind that overlapping is typically expressed with reference to the *median* of one group, not with reference to the *lowest* score. Thus even if 0% of Group A reaches or exceeds the median of Group B, there are still some individuals in Group A who equal or exceed the performance of certain individuals in Group B. Moreover, in distributions that deviate substantially from the normal curve, it is possible for the ranges to be identical even when the percentage of one group reaching or exceeding the median of the other falls considerably below 50.

In my *Differential Psychology* text (Anastasi, 1958a, p. 455) I reproduced a graph from an early study (Schiller, 1934), showing the complete

distribution curves of 189 boys and 206 girls in the third and fourth elementary school grades on a test of arithmetic reasoning. The mean difference in favor of the boys was substantial, amounting to 4.6 points out of a total of 60 points. This difference, significant at the .01 level, was equivalent to more than half a standard deviation of either group. In terms of the usual index of overlapping, only 28% of the girls reached or exceeded the boys' median. Yet the ranges of the two groups were virtually identical. In fact, the range of the boys' scores extended from 10 to 55, that of the girls from 13 to 55. The data about overlapping distributions provide a sort of empirical, statistically respectable underpinning for the current efforts to regard women, as well as men, as individuals. These data should serve as a corrective against the traditional tendency to "homogenize" women into narrowly defined sex roles—to use a very apt term introduced by Sandra and Daryl Bem (1970, p. 94).

Nature of Measuring Instruments

The third category concerns characteristics of the measuring instruments employed to compare the sexes. In drawing conclusions from any one study—and especially in pulling together the results of different studies to summarize what is known about sex differences—we need to examine carefully the particular tests or other observational procedures used by each investigator. Many features of the assessment techniques are obviously relevant to an evaluation of the findings. For example, when ratings are employed, the probable influence of social stereotypes on raters' judgments must be considered. The reliability of instruments (whether ratings or objective tests) must surely be looked into.

Item-selection procedures. Certain questions regarding the construction of instruments are also especially pertinent. For instance, were items selected with reference to sex differences in responses? If so, were they chosen so as to minimize or maximize sex differences? There are examples of both practices in published tests. The Stanford–Binet illustrates the former, while all the early masculinity–femininity (M–F) scales illustrate the latter (Constantinople, 1973; Terman & Miles, 1936). Among the best known examples of these early M–F scales are those incorporated in the following inventories: Strong Vocational Interest Blank, Minnesota Multiphasic Personality Inventory, California Psychological Inventory, and Guilford–Zimmerman Temperament Survey. The implications of these item-selection procedures is that neither type of instrument is appropriate for assessing sex differences in the variables it measures. The Stanford–Binet was constructed so as to obliterate sex differences in performance; the M–F scales were constructed so as to exaggerate sex differences in responses, since any items that failed to yield significant sex differences in the standardization sample were rejected. It should be

noted that these comments are not intended as criticisms of the test-construction procedures employed. In both cases, the procedures were appropriate for the purposes for which the instruments were designed. But failure to take these procedures into account could lead to misuses of the tests and misinterpretation of the results.

Global scores. A related interpretive problem arises from the use of global scores on any heterogeneous test, however constructed. Shortly after the development of the first intelligence tests, there was a flurry of studies comparing the sexes in "general intelligence." The resulting sex differences, whether reported as an IQ on an individual scale such as the Binet or as some other global score on group tests, were difficult to interpret. The direction and amount of sex difference depended in part on the extent to which test items were drawn from different content domains, such as verbal, numerical, or spatial. Since the early, traditional intelligence tests were quite heterogeneous in their content coverage, the resulting composite scores tended to blur or wipe out sex differences in performance. Apparently it was partly because of the failure of traditional intelligence tests to reveal large sex differences that some early investigators turned to the hypothesis of sex differences in variability.

With the development and widespread application of factor analysis, intellectual functions were sorted into distinct group factors or aptitudes (see Anastasi, 1982, Chap. 13). The most influential findings on the identification of such aptitudes undoubtedly stem from the research of Thurstone, which yielded a dozen or so primary mental abilities (Thurstone, 1938; Thurstone & Thurstone, 1941), and that of Guilford, which yielded over 100 abilities organized into his structure-of-intellect model (Guilford, 1967; Guilford & Hoepfner, 1971). It is certainly more fruitful to investigate sex differences in terms of these more clearly defined traits than in terms of the composite of diverse functions sampled by traditional intelligence tests. Both test development and studies of sex differences have made increasing use of the traits identified through factor-analytic research.

Broad group factors. Nevertheless, there are two questions that remain to be considered in assessing available sex-difference research. The first pertains to the breadth of group factors investigated in the comparison of male and female performances. If the categories into which intellectual functions are grouped are too broad, we may obtain the same canceling out or blurring effect found with the early intelligence tests. This blurring may occur within individual studies, if comparisons are limited to total scores in such broad areas as verbal, numerical, or spatial aptitudes. It is even more likely to occur, however, in surveys of different studies. Here, for example, the results obtained with many different kinds of numerical tests may be combined and a conclusion drawn about sex differences in a

broad area, which is labeled "quantitative aptitude" simply because all the tests involved numbers. Yet even in Thurstone's research, which yielded a relatively small number of abilities, separate factors were identified for number computation and for quantitative reasoning. There were also at least two factors in the verbal domain, including verbal comprehension and word fluency. The spatial-perception domain has shown evidence of several identifiable factors, including perceptual speed and accuracy, spatial orientation, and the manipulation of spatial relations, among others.

It is noteworthy that several studies have found sex differences in opposite directions within these content domains. For instance, while males excelled in quantitative reasoning, females excelled in numerical computation. Females excelled in word fluency and in language usage; but verbal comprehension tests yielded either no sex difference or a difference in favor of males, especially at the older age levels. Females excelled in perceptual speed and accuracy, while males excelled in most other spatial tests. These findings are still tentative and are themselves influenced by other concomitant variables such as age and educational level. But they recurred often enough in studies conducted before 1970 to merit careful scrutiny and further investigation (Anastasi, 1958a, Chap. 14; Garai & Scheinfeld, 1968, Chap. 7; Tyler, 1965, Chap. 10).

Trait organization. The second question is more basic. It concerns the actual organization of abilities and the identification of factors in males and females (Anastasi, 1970; Hyde, Geiringer, & Yen, 1975). There is an increasing body of data indicating the role of experiential conditions in the very formation of group factors. It is not only the level of performance in different abilities, but also the way performance is organized into distinct traits, that is influenced by experiential background. Differences in factor patterns have been found to be associated not only with sex but also with different cultures or subcultures, socioeconomic levels, age, amount of education, and types of school curricula (Anastasi, 1970). In general, there is evidence to suggest that greater differentiation of abilities into more specialized factors occurs in those domains in which the particular group excels. For example, several differentiated verbal factors may be identified in the female, but only one in the male; similarly, several spatial factors may emerge for males, but only one for females.

The factorial composition of the same objective task may differ among individuals with diverse experiential backgrounds (Anastasi, 1982, Chap. 13; Burns, 1980; Frederiksen, 1969; French, 1965). One reason for these individual differences may be found in the use of different methods to carry out the same task. Individuals with highly developed verbal abilities, for example, are likely to utilize verbal mediators to solve a mechanical or spatial problem; those whose experiences have been pre-

dominantly mechanical, on the other hand, are likely to follow a perceptual or spatial approach in solving the same problem.

The examples cited thus far have been limited to the cognitive domain. But the same two questions regarding excessive breadth of trait categories and sex differences in trait organization can and should also be raised about personality variables. Moreover, in the personality domain, it is often necessary to get down to the item level in research on both group differences and trait organization. In an early study of sex differences in introversion–extroversion, for example, Heidbreder (1927) found no significant sex differences in total scores; but she did find a significant difference when items were regrouped into those interfering with interpersonal relations and those interfering with efficient work. It should be noted that several recent examples of such behavioral specificity—and of the resulting need for caution in combining sex-difference results from different personality studies—can be found in an excellent article by Block (1976b).

When dealing with personality variables, we also have to consider the problem of situational specificity (Anastasi, 1982, Chap. 17; Mischel, 1968, 1969, 1973, 1979). While not a major concern in the assessment of aptitudes, the variability in an individual's behavior from one situational context to another has been receiving increasing attention in research on personality. And it is quite likely that situational context interacts with sex in the personality domain.

Conjoint Analysis of Procedural Variables and Reported Sex Differences

In concluding this discussion of methodological and interpretive insights that can be gained from early research, we may consider how published studies might be most effectively utilized. It is apparent that a box score or a simple tallying of studies that found a significant sex difference and those that did not is inappropriate. A better approach is the systematic analysis of the significance and magnitude of reported sex differences in relation to certain major procedural features of each study. For the purpose of this analysis, the sex difference can be treated as the dependent variable, while the independent variables are the relevant characteristics of each study. The latter could include such variables as number of cases; age, education, occupation, socioeconomic level, and ethnic background of subjects; date when behavior was assessed; reliability and construct validity of instruments employed; and degree of specificity of behavioral assessment (e.g., analysis of subtest scores, item responses, situational context).

This approach is quite similar to what Glass calls "meta-analysis" (Glass, 1976, 1977; Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Kulik, Kulik, & Cohen, 1979; Smith & Glass, 1977). To be sure, available

published research on sex differences may not permit extensive application of the precise statistical procedures of meta-analysis described by Glass. For one thing, much of the information on sex differences was obtained incidentally in studies designed for other purposes. Procedures are quite diverse, and any ordering of studies in terms of procedural variables would probably be too crude for quantitative analyses. Nevertheless, at this stage we can at least be guided by the general orientation of meta-analysis in evaluating and utilizing the findings of particular studies.

SEX DIFFERENCES IN RELATION TO SOCIETAL CHANGES

If we want to advance beyond the description of existing sex differences to an explanation of the origins of these differences, we need to consider the operation of what has been variously designated as heredity and environment, or nature and nurture, or biology and culture. During the first half of the 20th century, the changing model of the heredity–environment relation led to heated controversies, first in terms of *which behaviors* were inherited and which acquired, and more recently in the effort to estimate how much of the variance of particular traits was attributable to heredity and how much to environment (Cravens, 1978). In the 1950s, I suggested that a more fruitful approach was provided by the question "How?" (Anastasi, 1958b). What can be learned about the actual *modus operandi* of specific hereditary and environmental factors in the development of behavioral differences? What is the chain of events whereby hereditary and environmental variables interact in the etiology of a particular behavioral outcome? The study of sex differences offers some promising opportunities to explore such etiological mechanisms.

Examples of Proposed Etiological Hypotheses

A consideration of biological sex differences suggests some intriguing possibilities regarding the role that such physical differences may play in the development of aptitude and personality differences within particular cultural contexts. Some of these possible etiological mechanisms were proposed in early surveys of sex differences (Scheinfeld, 1943; Seward, 1946). The operation of such mechanisms in turn suggests changes in trait differences between the sexes that may occur as societal conditions change.

One example is provided by the developmental acceleration of the female, a well-established sex difference that begins before birth and extends to maturity (Anastasi, 1958a; Garai & Scheinfeld, 1968; Tyler, 1965). The psychological effects of this difference in developmental rate probably vary widely from trait to trait. For instance, the developmental acceleration of girls in infancy has been repeatedly proposed as a signifi-

cant factor in their more rapid progress in the acquisition of language, which may give them a head start in verbal development as a whole. Another possible implication of developmental acceleration is a social one. Because of their physical acceleration, adolescent girls have traditionally tended to associate socially with boys older than themselves (Scheinfeld, 1943). This probably accounts also for the common age discrepancy in marriage. Since the girl was generally younger than the boys she dated—and younger than the man she married—she was surpassed by most of her male associates in amount of education and general experience. The resulting age differences in knowledge and information may have been perceived and fostered as a sex difference and could thus be at the root of many social attitudes and sex stereotypes. With the sharp increase in coeducation over the decades, the resulting daily contact with age peers of the opposite sex may serve as a corrective for this traditional misperception and may eventually be reflected in an equalization of expectations, attitudes, and self-concepts.

Another conspicuous set of biological sex differences pertains to general body size, muscular strength, and speed and coordination of gross bodily movements, in all of which males excel (Anastasi, 1958a; Garai & Scheinfeld, 1968; Tyler, 1965). Of course, in this case as in all other traits, we must not lose sight of the overlapping of distributions. But the *mean* sex differences in these physical characteristics are certainly large and consistent. Sex differences in gross motor coordination, for example, have been noted from infancy and tend to increase throughout childhood (Gesell et al., 1940).

In addition to the probable contribution of these physical differences to the development of aggressiveness, in which sex differences have been consistently reported (Anastasi, 1958a; Garai & Scheinfeld, 1968; Maccoby & Jacklin, 1974; Tyler, 1965), some early writers discussed their implications for occupational specialization. Surveys of the occupational distribution of males and females in contemporary primitive cultures, as well as historical records, support the relation between the physical demands of occupational roles and their frequency among males and females (Murdock, 1937; Scheinfeld, 1943; Seward, 1946). The authors of these surveys, among others, further observed that in the more advanced cultures, the physical demands of occupations have steadily decreased following the introduction of machinery. In modern technological societies, occupational achievement and success show a progressively diminishing dependence on sex-related physical skills. What will be the effect of such a change in occupational criteria on the psychological concepts that men and women have about themselves and others? This suggests another area for comparative exploration over time.

From a different angle, one could also investigate changes in male and

female scores on the traits of aggressiveness and dominance as assessed by self-report inventories, in relation to a diminishing societal tolerance for violent methods of conflict resolution. A plausible hypothesis would be that male means in aggressiveness should drop and female means in dominance rise, with a consequent decrease in the sex difference in both traits.

Another type of relationship that may be reflected in sex differences is that between motivational, emotional, and attitudinal variables on the one hand and aptitudes and achievement on the other. As the former alter under the impact of societal changes, they may lead eventually to corresponding modifications in the latter. That not only immediate achievement but also the long-term development of aptitudes is influenced by the individual's motivation and related noncognitive variables has been repeatedly demonstrated, notably in the continuing research of Atkinson and his associates (Atkinson, 1974; Atkinson & Birch, 1978, Chap. 4; Atkinson, O'Malley, & Lens, 1976; see also Anastasi, 1982, Chap. 12). According to Atkinson's schema, motivation affects both the efficiency with which a task is performed and the time the individual devotes to the task (e.g., studying different subjects, carrying out a job-related activity). The final achievement or product shows the combined effects of level of performance and time spent on task. An important consequent of level of performance $\times$ time spent on task is the lasting cumulative effects of this activity or experience on the individual's own cognitive and noncognitive development.

With particular reference to sex differences, several investigations have yielded suggestive data on the relationship between motivational or attitudinal variables and test performance. For example, an early study on sex-role identification and problem-solving ability (Milton, 1957) found that total problem-solving scores correlated significantly with Terman–Miles masculinity–femininity scores, not only in the combined-sex group but also within each sex. In other words, the women with the more masculine scores on the M–F scale were better problem solvers than were those with the more feminine scores; and the same relation held among the men.

More recent studies have contributed further data on the relation between the individual's attitudes, sex-role standards, and sex-role identification and such variables as academic achievement in reading and arithmetic. The procedures of these studies are extremely varied; and the results are scattered and disparate, but they nevertheless provide provocative leads. For instance, in a study of schoolchildren from the 2nd to the 12th grade, reading and arithmetic scores on standardized achievement batteries proved to be more a function of the child's perception of these areas as sex-appropriate or sex-inappropriate than of the child's own

sex, individual preference for masculine or feminine sex role, or personal liking or disliking for reading or arithmetic (C. A. Dwyer, 1974). In another study (Fitzpatrick, 1978), the achievement of bright 10th-grade girls in mathematics, as assessed by both grades and standardized tests, was found to be significantly related to the students' attitudes toward various aspects of the female role. Those girls showing a more liberal orientation on the women's role scale performed better in mathematics than did those with a more traditional orientation. It would thus seem desirable to investigate also aptitude changes over time, since these changes may be the indirect result of socially instigated attitudinal and motivational shifts.

Some Scattered Comparisons of Sex Differences over Time

A few investigators have followed a retrospective approach by gathering sex-difference data that were roughly comparable to available earlier data. For instance, children's preferences for play activities surveyed in the 1950s were compared with the findings of a similar survey conducted in 1926 (Rosenberg & Sutton-Smith, 1960). In the later study, the girls retained their interest in most of the female play activities of the earlier study, but in addition showed an increased preference for some traditional male activities. Boys' preferences, however, revealed no such broadening, but actually had become somewhat more confined. The authors interpreted the results as consistent with an expansion in the girls' role concept to include more traditional masculine activities, without a corresponding change in the boys' role perception.

In the early 1970s, Spence (1974) used TAT-like stories, followed by a structured questionnaire, to investigate women's attitudes toward achievement. The results were compared with those obtained a decade earlier by Horner with the same TAT-like stories (see Horner, 1972). The more recent findings suggested that the so-called fear of success, or motive to avoid success, was not so prevalent in the 1970s as in the 1960s.

A third example is from the cognitive domain (Cunningham & Birren, 1976). Although the number of cases employed in this study is small, the experimental design is well controlled and the findings are suggestive. Essentially, the study involved a three-way comparison of 32 cases. Out of a sample of 485 undergraduates tested in 1944, 32 were identified and retested in 1972. In addition, the same test was administered in 1972 to 32 undergraduates individually matched with the original 32 in sex and in total standard score within their own cohort. Although the test used was the Army Alpha, performance was analyzed in terms of a verbal factor, a highly speeded relations factor, and a number factor. Only the number factor showed a different developmental pattern for the sexes. In this factor, the female scores were 1 *SD* higher for the 1972 cohort than for the 1944 cohort, while those of the males showed no significant difference.

Neither males nor females showed any significant change in the number factor when the 1944 cohort was retested in 1972. The authors suggest that the cohort difference in the females may reflect a decline in "traditional stereotypes regarding female occupational roles, with a consequent shift in patterns of education" (p. 82).

The study just described utilized a simplified version of a sophisticated experimental design first fully described by Schaie (1965). Designated as the cross-sequential method, it involves the testing of a representative sample of different age groups on two or more occasions. This permits *cross-sectional* comparisons of different age groups tested at one time period, *longitudinal* comparisons of the same individuals tested at two or more time periods, and *time-lag* comparisons of samples of equal age drawn from different cohorts tested at the different time periods—for instance, 30-year-olds tested in 1950 could be compared with 30-year-olds tested in 1960. The last comparison is similar to what was described earlier in this paper as the longitudinal study of populations.

A large-scale pioneer application of the cross-sequential model is illustrated in an investigation by Schaie and his co-workers (Schaie & Labouvie-Vief, 1974; Schaie, Labouvie, & Buech, 1973; Schaie & Strother, 1968). The study began in 1956 with the administration of the SRA Primary Mental Abilities test and Schaie's Test of Behavioral Rigidity to a stratified-random sample of 500 adults. The population from which this sample was drawn consisted of approximately 18,000 members of a prepaid medical plan, whose membership was fairly representative of the census figures for a large metropolitan area. The sample included 25 men and 25 women at each 5-year age interval from 20 to 70. All the original participants who could be located 7 years later were contacted and 302 of them were given the same tests again. A second 7-year retest was administered in 1970 to 161 of the original participants (Schaie & Labouvie-Vief, 1974). In addition, the same tests were given to three independent age-stratified samples drawn from the same population in 1956, 1963, and 1970 (Schaie et al., 1973).

Although the principal object of this study was to analyze age changes, data were also available on sex differences in each cohort. The results revealed virtually no significant change in the relative performance of the sexes either across age or across cohorts. The pattern of sex differences followed the traditional results. For example, the most conspicuous sex differences were in spatial aptitude in favor of males and in verbal fluency in favor of females. The authors themselves suggest that their findings are consistent with the hypothesis that sex differences in cognitive functions "are either genetically determined or established by early imprinting and when thus established are maintained throughout life" (Schaie et al., 1973, p. 164).

In connection with the "early imprinting," it is relevant to note that the

mean birth dates of the cohorts covered by this study ranged for 1889 to 1945. The participants were therefore reared during periods when changes in sex roles were minor relative to the changes characteristic of the late 1960s and the 1970s. It would be of particular interest to investigate sex differences in cohorts born between 1945 and 1965. In their discussion of age and cohort differences, Schaie and his associates point to the need for research involving a conjoint analysis of environmental and behavioral change (Schaie et al., 1973; Schaie & Labouvie-Vief, 1974). This is the type of analysis that would be especially appropriate in the assessment of sex differences during a period when relevant societal attitudes and opportunities were undergoing conspicuous change.

More recent work by some investigators following the general cross-sequential approach was focused more directly on sex-role changes. Urberg and Labouvie-Vief (1976) administered the Adjective Check List (Gough & Heilbrun, 1965) to 40 males and 40 females in each of three groups: 7th-grade students, 12th-grade students, and students in adult education classes (ages 20 to 74). In a balanced design, half of the females and half of the males in each of the three groups were instructed to check the adjectives describing their concept of an ideal female; the other half were asked to do the same for an ideal male. The study was replicated with an independent sample 2 years later (Urberg, 1979). With only a few significant exceptions, the results showed the pervasiveness of traditional sex-role stereotypes across all age groups; nor was any change in the prevalence of these stereotypes found in the 2-year follow-up. In the self-descriptions, which were obtained only in the follow-up, however, the male descriptions did not differ significantly from the female descriptions (with a single minor exception). As for the lack of change in the ideal descriptions, it should be noted that 2 years represents a very short period over which to investigate cohort changes.

Cohort Studies and Societal Change: Future Prospects

If we wish to investigate the effects of a changing societal context on sex differences in behavior, we ought to compare cohorts assessed by equivalent procedures at different historical periods. Nor need this approach be limited to sex differences. Large-scale cohort studies, with representative samples and well-designed instruments, could serve a useful function in providing *psychological indicators*. They could show what is happening to the attitudes, interests, values, and psychological well-being of a population over time.

We have become accustomed to economic indicators, which record changes in the gross national product, the consumer price index, and other measures of the economic health of the nation. We are beginning to witness the emergence of social indicators, concerned with the quality of

life and the general well-being of the population (Sheldon, 1976; Strumpel, 1976), and educational indicators, concerned with the state of the nation's educational system and the knowledge that young people have acquired in selected subject-matter areas (Gooler, 1976; Womer, 1970). Perhaps the addition of psychological indicators to this list would help to unify and add meaning to the other indicators, insofar as what happens to societal variables is both a result and a cause of what happens to the behavioral variables of the constituent individuals.

REFERENCES

Anastasi, A. *Differential psychology* (3rd ed.). New York: Macmillan, 1958. (a)

Anastasi, A. Heredity, environment, and the question "How?" *Psychological Review*, 1958, **65**, 197–208. (b)

Anastasi, A. On the formation of psychological traits. *American Psychologist*, 1970, **25**, 899–910.

Anastasi, A. *Psychological testing* (5th ed.). New York: Macmillan, 1982.

Atkinson, J. W. Motivational determinants of intellective performance and cumulative achievement, In J. W. Atkinson & J. O. Raynor, *Motivation and achievement*. Washington, D.C.: Winston, 1974. Chap. 20.

Atkinson, J. W., & Birch, D. *An introduction to motivation* (2nd ed.). New York: Van Nostrand, 1978.

Atkinson, J. W., O'Malley, P. M., & Lens, W. Motivation and ability: Interactive psychological determinants of intellective performance, educational achievement, and each other. In W. H. Sewell, R. M. Hauser, & D. L. Featherman (Eds.), *Schooling and achievement in American society*. New York: Academic Press, 1976. Chap. 8.

Bem, S. L., & Bem, D. J. Case study of a nonconscious ideology: Training the woman to know her place. In D. J. Bem, *Beliefs, attitudes, and human affairs*. Belmont, Calif.: Brooks/Cole, 1970. Pp. 89–99.

Benjamin, L. T., Jr. The psychological round table: Revolution of 1936. *American Psychologist*, 1977, **32**, 542–549.

Berry, J. W. Temme and Eskimo perceptual skills. *International Journal of Psychology*, 1966, **1**, 207–229.

Block, J. H. Debatable conclusions about sex differences: Review of E. E. Maccoby & C. N. Jacklin, *The psychology of sex differences*. *Contemporary Psychology*, 1976, **21**, 517–522. (a)

Block, J. H. Issues, problems, and pitfalls in assessing sex differences: A critical review of *The psychology of sex differences*. *Merill-Palmer Quarterly*, 1976, **22**, 283–308. (b)

Book, W. F., & Meadows, J. L. Sex differences in 5925 high school seniors in ten psychological tests. *Journal of Applied Psychology*, 1928, **12**, 56–81.

Boring, E. G. *A history of experimental psychology*. New York: Century, 1929.

Burns, R. B. Relation of aptitudes to learning at different points in time during instruction. *Journal of Educational Psychology*, 1980, **72**, 785–795.

Cattell, J. McK. A statistical study of eminent men. *Popular Science Monthly*, 1903, **62**, 359–377.

Constantinople, A. Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin*, 1973, **80**, 389–407.

Cravens, H. *The triumph of evolution: American scientists and the heredity-environment controversy, 1900–1941*. Philadelphia: Univ. of Pennsylvania Press, 1978.

Cunningham, W. R., & Birren, J. E. Age changes in human abilities: A 28-year longitudinal study. *Developmental Psychology*, 1976, **12**, 81–82.

Dwyer, C. A. Influence of children's sex role standards on reading and arithmetic achievement. *Journal of Educational Psychology*, 1974, **66**, 811–816.

Dwyer, D. H. *Images and self-images: Male and female in Morocco*. New York: Columbia Univ. Press, 1978.

Ellis, H. *Man and woman, a study of secondary and tertiary sexual characteristics* (8th rev. ed.). London: Heinemann, 1934. (1st ed., 1894.)

Fitzpatrick, J. L. Academic underachievement, other-direction, and attitudes toward women's roles in bright adolescent females. *Journal of Educational Psychology*, 1978, **70**, 645–650.

Frederiksen, C. H. Abilities, transfer, and information retrieval in verbal learning. *Multivariate Behavioral Research Monographs*, 1969, No. 69-2.

French, J. W. The relationship of problem-solving styles to the factor composition of tests. *Educational and Psychological Measurement*, 1965, **25**, 9–28.

Garai, J. E., & Scheinfeld, A. Sex differences in mental and behavioral traits. *Genetic Psychology Monographs*, 1968, **77**, 169–299.

Gesell, A., et al. *The first five years of life*. New York: Harper, 1940.

Glass, G. V. Primary, secondary, and meta-analysis of research. *The Educational Researcher*, 1976, **10**, 3–8.

Glass, G. V. Integrating findings: The meta-analysis of research. In L. Schulman (Ed.), *Review of research in education*. Itasca, Ill.: Peacock, 1977.

Gooler, D. D. The development and use of educational indicators. *Proceedings of the 1975 ETS Invitational Conference*. Princeton, N.J.: Educational Testing Service, 1976. Pp. 11–27.

Gough, H. G., & Heilbrun, A. B. *The Adjective Check List manual*. Palo Alto, Calif.: Consulting Psychologists Press, 1965.

Guilford, J. P. *The nature of human intelligence*. New York: McGraw–Hill, 1967.

Guilford, J. P., & Hoepfner, R. *The analysis of intelligence*. New York: McGraw–Hill, 1971.

Heidbreder, E. Introversion and extroversion in men and women. *Journal of Abnormal and Social Psychology*, 1927, **22**, 52–61.

Hollingworth, L. S. Variability as related to sex differences in achievement. *American Journal of Sociology*, 1914, **19**, 510–530.

Hollingworth, L. S. Differential action upon the sexes of forces which tend to segregate the feebleminded. *Journal of Abnormal and Social Psychology*, 1922, **17**, 35–57.

Horner, M. S. The motive to avoid success and changing aspirations of college women. In J. M. Bardwick (Ed.), *Readings on the psychology of women*. New York: Harper & Row, 1972. Pp. 62–67.

Hyde, J. S., Geiringer, E. R., & Yen, W. M. On the empirical relation between spatial ability and sex differences in other aspects of cognitive performance. *Multivariate Behavioral Research*, 1975, **10**, 289–307.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 1981, **89**, 47–62.

Kulik, J. A., Kulik, C-L. C., & Cohen, P. A. A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist*, 1979, **34**, 307–318.

Lapidus, G. W. *Women in Soviet society: Equality, development, and social change*. Berkeley: Univ. of California Press, 1978.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, Calif.: Stanford Univ. Press, 1974.

Mead, M. *Growing up in New Guinea*. New York: Morrow, 1930.

Mead, M. *Sex and temperament in three primitive societies*. New York: Morrow, 1935.

Mednick, M. T. S. Social change and sex-role inertia: The case of the kibbutz. In M. T. S. Mednick, S. S. Tangri, & L. W. Hoffman (Eds.), *Women and achievement: Social and motivational analyses*. New York: Halsted Press, 1975. Pp. 85–103.

Mednick, M. T. S. The revolution that never was: Review of M. E. Spiro, *Gender and culture: Kibbutz women revisited*. *Contemporary Psychology*, 1981, **26**, 101–102.

Milton, G. A. The effects of sex-role identification upon problem-solving skills. *Journal of Abnormal and Social Psychology*, 1957, **55**, 208–212.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Mischel, W. Continuity and change in personality. *American Psychologist*, 1969, **24**, 1012–1018.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, **80**, 252–283.

Mischel, W. On the interface of cognition and personality: Beyond the person-situation debate. *American Psychologist*, 1979, **34**, 740–754.

Murdock, G. P. Comparative data on the division of labor by sex. *Social Forces*, 1937, **15**, 551–553.

Pressey, L. W. Sex differences shown by 2544 schoolchildren on a group scale of intelligence, with special reference to variability. *Journal of Applied Psychology*, 1918, **2**, 323–340.

Rosenberg, B. G., & Sutton-Smith, B. A revised conception of masculine-feminine differences in play activities. *Journal of Genetic Psychology*, 1960, **96**, 165–170.

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, **64**, 92–107.

Schaie, K. W., Labouvie, G. V., & Buech, B. U. Generational and cohort-specific differences in adult cognitive functioning; A fourteen-year study of independent samples. *Developmental Psychology*, 1973, **9**, 151–166.

Schaie, K. W., & Labouvie-Vief, G. Generational versus ontogenetic components of change in cognitive behavior: A fourteen-year cross-sequential study. *Developmental Psychology*, 1974, **10**, 305–320.

Schaie, K. W., & Strother, C. R. A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin*, 1968, **70**, 671–680.

Scheinfeld, A. *Women and men*. New York: Harcourt, Brace, 1943.

Schiller, B. Verbal, numerical, and spatial abilities of young children. *Archives of Psychology*, 1934, No. 161.

Schlegel, A. (Ed.). *Sexual stratification: A cross-cultural view*. New York: Columbia Univ. Press, 1977.

Scottish Council for Research in Education. *The trend of Scottish intelligence*. London: Univ. of London Press, 1949.

Seward, G. H. *Sex and the social order*. New York: McGraw–Hill, 1946.

Sheldon, E. B. The social indicators movement. *Proceedings of the 1975 ETS Invitational Conference*. Princeton, N.J.: Educational Testing Service, 1976. Pp. 3–10.

Shields, S. A. Functionalism, Darwinism, and the psychology of women: A study in social myth. *American Psychologist*, 1975, **30**, 739–754. (a)

Shields, S. A. Ms. Pilgrim's progress: The contributions of Leta Stetter Hollingworth to the psychology of women. *American Psychologist*, 1975, **30**, 852–857. (b)

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 1977, **32**, 752–760.

Sokal, M. M. APA's first publication: *Proceedings of the American Psychological Association, 1892–1893*. *American Psychologist*, 1973, **28**, 277–292.

Sokal, M. M. *Science* and James McKeen Cattell, 1894 to 1945. *Science*, 1980, **209**, 43–52.

Sokal, M. M. (Ed.). *An education in psychology: James McKeen Cattell's journal and*

letters from Germany and England, 1880–1888. Cambridge, Mass.: MIT Press, 1981.

Spence, J. T. The Thematic Apperception Test and attitudes toward achievement in women; A new look at the motive to avoid success and a new method of measurement. *Journal of Consulting and Clinical Psychology,* 1974, **42,** 427–437.

Spiro, M. E. *Gender and culture: Kibbutz women revisited.* Durham, N.C.: Duke Univ. Press, 1979.

Strumpel, B. (Ed.). *Economic means for human needs: Social indicators of well-being and discontent.* Ann Arbor, Mich.: Institute for Social Research, 1976.

Tavris, C., & Offir, C. *The longest war: Sex differences in perspective.* New York: Harcourt Brace Jovanovich, 1977.

Terman, L. M., & Miles, C. C. *Sex and personality: Studies in masculinity and femininity.* New York: McGraw–Hill, 1936.

Thorndike, E. L. *Educational psychology* (Vol. 3). New York: Teachers College, Columbia University, 1914.

Thurstone, L. L. Primary mental abilities. *Psychometric Monographs,* 1938, No. 1.

Thurstone, L. L., & Thurstone, T. G. Factorial studies of intelligence. *Psychometric Monographs,* 1941, No. 2.

Tiger, L., & Shepher, J. *Women in the kibbutz.* New York: Harcourt Brace Jovanovich, 1975.

Tyler, L. E. *The psychology of human differences* (3rd ed.). New York: Appleton–Century–Crofts, 1965.

Urberg, K. A. Sex role conceptualization in adolescents and adults. *Developmental Psychology,* 1979, **15,** 90–92.

Urberg, K. A., & Labouvie-Vief, G. Conceptualizations of sex roles: A life span developmental study. *Developmental Psychology,* 1976, **12,** 15–23.

Wolf, M., & Witke, R. (Eds.). *Women in Chinese society.* Stanford, Calif.: Stanford Univ. Press, 1975.

Wolf, T. H. *Alfred Binet.* Chicago: Univ. of Chicago Press, 1973.

Womer, F. B. *What is National Assessment?* Ann Arbor, Mich.: National Assessment of Educational Progress, 1970.

Woolley, H. T. The psychology of sex. *Psychological Bulletin,* 1914, **11,** 353–379.

RECEIVED: November 17, 1980; REVISED April 7, 1981