

Social-Scientific Sexism: Gilligan's Mismeasure of Man*

BY DEBRA
NAILS

Appeals to reason or to the nature of the universe have been used throughout history to enshrine existing hierarchies as proper and inevitable. The hierarchies rarely endure for more than a few generations, but the arguments, refurbished for the next round of social institutions, cycle endlessly.—Stephen Jay Gould, *The Mismeasure of Man*

DIFFERENCES become deficiencies to those who peer through the bifocals of gender differentiation. Exactly as scientific racism flourished in the atmosphere of acclaim for IQ testing¹—and, before that, for craniometry—social-scientific sexism grows strong on claims that females, by nature or by nurture, have an intellectual or moral makeup or development different from that of males.

Carol Gilligan offers a description of female moral development wherein women prove to be more responsibility-oriented while men are more rights-oriented. This description—especially as it is elaborated in *In a Different Voice*,²

¹ Stephen Jay Gould's mighty attack on biological determinism, *The Mismeasure of Man* (New York and London: W. W. Norton & Company, 1981), discusses "the claim that worth can be assigned to individuals and groups by measuring intelligence as a single quantity" (p. 20). Three prongs of his attack are suggested by this phrase; and all three will prove applicable to the subject at hand once we substitute the words "moral maturity" for the word "intelligence" in what follows.

² Carol Gilligan, *In a Different Voice: Psychological Theory and Women's Development* (Cambridge, Mass. and London: Harvard University Press, 1982).

SOCIAL RESEARCH, Vol. 50, No. 3 (Autumn 1983)

which has captured popular attention—can and will be used as evidence for the inferiority of women unless (i) males are shown to undergo the same development, or (ii) the female morality is proven superior, or (iii) the entire enterprise is undermined as fallacious and unreliable. In a well-known dialogue between Lawrence Kohlberg and Owen Flanagan, Kohlberg addresses the first point: “The usefulness of Gilligan’s ideas in casting light on puzzles in our own data was a major opening for us, and we adapted her concept of ‘responsibility’ to mediate between both male and female judgment and action. . . .”³ Flanagan, taking up the second point, both criticizes Kohlberg’s claim that each successive stage of moral development is more “adequate” than all stages preceding it and points out that Gilligan must prove greater adequacy for the different moral development she claims exists for women if she is to escape the charge of female inferiority.⁴ This paper assesses the third point, looking critically at both Gilligan’s method (presentation of data and interview procedure) and her presuppositions: reification of moral maturity, scaling from negative to positive, and measurement of moral (rather than cognitive) skills; concluding with a few remarks about the danger of generalizing for groups and applying one’s generalizations to the individuals within the group.

It will be clear to those who are familiar with *The Mismeasure of Man* that this outline follows closely Gould’s lead of questioning research which itself leads to the oppression of disadvantaged groups. As Gould notes, he has reanalyzed data and found “a priori prejudice, leading scientists to invalid conclusions from adequate data, or distorting the gathering of data itself.”⁵ This is the point of the sections labeled “Presentation

³ Lawrence Kohlberg, “A Reply to Owen Flanagan and Some Comments on the Puka-Goodpaster Exchange,” *Ethics* 92 (April 1982): 514.

⁴ Owen Flanagan, “Virtue, Sex, and Gender: Some Philosophical Reflections on the Moral Psychology Debate,” *Ethics* 92 (April 1982): 502–507. See also Flanagan’s article with Adler in the present volume.

⁵ Gould, *The Mismeasure of Man*, p. 27. The author is grateful to Mary Ann O’Loughlin for suggestions about this section.

of Data” and “The Interviews.” Two of the three “Presuppositions” sections and the concluding comments are very close to Gould’s findings in his analyses of studies measuring intelligence. Yet whereas Gould’s studies are primarily historical, the present undertaking concerns a vital contemporary research project and a social scientist who seeks to liberate rather than further to oppress a disadvantaged group.

Method

A perfectly appropriate response to the suspicion voiced above, that the purporting of difference for female moral development leads to the claim of inferiority for that development, is: The proper battle against the claim of deficiency does nothing to undermine the fact of existing *differences*; Gilligan’s descriptions are presented as evidence of differences, and it would be irresponsible to ignore these empirical findings. Thus one begins empirically, examining the Gilligan descriptions.

Presentation of the Data. Given the admitted difficulties of managing large bodies of data, of featuring portions of case studies without misleading consequences, one still expects reliable accounts from social scientists. Without confidence that the data one reads are not distorted, that there is accurate sampling in statistically based sciences, one cannot hope to make informed judgments about the conclusions, much less the theories, drawn from the data and their interpretation. If, on the other hand, a social scientist holds that it is his or her proper role to highlight certain data and exclude others in the belief that this will illustrate more accurately some underlying truth heretofore unrecognized by others, deliberately to perform a hermeneutic “interpretation” in the process of presentation, the social scientist must make this known so the reader can approach the material forewarned. When one method poses as the other, the reader is treated unfairly.

A painter or a novelist, describing someone's transition, may portray the person as starkly at each end of that transition as is necessary to make some point; but what one expects of the social scientist is a description which reflects the person as accurately as possible, even if that means the reader may have great difficulty to recognize that a transition has occurred at all. One expects not examples which argue a particular view most successfully but representative examples.⁶

Carol Gilligan's description of the transition of a woman she calls Betty⁷ seemed a little too neat—more like a caricature than a characterization; so corroboration was sought in the dissertation of Mary F. Belenky,⁸ who performed an abortion-decision study with Gilligan and whose project allows her to present somewhat more data from the cases than does Gilligan.⁹

From the Belenky material, one learns that Betty was assigned to the Gain Group after her initial interview; this

⁶ James C. Walker takes issue with the author's distinction between interpretation and empirical science, moving the argument a step further: "I think that the work of painters and novelists on the one hand and of social scientists on the other are both theory-laden. What counts as 'accuracy' will depend on your theory. The question for the social scientist, and a question to which we have the right to a straight and frank answer, is: Do we all agree on the theory which functions to select data, methodologies, etc. The sin here is unacknowledged or innocent tendentiousness. Similarly for 'representative examples.' What is representative of a population will depend on our theory of that population—otherwise we are left with little alternative but random samples. We have little problem if our various theories, conflict though they may on some points, agree on those points germane to settling what counts as a 'representative example.' If they don't agree on this—i.e. if our theories entail contrary deductions on sampling, we must search for some other methodology if we wish to settle the issue" (personal communication).

⁷ Gilligan, *In a Different Voice*, pp. 75–76, 109–115.

⁸ Mary F. Belenky, "Conflict and Development: A Longitudinal Study of the Impact of Abortion Decisions on Moral Judgments of Adolescent and Adult Women," Ph.D. dissertation, Harvard University, 1978.

⁹ Between Belenky's writing of her dissertation and Gilligan's publishing of *In a Different Voice*, the two women collaborated on "A Naturalistic Study of Abortion Decisions," in R. Selman and R. Yando, eds., *Clinical-Developmental Psychology* (San Francisco: Jossey-Bass, 1980), pp. 69–90, wherein Betty warrants a lengthy discussion. The article is interesting as a sort of transition piece between the two longer works, but it is useless, because of its joint authorship, for illuminating the question why one researcher chooses to feature data that another chooses to discard.

means that she was expected "to bring a higher stage of reasoning to bear in the hypothetical moral dilemmas in the one year follow-up."¹⁰ This expectation was based on a comparison of two scores: her reasoning about her existing pregnancy (scored with reference to a manual developed by Gilligan, Lerner, and Belenky¹¹) and her reasoning about hypothetical moral dilemmas developed by Kohlberg and others (scored with reference to Kohlberg's revised manual¹²). Betty's reasoning about her own situation was "staged" higher than her hypothetical reasoning; hence, the positive expectations. A year later, as expected, Betty's Moral Maturity Score (MMS) had increased by 50 points, and she was said to have moved from predominantly Stage 2 to predominantly Stage 3.¹³

In Gilligan's abbreviated description, Betty's story "illuminates the potential for change in a seemingly sparse life," illustrating the transition from " 'selfishness' to responsibility." The "before" picture of Betty depicts her as "an adopted adolescent who had a history of repeated abortions, disorderly conduct, and reform school"; "Betty was sixteen when she went to an abortion clinic for a second abortion within a period of six months."¹⁴ In Belenky's narrative, Betty is said to have had one prior pregnancy—the result of a rape in which she had been beaten. She had afterward refused birth control because ". . . I had a real attitude about guys, like I said to

¹⁰ Belenky, "Conflict and Development," p. x. The twenty-four women who completed the first interview were divided into three groups. In addition to the Gain Group, there were the Stable and Loss groups, characterized by the expectations of no developmental change and possible regression, respectively. For a more detailed description of the group divisions and hypotheses about each, see *ibid.*, pp. 37-43.

¹¹ Carol Gilligan et al., "Moral Reasoning About Sexual Dilemmas: The Development of an Interview and Scoring System," in *Technical Report of the U.S. Commission on Obscenity and Pornography*, vol. 1 (Washington, D.C.: U.S. Government Printing Office, 1971).

¹² Lawrence Kohlberg et al., *Assessing Moral Stages: A Manual* (Cambridge, Mass.: Center for Moral Education, Harvard University, 1978).

¹³ Belenky, "Conflict and Development," p. 113.

¹⁴ Gilligan, *In a Different Voice*, p. 109.

myself that I was never going to be involved with a guy again. After I was raped, I thought all men were pigs. So I just said to the lady, she asked me if I wanted birth control, I told her no because I didn't intend to have sex with guys ever. But I did."¹⁵ Gilligan mentions the rape in passing two pages into her narrative.

Gilligan's negative characterization of Betty contains a description of her relationship with her boyfriend (bracketed ellipses are mine):

Given her expectation that if she went to bed with him, he would continue to meet her needs, her disappointment was great when she discovered that, "after I went to bed with him, he just [. . .] wanted me to do everything that he wanted to do. [. . .] I was more like a wife than a girlfriend, and I didn't like that." Describing the relationship as one of exchange, she concludes that he "was really one way," seeking only to meet his needs and disregarding "the fact that I wanted more freedom."¹⁶

On the basis of this description, Gilligan remarks on Betty's "concern about maintaining her freedom rather than having to do things for others."

Belenky's description is more complete, particularly in that it gives us a condition for Betty's dissatisfaction (interviewer's questions in capital letters, unbracketed ellipses are original author's):

At first he treated me really nice, but then he started to change. . . . (BECAUSE?) Because after I went to bed with him, he just treated me like shit. He was too possessive with me, he just wanted me to do everything that he wanted to do. Telling me to clean the house, to do this and telling me not to go here or there. I was more like a wife than a girlfriend. I didn't like it. . . . I talked with him, but he was ignorant. He didn't quite understand. He was really a one-way person. (HE DIDN'T UNDERSTAND WHAT?) The fact that I wanted more freedom.¹⁷

¹⁵ Belenky, "Conflict and Development," p. 102.

¹⁶ Gilligan, *In a Different Voice*, p. 110.

¹⁷ Belenky, "Conflict and Development," pp. 104-105.

Four things in the Belenky account make one uncomfortable with the Gilligan interpretation: (a) that ellipses are not used to indicate omissions of text;¹⁸ (b) that the apparent concrete reason for Betty's desire for freedom (being told to clean house etc.) is omitted; (c) that Betty's description of her boyfriend as "ignorant" and not quite understanding is reduced to Gilligan's word "disregarding"; and (d) that the whole episode is summarized to make Betty appear as selfish as possible ("freedom rather than having to do things for others").

However, when it comes time to describe the "dramatic change" in the transformed Betty, Gilligan's descriptions work in the opposite way, highlighting the positive. Compare the two descriptions of change in Betty's life. First, Gilligan:

I said I have to lose, and it was such a change for me, because I had been fat for so many years. And being thin, I never knew what it was like to be able to wear clothes that looked good. I just felt dynamite, because so many people and so many guys were trying to go out with me. It was the first summer I was able to wear a bathing suit.¹⁹

Then Belenky:

In June, I said I have to lose weight. It was such a change for me, being thin. I never knew what it was like to be able to wear clothes that were comfortable and stuff—that looked good. Guys were trying to go out with me. I said, "Wow, this is heavy," you know. It was dynamite. It was the first summer I was able to wear a bathing suit, too. . . . It took me all that time to get my shit together.²⁰

The latter quotation is less reserved, less grammatical, perhaps, but more like the former Betty. The reader is left

¹⁸ Solely from Gilligan's book and Belenky's dissertation, the possibility could not have been discounted that apparent omissions in the Gilligan passages were actually additions in the Belenky dissertation. Carol Gilligan kindly granted me access to (unpublished) transcripts of the interviews themselves. This was extremely useful not only in verifying portions of the text but in illuminating the issue of how certain data come to be featured in differing accounts.

¹⁹ Gilligan, *In a Different Voice*, p. 113.

²⁰ Belenky, "Conflict and Development," p. 112.

with the impression that Betty is doing better, but it is significant that neither researcher reports the *preceding* unhappy passage that belies the good report:

Last summer I lost a lot of weight and I felt really good about that. I was so happy. Like I used to be really, really fat, from fifth grade to eighth grade and I lost weight, but now I am putting it on again. I lost some weight last summer and I felt so good about myself. And this winter I just gained it all back, plus ten pounds. Now I am feeling pretty low about myself because I have gained so much weight.²¹

A page later, Gilligan has Betty saying "I am very sensitive" without the comment immediately following—reported by Belenky²²—"My self ideal would be not to be so sensitive. Like people can hurt my feelings so easily." For Gilligan, women typically are concerned not to hurt others instead of to avoid being hurt themselves.²³ But insofar as these are descriptive passages, their import is minor compared to another—and final—set of passages which is used as a basis for an interpretation by Gilligan.

At various times in *In a Different Voice* we are called on to consider responses to the Kohlbergian dilemma of Heinz. Heinz's wife is dying of cancer and a pharmacist has a drug that can cure her. However, he is charging more for the drug than Heinz can pay, so Heinz must consider whether to steal the drug. Betty's response to the dilemma at the time of her abortion as given by Belenky is:

Yah, the guy is ripping people off and his wife is dying so he deserves to be ripped off. (IS THAT THE RIGHT THING TO DO?) Probably, I think that survival is one of the first things in life that people fight for. Stealing might be wrong, but if you have to steal to survive, yourself, or even kill, that is what you should do. . . . The preservation of one's self is the most important thing—it comes before anything in life.²⁴

²¹ Transcript of the second interview, pp. 12–13.

²² Belenky, "Conflict and Development," p. 115.

²³ Gilligan, *In a Different Voice*, p. 65.

²⁴ Belenky, "Conflict and Development," p. 103.

Gilligan's quotations of the same material²⁵ contain only minor discrepancies compared to the passage above, easily attributable to two typists' transcriptions of the same voice. What Gilligan chooses to discuss from the passage is the emphasis on survival; Betty—as an adopted adolescent—“experiences herself as uncared for and alone.” But the neglected comment that the druggist deserves to be ripped off because he is ripping off others is not negligible here. Desert (and its negative aspect, revenge) is a concept within the sphere of moral discourse, albeit a concept more neatly fitted into the language of rights than into that of responsibility and relationship. This seemingly minor instance of neglect becomes relevant later when Gilligan describes the changed Betty's response to Heinz's dilemma:

The change in Betty's moral understanding is evident in her response to Heinz's dilemma. She now says that Heinz should steal the drug “because his wife is dying, near death, and he loves his wife.” Although she explains that she is going to “answer the same as before,” referring to the choice itself, the structure of her justification has fundamentally changed. Whereas previously she indicated the primacy of survival, now she emphasizes the importance of relationship. Where she spoke of entitlement, now she speaks of guilt. Heinz should steal “because he loves his wife, and if she dies, he is going to feel like he could have done something but he didn't.” Thus security, which she formerly saw as self-protection in an exploitative world where everyone gets ripped off, now depends on relationships with others, on the expression of love and care.²⁶

But Belenky's account quotes a longer and more informative segment of Betty's second response to the dilemma:

Because his wife is dying, is near death and he loves his wife. I think the druggist is really a jerk to sell it for that price. . . . He is heartless about it and if I was Heinz, I would steal the drug to[o]—just for being a heartless person. . . . He loves his wife,

²⁵ Gilligan, *In a Different Voice*, pp. 76, 110.

²⁶ *Ibid.*, pp. 113–114.

and if she dies he is going to feel like he could have done something but he didn't. If she dies he might feel really guilty, and wish that he had stolen it, had taken any means he could to have to get it.²⁷

The element of revenge returns here: Heinz should steal the drug just because the druggist is a heartless jerk. Whatever the lovely thoughts in Betty's second response, it is precisely those words that link this response to the former one, justifying her saying that she would "answer the same as before." Furthermore, the transcript of the second interview shows that the material that might have been included in the second ellipse of the paragraph is: "I think in a case of life and death, any means, he should take"—hardly a backing away from emphasis on the primacy of survival. It may be clear that Betty's second response is more measured than her first, less bitter, warmer. What is not clear is that the abortion decision itself was a significant and moral cause of the change.²⁸ It becomes no clearer when Betty says, "In a lot of ways this pregnancy has helped me . . .," because what she goes on to say is that the pregnancy caused her to stop getting high and drinking and "after a couple of weeks, I thought about it again, and I said, 'No, I can't have it, because I have to go back to school,'"²⁹ which seems as much a cognitive as a "moral" determination, at least in that one is in a better state for making decisions if one is not drugged (but more about the cognitive/moral distinction later). What is patently unclear is that the transition

²⁷ Belenky, "Conflict and Development," p. 113.

²⁸ The transcript of the second interview has Betty, at another point, giving credit to her new boyfriend for the change. Her special school also plays a prominent role in her description of the positive side of her life. But there is negative material as well, not just her having gained weight again, but in her complaint about her current boyfriend ("He thinks I should play my role and I don't like to play my role at all. I just want to go and do what I want to do," p. 20) and her flip-flops about various (factual) matters from one part of the interview to the next.

²⁹ Gilligan, *In a Different Voice*, p. 112. On p. 7 of the transcript of the first interview, Betty says she stopped drinking and getting high because doing so made her ill and undermined her attempts to lose weight. She later mentions her concern for the baby as a cause (p. 15).

between the old and the new Betty is as stark as it is painted by Gilligan. There is at least as much interpretation as empirical science going on in *In a Different Voice*. However prettily the literary criticism disguises itself as science, one cannot trust its conclusions.

The Interviews. Even if every case study were presented verbatim—the author's comments mere marginalia—the interview procedure itself, described in some detail in Belenky's dissertation, would remain a cause of concern. Granted, the developmental psychologist has a special set of problems which do not lend themselves easily to "standard" testing procedures. Still, without a control group, without a consistent test, without a rigorous scoring procedure, it is appropriate to question what procedures are employed. In the case of Belenky's and Gilligan's abortion study, Belenky offers the following description:

A semi-structured, open-ended interview format was used, beginning with general questions designed to discover the categories of the woman's own thinking and then progressing to questions which sought to ascertain how her judgement of the pregnancy dilemma was related to her concept of herself and her understanding of moral issues. Interview responses were probed until it was possible to ascertain the stage of reasoning used to justify each alternative choice the woman was considering as well as the way she herself judged these various alternative resolutions.³⁰

Why would one assume *a priori* that the interviewer will probe deeply enough to reveal the various types of moral reasoning that a person may be using at the same time; what would keep the interviewer from exploring one type, designating a stage to that type, and going on satisfied to other points without uncovering important evidence of other types of reasoning? Suppose, for example, Betty's interviewer had thought it important to explore with her why and in what

³⁰ Belenky, "Conflict and Development," pp. 48–49.

circumstances revenge (or retribution) is an appropriate response to someone's actions. Could Betty have been operating just short of the Categorical Imperative, from some maxim of exchange? Probably not. But it is just possible that the interviewer was deaf to a revengeful voice, finding comments about survival (or, later, love and guilt) more interesting to pursue. Needless to say, in an interview involving two people, it is not only the official follow-up question that determines the direction of the narrative³¹ but the more subtle communications of posture, eye contact, and facial expression, conceivably amplified by the inequality of a situation in which an insecure adolescent faces a mature academic researcher.³²

Belenky lists questions posed,³³ and both Gilligan and Belenky offer numerous examples of follow-up questions, but none of this is sufficient to overcome one's suspicion that the data collected are colored substantially by the interviewers' expectations and knowledge. And the interviewers are aware of some possible effects of interaction with their subjects. Belenky comments, "We conducted a number of pilot interviews with a wide range of women . . . until we were confident in our ability to be supportive to a person in the process of coping with this difficult issue."³⁴ If one wonders what "supportive" can mean in the context of such an interview, an example from the transcript of the first interview is: "I think you have a right to take care of yourself and do the things that will help you in your life." But also: "If the pregnancy went on it [the baby] would be formed. It has the potential for growing into a person, that doesn't mean you shouldn't have the abortion, that is the reality of the situation."

³¹ The interviewer, for example, might have talked with Betty about why she was presently *gaining* weight instead of pursuing the story of her having lost weight the preceding summer.

³² Near the end of the second interview, Betty says, "I was glad when you called back because I knew it would be good to tell you that things worked out so well because the last time I saw you I was a mess" (p. 22 in transcript).

³³ Belenky, "Conflict and Development," Appendix B.

³⁴ *Ibid.*, p. 49.

Presuppositions

Even if some way could be derived to remove the more obvious difficulties of the interview procedure, there would remain underlying assumptions so shaky that no amount of intellectual scaffolding could support them. They stay in place through the enormous power of cultural reinforcement; they *feel* right. This section sketches a critique of three related presuppositions which—although they have not yet disappeared from Gilligan's own work—are far more prevalent in the work of people around her. It is to Gilligan's great credit that she has placed some distance between herself and the tradition in developmental psychology from which she has emerged. If she continues to widen the gap, so much the better for developmental psychology.

Reification of Moral Maturity. Gould has argued eloquently that intelligence is not a single "thing" to be measured with IQ tests, yet the belief that it *is* continues to bolster the enormous amount of IQ testing (and accompanying misinterpretation) that goes on; the persistent and comfortable picture of intelligence as locatable within the brain and quantifiable along a scale continues to inform educational systems as well as individual expectations. The assumption that there exists a link between intelligence and moral maturity continues as uninterruptedly. Consider Gilligan's comment on the selection of a sample of sixth graders: "The sample selected for this study was chosen to focus the variables of gender and age while maximizing developmental potential by holding constant, at a high level, the factors of intelligence, education, and social class that have been associated with moral development, at least as measured by existing scales."³⁵ Also, Kohlberg's description of some of his longitudinal data: "A factor analysis of correlation in stage used on various dilemmas and issues indi-

³⁵ Gilligan, *In a Different Voice*, p. 25.

cated a single general 'moral development factor' analogous to the general 'intellectual functioning factor' found in tests of cognitive ability and reasoning."³⁶

The reification of moral maturity is as unjustifiable as the reification of intelligence. From the Latin *res* for "thing," to reify is to posit the physical existence of something. Unicorns exist in tapestries and literature, but one has reified them if one expects to find them gamboling in the forest. Ideal triangles exist in geometry proofs but they do not lie about on pieces of paper, nor do surveyors' instruments provide them for property owners. Factors, such as those mentioned by Kohlberg, are mathematical abstractions employed in factor analysis—a deductive mathematical technique—but whether the factor in the analysis represents something that actually exists and is causal is yet at issue.

A fundamental part of resolving the issue is to examine the search for very strong correlations among data. Kohlberg correlates stages: a person's responding at exactly the same stage to all the various hypothetical dilemmas presented constitutes perfect correlation. (Never mind the problem of sorting through the responses for material of the properly "moral" character before assigning stages to the responses.) Now, some high correlations occur because one event directly causes another: the repeated flicking on of a light switch *a* causes the light bulb *b* to glow repeatedly (assuming the circuit and Boston Edison are functioning properly). Some occur because there is an underlying cause of both the effects being correlated: if two bulbs *c* and *d* are on the same (properly functioning) circuit, then the repeated flicking of a switch will cause both to glow repeatedly. And some occur although there is no connection between the events: the Dow Jones industrial average *e* has been gaining in the past several days and the Red Sox *f* have been winning. The correlation of *a* to *b*, of *c* to *d*, and of *e* to *f* may be the same. The point is that the claim that one event causes another or that there is one underlying

³⁶ Kohlberg, "A Reply to Owen Flanagan," p. 517.

cause of two events must be based on something other than correlation alone; evidence must come from outside the immediate situation, much as one adds the qualifier that a circuit be functioning properly. If the switch one flicks is not wired to the bulb that glows, then one expects no correlation between flicking and glowing. The Kohlbergian model presupposes that, if various cases of reasoning about moral issues are all glowing with the same intensity, then the cause must be that they are all wired together in one circuit with one switch: call it the "moral development factor," having employed the useful technique of factor analysis to explain and simplify the matrix of correlation coefficients.

No independent evidence is offered that the mathematical abstraction that helps one to understand points on a graph exists physically as a single quantifiable entity in human beings. As different bulbs may have different switches or separate circuits, may be on automatic timers or attached to battery packs, a person's moral development may involve a whole constellation of skills, beliefs, capacities, memories, emotions, lacunae and much more, each having its own origin, and each involving its own cluster of qualities and quantities. There is nothing *a priori* to mandate one description over the other or over any one of several other descriptions that could be given. Mere correlation of stages provides no information about cause, no evidence that a measurable moral maturity factor actually exists underlying a group of responses.

Scaling from Negative to Positive. Moral Maturity Scores (mentioned above in connection with Belenky's and Gilligan's expectations for Betty) differ from IQ scores in that the former typically increase over time while the latter are supposed to remain relatively stable. This difference has the effect of making moral maturity a little less likely to be associated with heredity—a factual plus in this bigoted world. But left intact and reinforced is the assumption that the MMS measures a unilinear quantity, that a single number can be attached to a

person's moral maturity and thus to his or her moral worth. Let there be no misunderstanding: the measurement of moral maturity—regardless of the intentions of those who have developed the field to its present state—does imply the measurement of worth. Persons at the negative end, persons whose moral reasoning is scored in the lower stages, are presented as presently morally deficient, even if having potential for improvement. Part of Gilligan's original disagreement with Kohlberg was a reaction to the fact that, by his scoring procedure, women tended to become stuck in the lower stages;³⁷ but Gilligan's description of female moral development offers nothing more than a fork in the road to the same destination, a "branching" of the once-unilinear sequence. Gilligan still speaks of sequence, of charting progression, of direction of change. It is implicit in both authors that the most moral (virtuous, ethical) people are those who have progressed farthest in the right direction in the established sequence.

Doubtless it is common to view some people as more moral than others, but the determination of *which* people, types of reasoning, and personal characteristics are *more* moral involves much more widely varied information than indicated by responses to a Kohlbergian moral dilemma, even as supplemented by Gilligan. Who establishes the sequence, determining that Kantian types should occupy the uppermost reaches of morality? Kohlberg? Gilligan? Particular reservations aside, the ranking involved in any stage model of moral development requires criteria. But these criteria have, for centuries, been controversial, often socially determined and nebulous. It *would* be convenient to have a number for the purpose: the MMS as the single measure of rank in the moral order.

However convenient, however comfortably fitted to anyone's craving for scientific exactitude, however warmly one embraces the notion of progress, one still must question

³⁷ Gilligan, *In a Different Voice*, p. 18.

whether a single (or bifurcated) scale from negative to positive most accurately reflects differences in human moral development. Gilligan, indeed, questions just this, refusing at one point to scale differences from better to worse,³⁸ and speaking of the "bias of developmental theory toward ordering differences in a hierarchical mode."³⁹ But she later retreats, as noted above, into the language of hierarchy.

Scaling reflects comparison; they say you can't compare apples and oranges. Imagine getting agreement on a Goodness Score applicable to all fruits. Apart from the old controversy, "*de gustibus . . .*," pieces of each variety of each fruit would have to be compared with their fellows for nutrient content and shape and color and texture and size and shelf life and much more. It would be relatively easy to run a battery of tests on various varieties of various fruits; and it would be simple to subject the results to factor analysis, to stipulate a scale, and to assign a GS to each variety. Grading plays a role in determinations of some of the qualities mentioned above, but neither growers nor vendors nor consumers are tempted to perform such a reduction to a single scale from negative to positive. And how much more complicated are humans. It may well be that the many multifariously intricate and elusive qualities we properly associate with "good individuals" do not lend themselves to scaling, that variation should be preserved rather than eliminated in descriptions of the moral development of persons. Conservation of variation may preclude not only one's justifiably being able to rank order the members of a group of individuals by which qualities they possess and in what quantity (that is, to assign a single number, a MMS), but may preclude as well the idea of progression in an individual's moral development through a sequence of stages. A model emphasizing qualitative rather than quantitative difference (a cluster or constellation model, for

³⁸ *Ibid.*, p. 25.

³⁹ *Ibid.*, p. 33.

example) may well be much better suited to descriptions of moral development than a hierarchical model.

Cognitive versus Moral Determinations. Gilligan assumes that she is measuring moral development, not some other skill. It has been extensively argued by others that a distortion in the present conception of moral maturity is that it is measured by examining a person's *words*, without attention to his or her moral *deeds*. If this is so, then Gilligan's approach—with its intended interweaving of choice in *actual* moral dilemmas—might have been a methodological improvement. A possible explanation for the fact that it is not is that Gilligan—like Kohlberg—measures *cognitive* skills. One would suppose that persons who measure moral development would defend the proposition that moral maturity is something distinct from cognitive skill, although there might be overlap, dependence, or some other relationship between the two; it is important to *articulate* that relationship. With its deep conceptual and philosophical complexity, this tangled problem cannot be solved standing on one foot, but one suspects that it is, in part, the confusion of cognitive with moral statements that causes much misunderstanding of the data. The literature in developmental psychology is replete with examples of research on the relationship between moral and cognitive judgments and moral judgments and deeds. Lest one be accused of doing an injustice to a complex subject, may it suffice to join the dialogue between moral psychologists and moral philosophers with a digression into how the philosophers sometimes proceed empirically:

Bernard Rosen has developed an exercise⁴⁰ which invites students to agree strongly, to agree, to disagree, to disagree strongly, or to say “can't answer” to a wide variety of state-

⁴⁰ The exercise and key (which I have used as a no-credit ungraded exercise given at the beginning and at the end of introductory ethics courses) are adapted from unpublished materials provided by Bernard Rosen (Ohio State University) in 1976.

ments (fifty-one in all). For example, a few of the interspersed statements about motivation: "Hitler's motives, let us suppose, were to improve European civilization, eliminate crime, reduce unemployment, and to restore a sense of pride in Germans. In spite of these good motives his actions with respect to the Jews and Slavic countries were wrong." "The only motive anyone has in doing anything is to get something for himself. Even when you help others it's only because it makes you feel good." "If someone sees a person drowning and is motivated to try to save the person, then whether the effort is successful or not, the action is just as praiseworthy." The "key" to the exercise provides certain labels: "denial of formalistic deontological theory of obligation," "psychological egoism," and "denial of teleological theory of obligation," for example; and it provides a cross-reference of statements, agreement with which would be inconsistent with, or contradictory to, agreement with other statements.

It will come as no surprise to anyone who has taught introductory ethics that students quite typically have a high incidence of contradictory and inconsistent responses when the exercise is first given; and no surprise either that there are high percentages of "relativists" and "egoists" among freshmen and sophomores. After a semester of wrangling with a succession of worthy opponents—Plato, Aristotle, Hume, Kant, Mill, Nietzsche, and Moore—these same students not only are able to eliminate their contradictions, explain their inconsistencies, and answer the exercise on behalf of each of the philosophers they have studied, but call themselves "act utilitarians," "deontologists," and other such names—and often, they say, without having changed a whit how they would behave in any *actual* moral dilemma from one end of the term to the other.

Reading Kohlberg's and Gilligan's hypothetical dilemmas and questions for interviewees, one sees that the student with facility in the language of moral discourse would be adept at addressing their issues. No wonder philosophers invariably

score in Kohlberg's highest stages,⁴¹ whatever the quality of their actions. One can accept the notion that there is more to moral maturity than cognitive skill and still deny that the moral psychologists are measuring anything other than cognitive skill.

Concluding Remarks

Gilligan offers an early disclaimer:

The different voice I describe is characterized not by gender but theme. Its association with women is an empirical observation, and it is primarily through women's voices that I trace its development. But this association is not absolute, and the contrasts between male and female voices are presented here to highlight a distinction between two modes of thought and to focus a problem of interpretation rather than to represent a generalization about either sex.⁴²

But her book is characterized by generalizations about the sexes, offered as descriptions of differences. Examples: "... women replace the bias of men toward separation with a representation of the interdependence of self and other, both in love and in work." "In view of the evidence that women perceive and construe social reality differently from men and that these differences center around experiences of attachment and separation, life transitions that invariably engage these experiences can be expected to involve women in a distinctive way."⁴³

A danger of the Gilligan-type description of female moral development is that it has the power to exaggerate existing differences, or even to create expectations that reward par-

⁴¹ Flanagan, "Virtue, Sex, and Gender," p. 509.

⁴² Gilligan, *In a Different Voice*, p. 2.

⁴³ *Ibid.*, pp. 170, 171.

ticular behaviors. It does not much matter whether one who believes in the accuracy of the Gilligan descriptions believes as well that the differences occur through hereditary or environmental causes. The description itself can erect a set of boundaries for female moral development, a set of limits on behavior: a girl child who sees a moral dilemma as “sort of like a math problem with humans” (a response of one of Gilligan’s male sixth graders) is viewed as somehow less feminine than one who emphasizes the relationships among the various characters of a hypothetical dilemma. And it does not much matter whether the researcher herself or himself has the good intention of giving voice to a disadvantaged group. The story of the original purposes for which Alfred Binet intended his scale (later IQ) makes sobering reading.⁴⁴

For the sake of argument, let us assume that Gilligan correctly identifies a fact of nature or environment, that there are two different modes of thought in moral reasoning—call them rights-oriented and responsibilities-oriented—empirically observable to be strongly predominant in males and females respectively (typically represented by overlapping bell curves). This “fact” would be useless in predicting the orientation of any particular individual. If one considers characteristics *pre-dominant* in one gender (as opposed to definitive characteristics such as “having testes” or “having ovaries”—which are virtually mutually exclusive), one cannot predict what will be the case for *any* particular individual. As Gould puts it, “variation among individuals within a group and differences in mean values between groups are entirely separate phenomena. One item provides no license for speculation about the other.”⁴⁵ Aristotle (the fellow in whose list of opposites women fared so poorly) thought that reasoning from facts could lead to actions, that the conclusion of a practical syllogism is an action.

⁴⁴ Gould, *The Mismeasure of Man*, p. 146–158.

⁴⁵ *Ibid.*, p. 156.

But what actions could be proposed on the basis of Gilligan's fact by those who believe in it? Should one legislate or educate with her gender distinctions in mind, or especially encourage boys to acknowledge responsibilities and girls rights? Or should the debate remain within the hallowed halls?

Let the red flags go up in warning: this type of research is social science at sea without anchor, and no one is out of danger. Ideology—even ideology in the service of the oppressed—is a poor underpinning for research. Let us beware most of all, and criticize most effectively, those with whom we yearn to agree. Worth must not be assigned to individuals and groups by measuring moral maturity as a single quantity.

*An earlier version of this paper was presented at the Cuban Academy of Science, Havana, in February 1983. I should like to thank the following for their criticism: Bob Cohen, Carolyn Fawcett, Deborah Johnson, Berend Kolk, Mary Ann O'Loughlin, Katie Platt, and Jim Walker.