the technique of testing hypotheses is vastly overrated in statistics as a method. It isn't so much that the classical methods give the wrong answers, as Berger and Delampady correctly show, as it is that I find the problem ill-suited to help me do statistics better. Thus, I find myself in agreement with Berger and Delampady that "when testing precise hypotheses, formal use of P-values should be abandoned." On the other hand, I do not expect to test a precise hypothesis as a serious statistical calculation.

ADDITIONAL REFERENCES

KADANE, J. B., LEWIS, G. and RAMAGE, J. (1969). Horvath's theory of participation in group discussion. *Sociometry* **32** 348–361.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

Rejoinder

James O. Berger and Mohan Delampady

We are grateful to the discussants for their comments. All raise interesting issues that are highly deserving of discussion. As usual, we will focus on disagreements in our rejoinder.

REPLY TO COX

Professor Cox questions our argument that P-values do not have a valid frequentist interpretation, stating that the "hypothetical long-run frequency interpretation of a significance level seems totally clear and unambiguous." Over many years of trying to understand what makes a valid frequentist interpretation, we have come to agree with Neyman's view that one must have a stated accuracy criterion, a stated procedure and determine the expected accuracy of the procedure in repeated use; thus, an $\alpha = .05$ level test will indeed reject true nulls only 5% of the time in repeated use. A P-value has no such *real* frequentist interpretation. It has various pseudofrequentist interpretations (cf. Cox and Hinkley, 1974), but these are somewhat contorted so that their impact, or persuasiveness, is much less than that of the *real* frequentist justification. Also, a thorough study of our Example 6 is, we feel, very important in understanding the role of frequentism here.

The reaction of Cox to our claim, that ". . . inclusion of all data 'more extreme' than x_0 is a curious step and one we have seen no remotely convincing justification for," is to say that he finds the reasoning clear and precise and at least sometimes relevant. He, of course, is well aware of the many examples in statistics (some due to Cox himself) where inclusion of "other data" in the calculation leads to nonsense. We submit that this is one of those situations, and indeed can marshall (following Jeffreys) purely intuitive arguments against including more extreme data: is it really fair to H_0 to hurl against it not just the (mild) evidence x_0, but also all the much stronger "extreme" values, when these extreme values *did not occur*?

We, for the most part, agree with the remaining comments of Cox. Our statement that "formal use of P-values should be abandoned" was directed to the formal use of P-values in providing quantitative measures of doubt of H_0. At the beginning of Section 5 we agreed that the informal use of P-values "as a general warning that something is wrong (or not) . . . " (to use Cox's phrase) is perhaps reasonable; this informal use in data analysis may well justify the teaching and consideration of P-values.

In regard to "sensible uses of P-values," it is worth considering an earlier comment of Cox to the effect that for "dividing hypotheses . . . the apparent disagreements between different approaches are normally minor." We used to think this, but the discussion of Carl Morris to Berger and Sellke (1987) shows that such may well not be so.

Finally, our response to Cox's Rejoinder 8 or 4′ is what would be expected of Bayesians: We feel that using the Bayesian paradigm will give misleading answers less often than use of alternative paradigms.

REPLY TO EATON

We agree with just about everything in Professor Eaton's discussion, leaving us little to do but applaud the further insights provided. The objectivity issue is indeed a fundamental concern. Eaton argues that objectivity is a vague, ill-defined concept, and may not exist. We agree; indeed, one of the major purposes of the paper was to show that Opinion 2 in the introduction is wrong. Testing a precise hypothesis is a situation in which there is *clearly* no objective Bayesian analysis and, by implication, no sensible objective analysis whatsover. In other problems, arguments about whether noninformative priors are, or are not, objective tend to be inconclusive, but here there simply is *no* prior that can even be called noninformative.

Although the precise hypothesis testing scenario was used to demonstrate that objectivity is at least