Comment: Causal Mechanism or Causal Effect: Which Is Best for Statistical Science?

Paul W. Holland

My current hobbyhorse is to promote the view that statistical science does more good in the world when it concentrates on the careful *measurement* of the *effects* of causes than when it attempts to *explicate* the *causes* of effects. Well-founded measurements of causal effects are the building blocks of the successful identification of causes. Causal effects come first, not last, in the difficult process of causal inference (Holland, 1986b). In this admirable contribution to the statistics of employment discrimination Dempster seems to be riding in the opposite direction. He gives, in his words, "an explicit view of the basic mechanism of reward determination which is at best left implicit in traditional econometric models." This is a causal mechanism, i.e., an explication of the causes of the salaries that employees receive. Does Dempster's paper convince me that I should turn my hobbyhorse around and ride off with him, identifying causes at every opportunity? I don't think so, and I shall try to show how a significant portion of what Dempster accomplishes can be articulated within the structure of what I call Rubin's model (Holland, 1986a; Rubin, 1978) and does not really require explicit causal mechanisms.

A notion of employment discrimination can be developed along the lines of Rubin's model that illustrates how difficult it is to justify much of what passes as statistical or econometric analysis of this problem. The idea is quite simple—the effect of discrimination on a person's salary is the difference between their salary and what their salary would be if there were no discrimination. Such a position assumes (a) that the person's current salary is obtained under conditions of some relevant amount of discrimination, and (b) that a "control" condition of "no discrimination" can be conceived of in which the person would get a possibly different salary.

To develop some notation, let U be a population of employees and let $u \in U$ denote a particular employee. Then we have

$Y_d(u) = u$'s salary under the current, possibly
discriminatory, system $= d$, and

$Y_c(u) = u$'s salary if there were no discrimination

Paul W. Holland is Distinguished Research Scholar and Director of the Research Statistics Group, Educational Testing Service, Princeton, New Jersey 08541.

(c is for "control" in the sense used by Dempster). We also let

$$G(u) = \begin{cases} 1 & \text{if } u \text{ is male,} \\ 0 & \text{if } u \text{ is female.} \end{cases}$$

Dempster's Y is my Y_d, his Y^* is my Y_c and we both use G to denote gender. The primary difference between Dempster's approach and mine is our attitude toward Y_c. For me, $Y_c(u)$ is a number that is typically not observed. For Dempster, $Y_c(u)$ is an employer's posterior mean of another unobserved quantity, $Y^{**}(u)$, which is u's "true worth" to the employer. For both of us, $Y_d(u)$ is u's salary, and is a known value.

In terms of Rubin's model the causal effect of discrimination is the difference between $Y_d(u)$ and $Y_c(u)$, i.e.,

(A) $$D(u) = Y_d(u) - Y_c(u).$$

Thus, $D(u)$ is the difference between u's current salary and what u's salary would be if there were no discrimination. I can think of no clearer definition of the effect of discrimination on u's salary.

In the structure of Rubin's model, *causal theories* are specifications or partial specifications of the values of the responses, Y_d and Y_c. Dempster's equation (4) is a very simple causal theory; it is

(B) $$Y_d(u) = Y_c(u) + \alpha' G(u) \quad \text{for } u \in U.$$

Dempster's causal model (B) yields these causal effects of discrimination:

(C) $$D(u) = \begin{cases} \alpha' & \text{if } u \text{ is male,} \\ 0 & \text{if } u \text{ is female.} \end{cases}$$

Hence, due to the way Dempster has parameterized the problem, there is no causal effect of discrimination for females, whereas males have a constant discriminatory increment, α', added to their control salaries, Y_c, to produce their current salaries. It is not my purpose here to criticize this simple model but merely to show what Dempster's equation (4) means in terms of causal effects.

The question then arises as to what can the data say about α'? To begin, what are the data? We can certainly measure $Y_d(u)$ and $G(u)$. Unfortunately, $Y_c(u)$ is not directly observed in typical employment discrimination cases. Dempster also includes a vector, $X(u)$, of other measured variables thought to be

---