We offer a final thought. In many contexts S^* does not arise on the basis of data x, but rather as a 'natural' assumption. Examples include conditional independence, exchangeability, canonical links, identity (or linear) calibration functions. Then, x would be used to check the adequacy of S^* by using predictive distributions.

Andrew Gelman (University of California, Berkeley) **and Xiao-Li Meng** (University of Chicago): We enjoyed this paper and largely agree with the author, even though we would emphasize some of the points slightly differently.

We agree that averaging across competing models is better than choosing just one model, but it is often even better to consider the models as a continuous class (as the author briefly notes at the beginning of Section 5), which to do seriously often requires additional work to ensure that the individual model parameters make sense in the supermodel. We prefer the term 'model improvement' (Gelman and Meng, 1994) instead of mere 'averaging' to indicate the additional information and consideration that goes into creating a sensible larger model.

The author briefly mentions sensitivity analysis as a qualitative method that is improved on by model averaging; we believe that learning about sensitivity of inferences to model assumptions is often an important goal that is not achieved by merely looking at the combined inference. The author seems to recognize this implicitly in Fig. 7. For another example, in the context of dealing with incomplete data, it is typically important to display sensitivity to assumptions about the missing data mechanism; an illustration in an applied context is in Heitjan and Rubin (1990).

The penultimate paragraph of this paper offers a qualitative view of not considering all possibilities in a model. In many cases, we can detect aspects of poor fit by comparing the observed data with their predictions under the model; in Bayesian terms, posterior predictive checks (Rubin, 1981, 1984; Gelman *et al.*, 1995; Gelman and Meng, 1994). Even in the Bayesian context, model checking falls outside the 'model uncertainty' framework. In addition, the larger model used in the uncertainty analysis can, and should, be checked against the data. Even if the large model fits, we must keep an open mind about other possibilities.

Finally, the final paragraph of the paper might leave the impression that there is a trade-off between greater accuracy in a smaller model and better calibration in an expanded model. In both the short and the long term, the expanded model should be superior in both accuracy and calibration. The use of 'accuracy' in Section 7 seems also to include the concept of 'precision', which has quite different implications when considering the desirability of prediction procedures. The paper is somewhat unclear on such a distinction (e.g. Section 7.1). It is not that a statistician wants to 'widen the bands' to cover his neck; it is that real uncertainty governs the widths of the bands. This is especially important when explaining a statistical analysis to decision makers, such as those who authorized the launching of Challenger.

C. A. Glasbey and G. J. Gibson (Scottish Agricultural Statistics Service, Edinburgh): This paper states the obvious, but it is apparent that these are issues which we all need to be reminded of. It is interesting to see a Bayesian approach being used to increase, rather than to decrease, the uncertainty in prior models. In our modelling work in the Scottish Agricultural Statistics Service (SASS) we must often consider uncertainty.

(a) In most models of agricultural systems, scant attention is paid to uncertainty in any guise. The SASS undertook a study of model uncertainty for the Agricultural and Food Research Council (Gibson *et al.*, 1993). We found a range of examples in which failure to take account of uncertainty led to models whose predictions are as misleading as those considered in this paper. We identified a fourth source of uncertainty in addition to those explicitly mentioned by Draper—the intrinsic stochasticity of many models. For example, the number of failed rings on the shuttle has a binomial distribution.
(b) In an on-going study on the feasibility of developing systems models to aid decision-making in agricultural policy, Gibson (1994) considered predicting the detrimental effects associated with the release of genetically modified micro-organisms into the environment. In this case it was apparent that the various sources of uncertainty in the system, not least of which lay in the form of models for the survivability and transmission of organisms, meant that the uncertainty in model predictions would be so great that its utility was negligible.
(c) Glasbey (1987) investigated the effects on estimates of ED50 (the dose at which 50% of subjects respond) from quantal dose–response data of relaxing assumptions about the tolerance distribution.