of the plaintiffs' expert?" He said to me: "You don't understand. If the plaintiffs' expert hadn't been busy running multiple regressions she might have taken a closer look at the employee manual which describes what in essence is a two-tiered job system. Men are channelled into one tier and women into the other. After that, virtually all employment decisions follow as a matter of course. When our expert responded by running his own regressions, the lawyers were quite pleased. They believed that the outcome would have been far worse if he had explained to the court what we really do because then the judge could easily have concluded that our system was discriminatory on its face."

Within Dempster's framework, I had special difficulty in understanding the distinction he attempts to draw between judgmental discrimination and prejudicial discrimination. For me, attributing judgmental discrimination to "a presumed honest attempt to assess productivity" is ignoring the realities of the legal meaning of discrimination and the judicial injunction that statisticians cannot use intrinsically tainted carriers of discrimination as predictors in their statistical models. It is all well and good for Dempster to say that his definition of fairness implies that "there is no restriction at all on the variables admitted to X^*," but it won't do him much good if he attempts to take his framework into the courtroom. This is the problem I alluded to at the beginning of this comment. When statisticians use labels with nonstatistical, value-laden meanings to interpret coefficients and variables in an abstract statistical model, they cannot hope to advance statistical science. Nor can they expect agreement on the interpretation of their statistical efforts in adversarial settings.

ACKNOWLEDGMENT

This work was facilitated in part by Grant SES-87-01606 from the National Science Foundation.

ADDITIONAL REFERENCES

FIENBERG, S. E. (1985). Comments on and reactions to Freedman, statistics and the scientific method. In *Cohort Analysis in Social Research* (W. M. Mason and S. E. Fienberg, eds.) 371–383. Springer, New York.

FIENBERG, S. E., ed. (1988). *The Evolving Role of Statistical Assessments as Evidence in the Courts*. Springer, New York. To appear.

FREEDMAN, D. A. (1985a). Statistics and the scientific method. In *Cohort Analysis in Social Research* (W. M. Mason and S. E. Fienberg, eds.) 343–366. Springer, New York.

FREEDMAN, D. A. (1985b). A rejoinder to Fienberg's comments. In *Cohort Analysis in Social Research* (W. M. Mason and S. E. Fienberg, eds.) 385–390. Springer, New York.

MICHELSON, S. (1986). Comment on "Regression analyses in employment discrimination cases" by D. A. Conway and H. V. Roberts. In *Statistics and the Law* (M. H. DeGroot, S. E. Fienberg and J. B. Kadane, eds.) 169–181. Wiley, New York.

Rejoinder

Arthur P. Dempster

1. FRANKLIN FISHER

Much of Franklin Fisher's commentary consists of adversarial argumentation of a sort often heard in courtrooms. In my paper, I mainly kept discussion of active legal processes in the background, because the issues I was discussing were intended to be primarily scientific. But I accept that it is fair tactics on his part, given that our relationship apparently continues to be adversarial in the scientific realm, to bring out that my practical experience was primarily in advising counsel and testifying on behalf of defendants (i.e., employers), while he served on behalf of plaintiffs (i.e., in some cases one or more employees who believed themselves to be victims of discrimination, or in other cases the government acting on behalf of a protected class of employees whether or not grievances had been registered).

That we chose sides as we did is presumably not a chance result. For my part, I believe that the explanation has nothing to do with a predilection to find for one side or the other. Rather, my preference resulted from a conviction that the statistical strategies typically pursued by plaintiffs in employment discrimination cases were serious flawed, as I continue to believe. No doubt Fisher can offer a parallel explanation for his choice of side. But the symmetry ends there, for he evidently feels that the validity of direct regression methods is such that plaintiffs' cases are often proved by statistical arguments, whereas my expert view of the epistemic deficiencies of many plaintiffs' experts' statistical arguments suggests that no statistically based judgments should be reached until the defects in the arguments are repaired. The repairs will be difficult and demanding in terms of commitment of professional resources, because they require the construction of credible Bayesian formulations, but I surely hope that the required investments are made as soon as possible.

As I see our debate, Fisher's position crumbles if inferences about discrimination cannot reasonably be drawn from ordinary direct regression analyses. So the central need is to examine carefully the arguments pro and con this issue. First I will analyze Fisher's treatment of my theory, and show that he made a simple algebraic mistake which caused him to misinterpret my formulation in terms of a different formulation more familiar to him. Hence he neither addressed the substance of my theory, nor found a flaw in its consequent formulas for statistical bias in estimates of discrimination effects, whether the estimates are of the direct or reverse regression varieties. Second I will review the argument he does make for the unbiasedness of direct regression estimates, and argue that he comes perilously close to being self-contradictory.

Fisher opens with the assertion that it is "not hard" for him to "see what it is" that I am "really saying." He proceeds to exposit the initial formulation of my theory, and to reproduce my formulas. My (1) becomes his (1), my (3) becomes his (2), and my (3*) becomes his (4). So far so good. He then combines his (3) and (4) to obtain his (5), while my (6) resulting from the same combination differs from his (5) in that no e^{**} appears in my (6). My algebra is correct and his is wrong. Moreover, my (6) concisely expresses exactly what my theory says that the decision makers are doing when they Bayesianly set reward Y, whereas his (5) adds a dice-throwing component to the reward-setting process which is not in my theory.

It is a reasonable speculation, I believe, that Fisher's equation (5) did not send him warning signals because it has exactly the form that users of what I called in my paper "chance mechanism" models expect. It is even commonplace for users of these models to identify the error term with the effects of unavailable information, which agrees with my definition of e^{**}. My analysis indicates, however, that the chance mechanism formulation cannot be so easily motivated, and indeed that careful probabilistic reasoning can lead to something rather different. I am laying all this out in detail in part to encourage readers to take seriously the idea that there are some simple but fundamental issues here, which my paper is designed to expose, and which should repay study.

Next, although Fisher's argument is not mine, let us follow his reasoning, starting from the chance mechanism model in his equation (5). I find that Fisher invites trouble by espousing opposing positions, one for errors in variables, and another for omitted variables (aptly abbreviated to EV and OV by Joseph Gastwirth). Fisher and I agree that direct regression estimates of discrimination effects suffer from statistical bias under the EV model. He argues that the bias problem goes away in the OV case, while I argue that it remains. One reason to expect consistency is that the EV model is just a special type of OV model, namely, one where the true X^* is an omitted variable.

It is important to review the arguments in detail. I start by juxtaposing Fisher's contradictory arguments. The argument for the presence of bias in the case of EV starts from his equation (6), where the measurement error v is judged independent of the true X^* and G. Substituting his (6) into his (5) yields his (7), having the form of a regression equation of Y on G and X whose true gender coefficient is the desired discrimination effect. But ordinary regression analysis yields biased estimated coefficients because the error term in (7) is correlated with X. Proceeding to the OV case, Fisher writes his equation (9) which, opposite to (6), has X^* on the left side and X on the right side. Substituting (9) into (5) yields his (10), where now it paradoxically appears that the desired discrimination effect is unbiasedly estimated by ordinary regression analysis of Y on X and G. The paradox is resolved by observing that one can have simultaneously valid equations relating X and X^*, where one like (6) has error term v independent of X^*, while the other like (9) has error term W independent of X, but, unless both sexes have the same mean X and the same mean X^*, one or other (or more likely both as I would argue) of the relations (6) and (9) must have a gender term on the right side.

Fisher does not explicitly justify omitting the gender term from (9), but the appearance of the phrase "this is likely to mean that" preceding (9) suggests implicit justification by the author's authority. But does the emperor really have clothes? I have the impression that we subjects are being asked to accept (9) on the basis of some vaguely specified "causal" theory put forth by a guru. To me, that's not good enough, which is really why I felt the need to rebuild the theory from the ground up.

Fisher's positions on the central issues are to me almost perfect illustrations of several tendencies to fallacy which I referred to in my paper, and which I should say are widely found in many fields where statistical analyses of observational data are essential, whether the analyses are carried out by statisticians or subject matter specialists. One is the tendency to rush into probability models without thinking through the scientific bases and contexts that give the models meaning, so that subsequent cranking out of manipulations with the models quickly loses touch with the real questions and required uncertain judgments under analysis. A second type of fallacy derives from the unsupported introduction of causal arguments and language, which are then used implicitly as underpinnings for strong conclusions, without ever rigorously facing up to the need to assemble credible descriptions of or arguments for the adopted causal mechanisms.

A third sort of fallacy is illustrated by Fisher when he complains that my position on OV is "tantamount to finding for the employer in all cases," and again when he argues that the inability to separate prejudicial from judgmental discrimination from the data cannot be a valid issue because it would automatically cause all guilty defendants to be cleared. Fisher and those who think like him would apparently advocate accepting rules for deciding guilt and innocence on the basis of statistical data even when careful analysis of the rules indicates large sensitivity to assumptions that cannot be verified from available data. The argument about the guilty always going free is simply a scare tactic. Bayesian thinking shows the proper way. Every uncertain judgment is subject to dependence to some degree on assumptions unverifiable from hard data. The qualifier "to some degree" is the critical point at which judgment is needed, whether the context is scientific or legal. The truly irresponsible position is to advocate statistical rules in circumstances where dependence on plausible prior alternatives is evident and critical.

It is troubling that Fisher counts me among the "proponents" of reverse regression, when my lead into the discussion of the topic states that "we know that it is hopeless to try to solve the problem" of bias in direct regression estimates by switching to another method of estimation such as reverse regression. That is, we know in advance that the problem lies elsewhere, in the lack of required information in the data. It is true that I have written a little on theoretical aspects of reverse regression, and that I think reverse regression analysis can often provide exploratory insights into data, which in turn may contribute to making subsequently constructed formal models more adequately descriptive of observed variation. But on the main point of using reverse regression to estimate discrimination effects I am certainly not a proponent.

It is arguably true, as Fisher maintains, that I am "incorrect" in assigning priority to "fairness 1" versus "fairness 2," as opposed perhaps to "hiring 1" versus "hiring 2," as a motivation for Harry Roberts's independent discovery of reverse regression in 1979. My sense, however, is that the fact of a disagreement between regression lines of Y on X versus X on Y would hardly have seemed noteworthy absent an implicit sense that each embodied some standard of fairness, which in turn might be appropriate under different hiring models. Because my theory starts from first principles, without building on Roberts's concepts, I believe I should leave questions about his early perceptions of reverse regression to Roberts himself.

Finally, regarding Fisher's "plain English" punch line, I can happily cede to him priority for first discussion of the question of whether the statistician's inability to separate α^* into its two components constitutes a legitimate defense for an accused party. I did not address the matter in my paper, and I think it would be premature to do so now, because I think it raises a host of issues that have not been thought through. I do suggest in my paper that we might be wise to inform the makers of laws and policies concerning employment discrimination that the decomposition exists, might have unforeseen consequences and might lead to consideration of alternative policies.

2. ARTHUR GOLDBERGER

I appreciate the measure of agreement and the eloquence of Arthur Goldberger's comments, and I join him in deploring the sad state of communication between our disciplines. I accept that both sides possess and reflect distortions in their views of the other. Rather than argue about which distortions are more grievous, we should regenerate good will and take positive steps to repair the damage. The present situation does substantial harm to what is after all our common science.

The history of my interaction with Goldberger's test may shed light on the current relations between our fields. I first heard of the test early in 1984 during the *Cynthia Baran v. The Register Publishing Company* trial. Goldberger had presented his paper at the Christmas economics meetings, and the plaintiff's lawyer was quick to introduce it into the trial as a means of deflecting attacks on the validity of her side's direct regression studies. My role in the trial was not to defend reverse regression, but rather to try to educate the judge about the basically skeptical views expressed in my earlier paper (Dempster, 1984). Nevertheless, I was surprised that so transparently wrong a procedure was apparently being taken seriously. I wrote a short memo outlining my objections and sent it to Goldberger in April 1984, and he replied amiably and promptly about the neatness of my result, asking if our views could not be reconciled, suggesting that I should not hold all econometricians accountable for his sins, and telling me that his paper would soon appear in the *Journal of Human Resources.* Again I was surprised, this time at the rapidity of publication of a method that had been seriously challenged, but I felt I had done my duty, and passed on to more mainline concerns. My only subsequent connection with employment discrimination cases has been a very minor involvement with the second round of the Harris Bank case in the summer of 1985. I wrote my paper in the summer of 1986, and presented it as an invited lecture at the August statistics meetings. I believe that

I sent a copy to Goldberger and several other econometricians in September 1986. My only related subsequent contact with econometricians since then appears in the comments which I received from the editor of *Statistical Science* in February 1988.

I can only speculate about what factors affect the slow and uncertain process of communication. As Goldberger's commentary about the deficiencies of my perceptions of economists suggests, there is in any scientific culture a hidden semantics within the explicit language of scientific discourse, which the uninitiated must try to discern. As a statistician who deals with many areas of application, I think I am quite sensitized to this issue, not only regarding many relevant substantive aspects of various sciences and professions, but also in the many related but splintered mathematical sciences such as AI, control theory, decision science, risk assessment and of course econometrics. From a statistician's perspective, the discipline of economics is large and richly endowed, organized and entrepreneurially inclined to dominate its turf, relatively turned in on itself, and in particular not much concerned to understand the thought patterns of the most innovative statisticians. Clearly, there is room here for explorations and hopefully joint projects of potential mutual benefit.

On the narrow point of whether Goldberger "was there first" on the inadequacies of both direct and reverse regression, he says only that "it is easy enough to write down" other models. Am I missing secret language here? Surely the issue is what alternative model or models are to be taken seriously. My paper is my first attempt to specify and defend an alternative approach.

3. HARRY ROBERTS, DELORES CONWAY AND JOSEPH GASTWIRTH

I feel honored and grateful that these able colleagues who have struggled with the details of so many important real cases have taken the time to summarize and share their accumulated wisdom. I thank them for basically friendly commentary. Their remarks, together with the literature they cite, and along with Gastwirth's soon to appear two-volume work, could form the basis for a fine graduate seminar on the statistics of employment discrimination and related topics in statistics and the law. I hope that our profession can develop and maintain mechanisms whereby such people can inform policy makers and legislators about the empirical realities of the phenomena they seek to regulate. Otherwise, as seems to be happening in the current debate over handling undercount in the 1990 census, the statisticians speaking out and being heard are often academics with superficial knowledge of the details, while the toilers who best understand how various techniques will work in practice are not sought out and consulted.

4. GAIL BLATTENBERGER, JOHN GEWEKE AND PAUL HOLLAND

These commentators add conceptual depth to the discussion, each bringing in important new dimensions. Blattenberger and Geweke point out that my formulation is simply a shell, and requires development of substantive legal and economic underpinnings before any satisfactory account of how the shell might reasonably be used could be said to exist. My sense is that they agree with me that the shell is a sensible first step, and I welcome their stimulating contributions. Holland on the other hand warns that statisticians may be wandering out of their zone of maximum effectiveness if they engage in deep conceptual analyses of the sort advocated by Blattenberger and Geweke. But I think we need hobbyhorses running in several directions.

I do not have a sense that Blattenberger is more extremely personalist than I, unless she means that I exhibit more philosophical interest than she does in the question of how and why people come to a measure of agreement about their probability assessments. I much appreciate her attempt to relate the "critical legal studies" perspective to the definition of discrimination, and hope to hear more about it.

In the sense in which it is meant, I have no trouble with Geweke's comment about "no causal model" in my article. On the other hand, looking back to the confusion created by even less mindful econometric models, I think it is important to retain (3) as a starting point for thinking about statistical models. I have read a little into labor economics and find only factors and concepts that I think many intelligent analysts would fairly quickly develop. Rather than pursue historical attempts to their existing limits, my instincts are to try to develop working definitions of Y^{**} for specific applied circumstances, and hope that some deeper understanding might emerge from such problem-solving efforts. But I am delighted when Geweke or anyone else tells me about fundamental concepts that I am missing.

Holland also wishes to operate in a problem-solving mode, but for reasons which I do not understand he advises me not to think about the real scientific meaning of the causal processes which underlie measured causal effects. I fear, however, that statistics will deservedly disappear as a discipline if all we do is conduct and analyze randomized sampling and experimentation. Nor will adding decription and prediction suffice to bring the best minds into the field. We need also to make contact with the powerful theories which drive the empirical sciences, and understand how to bind those theories to sound statistical thinking.

5. STEPHEN FIENBERG

I believe that Fienberg is making a point similar to that of Blattenberger and Geweke when he avers that he has difficulties following my "attempt" to relate my models to the phenomena of discrimination. I too wish I had been able to untangle all the complex issues involved. I am somewhat dismayed, however, that supposedly in contrast to me he "would argue for a framework which allowed the statistician to focus" on the decisions and input to decisions. For me, that precisely characterizes what my model does, no more and no less.

I am sorry that he has difficulties with my "language." If he feels that my term "judgmental" is "nonstatistical" and "value laden," then I invite him to suggest a better one. But, evidently, he would first need to overcome his "difficulty in understanding" the concept. He chides me for not explaining "the role of statisticians as expert witnesses." That was not my chosen topic, but the explanation, if one is needed, is that the expert tells it like it is, subject to the limits of his or her abilities and resources.

I am not surprised that Fienberg has "ongoing concerns" about statistics in the social sciences and in litigation. But surely he has been around long enough not to react in "horror" or be "appalled" at widespread misuses of statistical methods. The story about the "large southern employer" is rather banal, and I did wonder if the long quote was really a quote. But mostly I am puzzled about the point. Surely it is the lawyers for the plaintiffs who should have turned up the smoking gun, and realized they would not need a statistician.

In short, Fienberg's style is provocative, but the other discussants contributed more substance in raising and debating core issues. Again, thanks to all for their participation.