diagnostics are the residual and the leverage measure. The first diagnostic is fully discussed in Section 4.5, but the second is hardly mentioned in the paper. It is well known that the leverage measure also plays an important diagnostic role in many different kinds of linear models such as the mixed effect linear model (Christensen *et al.*, 1992) and the generalized linear model (Preisser and Qaqish, 1996). Can the author comment on the possible construction of the measure in hierarchical models?

The author also discusses the Box–Cox transformation for the outcome variable. Atkinson (1985) has described some alternative parametric transformations. Recently, Cook and Weisberg (1994) and He and Shen (1997) proposed a very general class of transformations which can be estimated by using a variety of smoothers and the *B*-spline functions. Their methods often work well and give very similar results in practice (Shi and Fung, 1998). Does the author have any comments about applying these transformations in hierarchical models?

Andrew Gelman (*Columbia University, New York*) **and Phillip N. Price** (*Lawrence Berkeley National Laboratory*)
This paper makes the interesting and important contribution of viewing the pseudodata expression of the hierarchical linear model as a unifying tool in model checking. (The analytical and computational use of the formulation for Bayesian inference is well known; see, for example Dempster *et al.* (1991) and Gelman *et al.* (1995), chapter 13.) More generally, the author notes that hierarchical regression models are more difficult to understand than we might imagine, especially if predictors appear at more than one level in the hierarchy. We would like to echo that point with a statistical anecdote of our own (see Price *et al.* (1996)).

We fit a hierarchical linear model to the logarithms of home radon measurements in Minnesota, with random effects for the counties of measurement. The model thus had variance components at the within- and between-county levels, which we examined to obtain an idea of the model's precision. We then added an individual level predictor — an indicator for whether each house had a basement. Adding the predictor caused the estimated (posterior median) within-county variance to decrease (as expected; houses with basements tend to have higher radon levels), but the estimated between-county variance substantially *increased*, which was completely unexpected. What was happening? After some thought, we realized that the counties with more basements happened to have higher random effects (in the second model). In the first model, much of the variation in county radon levels was cancelled by an opposite variation in the proportion of basements. The increased between-county variance in the second model indicates true variation among counties that happened to be masked by the first model.

This sort of pattern, caused by correlation between individual level variables and random effects, does not occur in non-hierarchical regression. We suspect that the tools developed in the paper under discussion will be useful in understanding this sort of data structure, and we look forward to future work in this area.

Finally, we disagree with the claim in Section 3.2.2 that 'Bayesian diagnostics . . . place an extra layer of mathematics between the analyst and the data'. Posterior predictive checks — comparisons of observed data with their predictive distribution under an assumed model — can be performed graphically and, in fact, can be simpler to interpret than classical methods such as Studentized residual plots (see Rubin (1984) and Gelman *et al.* (1996)).

Jian-Xin Pan (*Rothamsted Experimental Station, Harpenden*)
In what follows I concentrate on three topics that have been neglected by the author.

Masking and swamping effects of outliers
In ordinary regression diagnostics, masking and swamping effects are common (see, for example, Barnett and Lewis (1984)), in which one observation may conceal the importance of another. In the hierarchical models, these effects become more complex because the masking and swamping effects may apply across the hierarchies of the models. In the health maintenance organizations data analysed in the paper, for example, the outlying nature of a state may be masked and swamped by another outlying state or by a small number of plans within the state. For the hierarchical models, how can we effectively deal with masking and swamping effects either within a specific hierarchy or across the hierarchies?

Case influence on covariance Γ
The covariance matrix of the hierarchical models (2.12) is of block diagonal form: $\Gamma = \mathrm{diag}(\Gamma_1, \Gamma_2, \Gamma_3)$ where Γ_i ($i = 1, 2, 3$) are covariance matrices of the data, the constraint and the prior cases respectively. When Γ is unknown, the author suggested using Gibbs resampling methods to estimate Γ. Is it possible