1986

What We Don't Know About Sex Differences

Joseph Adelson

WHAT WE DON'T KNOW ABOUT SEX DIFFERENCES

*Joseph Adelson**

I can think of few activities more enervating emotionally than to survey the psychological literature on sex differences. I first did so about 15 years ago, soon after the birth of contemporary American feminism, and was taken aback by the gap between the actual, enfeebled state of our knowledge, and the dogmatic self-assertion of so much then being written on the topic.

Since that time, things both have and have not changed. What has changed is the sheer quantity of work done or in progress, a dearth having become a glut; what has not changed is our depth of understanding. The abundance of new data has produced no breakthroughs, no new insights, and few bases for changes of heart or mind. If you believed fifteen years ago that sex differences, the important ones, were at bottom biological in origin, you would have no compelling reason to believe otherwise today. If you were a doctrinaire environmentalist then, you would still be so today. In either case, you could muster far more support for your position than previously, as in fact you could for all positions between the extremes. So one major reason to be dispirited is the strong sense one gets of a discipline merely treading water—it is depressing to read through dozens of laborious articles reporting minor variations on this or that empirical theme, to find that in the end they add up to little in the way of enhanced understanding.

To some considerable degree, these problems reflect a larger set of problems we find in social science generally: given a complex topic, it is extraordinarily difficult to obtain secure, non-trivial findings, and to articulate compelling or even heuristic theoretical models. The optimism we once had about the powers of social science, our belief that it would soon clarify and help resolve a wide range of social and psychological troubles—that optimism, so strongly felt in the 1950's and 1960's, now seems heady if not utopian.

Yet these generic problems are compounded, within the do-

* Professor of Psychology, University of Michigan. This article originally appeared in the Spring, 1985, issue of NEW PERSPECTIVES.

main of sex differences research, by the tendentious intentions of so many investigators. It is not merely that so many seem drawn to the topic because of an overweening personal interest; beyond that, we see a steady erosion of that necessary line between scientific disinterestedness and ideological purpose. It is not true in all cases, perhaps not even in most; but it is common enough to force a wariness upon the reader who can no longer assume that some tainting or tilting either of facts or interpretation, whether witting or unwitting, is an event either rare or guarded against. Indeed, some scholars in this area have become quite open about their intentions, proclaiming that their aim is to support their "values" through the medium of research. Many others are not quite so bold publicly, but will talk freely in private about what they expect their findings to demonstrate, and how it will lead to the betterment of humankind. Still others take to wearing two hats, playing the role of the objective scientist part of the time, and that of the activist while speaking to the press or in public appearances. In all of these instances we find a touching faith in the power of the scientific method to help keep one's partisan passions at bay, a faith often misplaced. Yet in some cases, as we will see later, the belief in the scientific method itself has waned, there being the view that science is masculinist, thereby keeping us from larger and truer truths.

One would like to believe that so far not too much has been lost. The more carefully refereed journals, like *Psychology Bulletin*, are unlikely to print egregiously biased articles, and those journals remain the most prestigious. That is indeed comfort, but rather smaller than it may appear, since only a fraction of the total range of research is subjected to close scrutiny. Futhermore, neither trade nor textbooks are carefully refereed, from the point of view of scholarly balance, and these tend to receive favorable attention in both the scholarly and general press, as long as they follow current fashion. Still futher, those books tend to be selected for use in college-level courses in women's studies, many of which are exercises in political indoctrination. Thus we find texts which seem given over monotonously to complaints or rationalizations, all differences between the sexes, including those seemingly favorable to women, seen as connected to discrimination or to invidious forms of child-rearing. In one recent text, for example, the male advantage in spatial perception is seen as an outcome of girls' being dissuaded from participation in sports, whereas the female advantage in verbal ability evokes the comment that despite this advantage, women are not allowed to use these gifts to the fullest.

The problem of bias has been exacerbated by the dubious posi-

tions taken by some of the scholarly associations, now given over to the propagation of liberal or leftist political causes. Much of the American Orthopsychiatric Association's annual program is devoted to agit-prop; issues on the feminist agenda are not given objective discussion. That division of the American Psychological Association presumably devoted to the study of feminine psychology has shown a quick and unseemly devolution from the sponsorship of research to the promotion of feminist causes, including some which do not, logically speaking, have much to do with feminism *per se*, such as the support of both male and female homosexuality. The sad fact is that it has become nearly impossible to tell the difference between a scholarly meeting and a political caucus, and what is worse, there seem to be few within the societies willing to complain, or even call attention to what would have been deemed, just a few years ago, a scandalous situation.

Hence, much of social science scholarship, rather than helping to solve these inherently difficult questions, participates in or contributes to the irresponsibility with which these issues are discussed in public discourse. A certain frivolousness in dealing even with simple facts is now so commonplace as to be nearly normative. Consider the straightforward question of differences in earnings between the sexes. One hears a great many assertions made on this matter, nearly all false or misleading. A moment's thought leads us to the understanding that wage differences are due to the fact that the two sexes usually do not do the same work or have the same history of continuity or seniority in the job, two factors being the major determinants of wages in the marketplace. Yet these facts have not prevented feminists from pointing to the "59 percent wage gap" as evidence of discrimination in the market, despite the fact that this statistic is based on gross comparisons between male and female incomes.

This essay concerns itself not with wage rates or labor economics, but with psychological sex differences, questions having to do with variations in ability, emotions, drives, personal traits, and the like. The reason I mention differential earnings is to provide a foretaste of the problems lying in wait. Money is one of the simplest variables we can imagine—it is tangible, quantifiable, morally neutral, and universally understood. The literature on sex differences typically deals with variables extraordinarily complex and elusive—such qualities as aggressiveness or dependency or moral outlook. Such qualities are intangible, difficult to quantify, morally controversial or ambigous, and the source of considerable confusion and disagreement. In this area, questions which initially seem to be en-

tirely straightforward soon turn out to be maddeningly tortuous, and teeming with nearly insoluble problems of measurement and interpretation. Consider whether men are more aggressive than women. One would certainly think so, on the basis of common observation, or such indices as the statistics on assault and disorderly conduct, or preferences in spectator sport. Yet when we approach the question analytically, matters do not seem at all straightforward. What do we mean by aggression? Do we mean physical violence or verbal abuse? Do we mean violence alone or such qualities as competitiveness or assertiveness? And how do you measure them? These questions or others like them have come to dominate the literature on sex differences and aggressiveness, which has further evolved into a group of sub-literatures exploring and arguing about fairly narrow matters of research design and the like.

That is the inevitable evolution in all areas of research, and should not distress us. Yet one senses that on this topic the evolution is being guided not so much by a more-or-less disinterested wish to clarify the issues as by the wish to substitute a new set of stereotypes for the ones we already have in place. What we now find in the literature influenced by feminist doctrine is that aggressiveness-as-violence is thought to be a masculine quality, whereas aggressiveness-by-assertiveness is thought not to be differentiated by sex. One can make a plausible case for that construction; but then one can make equally plausible cases for a number of entirely different constructions. The state of the literature, here as elsewhere, is so jumbled as to allow the findings to be arranged and interpreted almost at will.

The "gender and aggression" topic is more or less typical, neither the worst nor the best example of the problems involved in obtaining secure knowledge on difficult topics, and of the compounding of those problems when ideological passions are on the loose. It has seemed to me that any survey of the sex differences literature, prepared for a general audience, entirely misses the point, since an honest report of most topics within the domain would have to say something like: "This is the little that we know. The rest is speculation or pretense or wishful thinking. You do well to take it all with a grain of salt." It seems far more useful to look closely at a representative topic, to examine the state of the knowledge, to look at the controversies and why they exist, and in general to introduce the reader to a necessary skepticism.

Let us consider a fairly simple question—sex differences in mathematical ability—where the dependent variable is relatively easy to measure, where the differences between the sexes are well

established, and where the major desideratum—an improvement in female achievement—is universally accepted. It is also a topic where we have had, again relatively speaking, an abundance of research, enough so that the conflicts and disagreements that have emerged cannot be written off as a result of our ignorance, as is so often the case in the sex differences literature.

Julian Stanley and Camille Benbow, psychologists at Johns Hopkins University, are our leading scholars in the study of mathematical precocity. For some years they have devoted themselves to the question of early mathematical achievement—how to recognize it, how to cultivate it, and beyond that, the lessons to be learned about the origins and nurturance of mathematical and scientific talent in general. They have been working for nearly fifteen years to discover ways of identifying young talent and understanding its evolution: academic programs chosen, the effectiveness of acceleration and enrichment courses, the progress through high school, college, and early career.

From the beginning the emphasis has been on precocity itself, not on gender, but the fact of sex differences has become salient for two reasons: (1) it became evident that talented girls were far less likely to skip grades or take advanced classes, some indication of an absence of drive or encouragement or opportunity; (2) there were far fewer girls to be found at the very top in mathematical aptitude. Although on the whole boys and girls do not differ by a wide margin, there are astonishing differences at the highest levels of aptitude. In early adolescence, boys are represented ten or twelve times more often than girls at the highest levels of the test employed (and in the latest figures reported the ratio is 17:1). They feel that this is a datum of great social importance, since innovators in science and technology are drawn from the ranks of the highly talented, most of whom were precocious. For that reason, the modest advantage in boys generally may be of less importance socially than the extraordinary differences at the top. They have argued for a concentration on that particular gender gap, given the need to cultivate and recruit a scarce supply of the mathematically talented required in the generations ahead.

Benbow and Stanley were in a good position to make this point. They were not primarily students of gender, but of education; they had done more work on the question of mathematical talent than anyone else; and—no small matter—they were female and male, and hence were not easily accused of malice or invidiousness or sexism. Nevertheless, that accusation was made. Since some of the popular media in reporting their findings had raised the

possibility of a genetic sex difference, it was argued that if mathematical aptitude were to be widely understood as genetic, it would act as a disincentive to girls, to their families, and to schools. Benbow and Stanley replied that first, they had done more to stimulate mathematical learning, in both boys and girls, than anyone else; second, that they had not themselves raised the issue of a "math gene," third, their own view held that a mixture of exogenous and endogenous influences were involved.

So despite the bitterness between the adversaries, both agree—indeed, both aver—that there almost certainly is an interaction between nature and nurture. They also agree that the relative degree of influence cannot be measured with any precision, now or in the immediate future. They also agree that efforts at remediation could make a difference, and ought to be tried. In short, they agree that, speaking practically, one must be environmentalist, that is, if one is to improve the performance in girls one must concentrate upon "social reality." And it is at this point that the truly difficult question appears, since it soon becomes evident that we have no clear idea what comprises that reality.

There are two major ways in which "environment" might influence mathematical performance: through socialization, the myriad ways in which the family and other institutions form the total personality; and through situation—the pressures, constraints, opportunities, and incentives of the here and now. When we look to socialization as the key, we must choose among a vast array of possibilities. Is the clue to be found in cognition—are boys and girls perhaps being encouraged to think differently, boys being rewarded for logical as against expressive thought, or for playing number rather than word games? Might the differences have to do with the motivation each sex is permitted—girls, let us say, being forbidden competitiveness? The problem may lie in expectations—teachers, believing that girls cannot do well in mathematics, communicate these expectations to them, thereby inducing a self-fulfilling prophecy. Or it may be that girls and boys, looking at the occupational world they are about to enter, make commitments of effort and ambition based on an appraisal of the opportunities.

One gets a sense of the problem by looking at some of the variables just mentioned. Think how difficult it is to measure almost any aspect of the differential rearing of the sexes. The most economical way is some survey or interview of the parents, but that is fallible for obvious reasons—false memory, self-deception, the wish to say the right thing, and so on. Or one might spend a great deal of time, as anthropologists do, in the close and more-or-less constant

observation of a small number of families; but the problems there have to do with the expense and effort required, and the limited samples available. Even so, there would be problems having to do with an unbiased assessment of the observations. Or one might have occasional meetings with a larger number of families. Whichever method we choose, we will be getting only a partial picture, since for a full account one would have to study families at different social levels, of different sizes, with different structures, and at different ages. Even then, one could not easily make the case that socialization, or any particular aspect of it, is genuinely influential in the development of a given talent. One might do so, if the results were decisively clear; but to my certain knowledge, that has never yet been the case in socialization research on any topic.

Looking at the situation as a source of influence presents its own formidable measurement problems. Thus, there has been some attempt to directly observe the interactions among youngsters, or between youngsters and significant adults, sometimes preserving more-or-less "true life" situations, but more often setting up controlled experimental situations and observing behavior within these. Reviewing the results of these experiments, we soon become aware that there are often no strong correlations among situations, or between experimental behavior and real life behavior, or between experimental behavior and various measures of traits or abilities. Furthermore, even the most carefully crafted laboratory experiment, one which finds stable differences between males and females, may not find those at different ages, or given different conditions (such as the tasks given), or when conducted by different experimenters (there being, it now appears, a tendency for both men and women researchers to emerge with findings favorable to their own genders).

Does this account exaggerate the complexities? If anything, it understates them. For example, the best current model of academic choice—itself only one part of the larger question of talent and its training—provides for eleven *general* categories of variables, most of these subdivided, making more than twenty that would have to be defined, measured reliably, with the interrelationships plotted. Even so, it omits several variables which would seem to be necessary for an adequate picture.

Even when we achieve a plausible map of the variables we need to know about, we meet another problem far more serious than is generally recognized: the instability of findings from study to study. Seemingly straightforward relationships tend to lead to murky findings. It has been widely believed, for example, that mathematical

talent has a great deal to do with spatial skills, but the evidence on that rather narrow, focused question turns out to produce no strong findings. "Thus it appears that the relation between spatial skills and mathematical achievement is not yet fully understood." Needless to say, the uncertainties and confusions increase when we deal with more complex relationships. Even when we find what seems to be a clear set of correlations, it is not at all clear how we ought to construe the causal sequences. One example: most (though by no means all) studies show girls to be less confident of their math abilities, take fewer advanced courses in math, do more poorly, and have lower expectations directed toward them by parents and teachers. One plausible construction of these findings holds that the indifference to math achievement in girls (or the active discouragement of it) communicated by the culture through significant adults is the primary source of lowered achievement and loss of interest. Yet one could turn that on its head without doing any violence to the facts, arguing that girls on the threshold of adolescence, watching boys suddenly move ahead of them in math achievement, lose interest and put their energies elsewhere. I should say that I find the first construction somewhat more plausible than the second, but then again when we look closely at the findings in this area, we find that there are many plausibilities which turn out not to be true. One would certainly believe that there is an association between the amount of math done by parents and their children's atttitudes toward math, and plans to enroll in courses; but there is not. Throughout the literature on this topic, we find the belief that parents and teachers expect less from girls in math; in fact, the better studies are unable to confirm that nearly universal expectation.

Another confounding element has to do with historical changes. When we deal with such variables as values, sex roles, socialization patterns, economic incentives, careers, and so on, we are dealing with matters which are highly vulnerable to changes, both real and symbolic, in the culture at large. Almost all of the literature I have surveyed on parental expectations for their children's schooling is over a decade old, and it is hard to believe that attitudes have not changed in that period of time, especially given the continuing increase of women in the work force.

Let us pause here to review what we know about sex differences and math ability. It amounts to very little. Boys and girls do not differ much until early adolescence, and even then the gap between them is not at all substantial, although the number of genuine prodigies is vastly disproportionate between boys and girls. We do

not know why this is so, nor why pubescence is the apparent turning point. A biological explanation would seem to account parsimoniously for what is known, (it is my own preference, by the way) but so would an entirely environmental explanation.

Once we get past these plain facts we find ourselves awash in findings, which add up to very little when examined closely. Does a child like math because he is good at it, or vice versa? Do math teachers pay more attention to boys because they are boys, or because they are better in math, or because they are believed to be better in math, or believed to be better when paid attention to? Here, as elsewhere, the findings we have can be read variously. They do not compel any specific model of how mathematical talent is evoked, or enhanced, or directed. An existing model is imposed upon the evidence, guides the interpretation of what is found, and directs the search for relationships as well as the search for new findings. Hence, research tends toward the confirmation of existing belief, and although the controls of science are meant to minimize that tendency, they do so only over the long run, and never easily or perfectly. Given strong beliefs and frail evidence, there is all the more temptation to employ a coercive model to order the evidence and to formulate its meanings.

With respect to mathematical talent, the common belief has it that sex differences are a function of differential (and invidious) processes of socialization, initiated in the family, reinforced by later agents, such as the schools, the intent of which is to inhibit expectations, and aspiration, and ultimately performance in areas deemed to be "masculine" such as mathematics and science. If the socialization processes against math achievement in girls are so powerful, why do they not work in childhood, when presumably there is a greater malleability to adult pressure? Why are the data on parental pressures so weak and uneven? To return to the original Stanley-Benbow question, why are there such huge differences in talent at the top, and not elsewhere? If there is indeed a conspiracy to draw boys toward mathematics and girls away, what is the point of it? Presumably to keep women "powerless." If that is the case, why are they "permitted" to be better than men in verbal performance? It is the lawyers and memo-writers who rule the world—ask any engineer.

It is discouraging to reflect that after so much work, we end up knowing little more than we would from common sense alone. Here, for example, is one of the conclusions of the most thorough review of this literature we now have: "Thus, if a girl likes math but feels that the amount of effort it will take to do well is not

worthwhile because it decreases the time she will have available for more preferred activities . . . she will be less likely to continue taking math. Similarly, if a girl sex-types mathematics . . . as masculine and not in line with her own sex values, she will be less likely . . . to continue her mathematical studies, especially if she does not expect to do well."

Of course we know that already. Furthermore, there is nothing at all sex-specific about that conclusion, since it also might apply to boys. Boys who like math but feel that the effort to do well is not worth the time, if it cuts into, say, football practice, will tend not to take math courses. And boys who consider academic study to be unmanly, will be less likely to put any effort into school work, especially so if they do not expect to do well.

The authors go on to argue that what counts is not so much reality as the youngster's perception of reality—an arguable proposition—hence, adults ought to "become more sensitive to their own attitudes toward mathematics and avoid perpetuating stereotypic views of math achievement and [quantitative] careers. . . as inappropriate for girls and women." Yet if we look more closely at that very modest bit of advice, we see that it embodies an idea of human action itself quite arguable, to wit, that youngsters choosing a career are easily dissuaded from doing what they truly want to do, thus easily persuaded to do otherwise by enlightened adults. Why not assume instead that youngsters, both boys and girls, are on the whole rational consumers of careers, choosing through a calculus made up of opportunities, incentives, values, and talent? Why assume only benighted teachers and parents determined to grind down the young? We have, after all, seen during the past two decades some remarkable changes in the rise and fall, or fall and rise, of gender distribution in a number of occupations, especially such elite vocations as law and medicine. These changes took place because of other changes—economic, demographic, and legal—which in turn produced still other changes, in opportunities and incentives. Yet much of the research treats the labor market and other realities almost as epiphenomena, certainly as secondary, giving its full credence to the idea that society is no more than a vast, coercive, relentless, and evil machine for the perpetuation of sexism, so powerful that it must be countered by a vast and continuing propaganda campaign. That image of the American social order lies behind most of the research on gender and talent—inspiring the questions it deems important to ask, the answers it expects to find, and the interpretations it imposes on findings which, as we have seen, are invariably weak or equivocal.

Even so, the problems in literature on mathematical achievements are indeed minor when weighed against what we have in most other areas. What do we have where the variable is intrinsically complex or ambiguous, or difficult to define and assess? A good example is the current state of thought on the question of sex differences in morality. To begin with there are a large number of disputes about what "morality" really is—whether it is behavior, or sentiment, or quality of thought. Beyond that, there are vastly complicated questions of how to approach each of these elements conceptually and empirically. The specific question most recent research has concentrated upon is whether men or women have "higher" or "lower" levels of moral thinking. Depending on the instruments employed, one can demonstrate (a) that one sex or the other is higher or lower; (b) that there are no sex differences; (c) that there are differences, but only in quality or direction, not in degree; or (d) that there are qualitative differences which prove that either one sex or the other has a higher or lower level of moral maturity. The reader unwilling to believe this account of the state of the research is advised to study a recent issue of the scholarly journal *Social Research*, devoted entirely to the question of women and morality, containing a dozen or so contributions, all of them focusing on essentially the same body of information, yet differing so remarkably in approach and interpretation that the reader soon imagines he has come upon a Tower of Babel.

The serious reader, trying to keep up with what is going on in the social sciences, must rely upon the better newspapers, the weekly news magazines, or those publications devoted to reporting science for a general audience. So he will pick up the *New York Times* or *Newsweek*, or *Psychology Today*, or *Discover*, and therein learn about the breakthroughs, the recent findings, the new perspectives. The accounts given will likely be accurate, yet quite as likely misleading, in that they rarely capture the provisional, tentative, often ephemeral nature of the work reported. If you were to see the same studies discussed in a technical journal—let us say, the *Psychological Bulletin*—you would probably learn that for every finding in one direction, one can discover another in the opposite direction; or that earlier work has not been repeated, or is repeated only under very special conditions; or that an entire genre of research has proved to be false because of newly discovered methodological errors. And it is important to bear in mind that the "discrediting" of earlier work is by no means an occasional event; far more often than not, the secular trend is for prior work to prove insubstantial or incomplete.

In short, secure knowledge is extremely difficult to achieve in the social sciences. Minor variations in procedure can produce major variations in outcome. When findings accumulate in a domain, they are often such a mixture of yeas and nays and maybes that the scholar must order data through an interpretation others may find false or idiosyncratic. When findings are unclear or uninteresting, or when they conflict with current belief, the investigator will be too disheartened to write them up, or the journals will be unwilling to publish them. That is not conjecture: studies in several areas confirm that research which disconfirms the conventional wisdom of the field is less likely to find its way into print.

These are the ordinary hazards of doing and using social science. They can be overcome, but only in the long run, when there has been a considerable accumulation of work; we have in fact seen that take place in such areas as psychotherapy and education, but only when we have had hundreds of studies on a limited range of issues. That is not yet the case with respect to sex differences, where the quotidian difficulties of research are compounded by the strong ideological interests at work. On these topics, the prudent citizen ought not to believe what he reads, not fully, and those responsible for public policy should keep themselves fanatically skeptical when instructed on the latest lessons from social science.