Rejoinder

James S. Hodges

Several themes emerged in the comments; this rejoinder is organized as a discussion of those themes. In the sequel, section numbers refer to the paper under discussion.

1. REPRESENTATION

The main idea that I gleaned from de Finetti (1974, 1975)—and the idea that was to be brought to practitioners—was the idea that all uncertainty can and should be represented as probability. This idea has practical and normative implications. On the practical side, it and the taxonomy of uncertainty in Section 2 suggest a strategy of allocating resources in analysis (to which Huber alluded, and which was discussed in Section 3.1). In this regard, what matters is not how large the various kinds of error are over one's lifetime (Madansky), but how large they are in the problem at hand—if you take care of the latter, the former will take care of itself. De Finetti's idea and Section 2 also provide a framework for communication among members of a team (Smith, 1986, and his comments above).

On the normative side, this central idea requires practitioners to acknowledge all of the sources of uncertainty in an analysis and to incorporate them explicitly in choices made in the course of the analytical work and in the products that arise from it. In the Air Force example (to use the expression of a RAND colleague, Jim Quinlivan), the noise *is* the signal, and it must be reported and used in decisions. This imperative does not imply a "black-box presentation" (Huber), or that one cannot form an attachment to some particularly elegant model (Geisser); nor is it clear that an exhaustive list of models is necessary (Geisser) for an adequate representation of predictive uncertainty. What is clear is that when the time comes for betting on what the future holds, one's uncertainty about that future should be fully represented, and model mixing is the only tool around.

In this sense, I am "more optimistic" about the Bayesian framework than Freedman: in the Bayesian approach all of the types of uncertainty can be represented and discussed in the same language and thus acquire the same importance. In the frequentist framework, this is not the case. But with this Bayesian advantage comes a disadvantage. Taken at face value, the approach generates an infinite regress (Huber, Geisser, Section 2.1) in that expressions of uncertainty are themselves often somewhat indefinite; at some level a Bayesian must make an assertion without further qualification (Huber, Section 2.1).

2. ADDING INFORMATION TO DATA

At this point one can no longer avoid a question that has been glossed over so far: what is being represented? Information—but plainly not just the information in the data (whatever that might mean). A data set, by itself, refers only to itself; in uttering a predictive or inferential statement, we necessarily add assertions to the data set. For one, we assert the relationship of the seen to the unseen, e.g., the relationship of the observations on experimental units to the properties of some unseen mechanism that produces the effect of a treatment. This assertion is usually slipped in implicitly, and it is justified (when it can be) by the design of the experiment, by knowledge of the experimental apparatus and protocols, and so on. But without the addition to the data of this assertion or something like it, any computations done using those data produce only descriptions of the data, not inferences or predictions about anything distinct from the data. (Holland (1986) gives an excellent discussion of different types of such assertions.)

For another example, we represent the relationship of the unseen future to the recorded past, usually with an explicit model. The data themselves do not and cannot support an assertion that future events will arise from the same mechanism as past observations or be otherwise comparable. This assertion must be *added* to the data; it is a judgment, perhaps difficult to criticize, but a judgment nonetheless. (Holland (1986) addresses this as well.) I think this explains de Finetti's argument (alluded to by Geisser) that it is unfair to criticize someone's predictions after the predicted events have passed; you can test predictions, but the legitimacy of the test as a gauge of future predictive power depends on an unverifiable assertion that the past—as represented by the collection of earlier predictions and the standard against which they are evaluated—is relevant to the future, for which a prediction is to be made. Even predictive validation is necessarily subjective.

Thus, I do not suggest dumping cross-validation (Geisser, Madansky), but I do suggest that cross-validators stop kidding themselves about getting something for nothing and figure out what information a cross-validation adds to the data on which it is performed. I am not sure what this information is, but it must involve exchangeability of future and past observables conditional on explanatory variables, for stratified cross-validations, and unconditional exchangeability, for unstratified cross-validations.