benchmark cusum path is comparable with the T cusum path in terms of smoothness of the path and size of the excursion, then we conclude that the sampler is mixing well [in the direction specified by $T(X)$, to be precise]. Otherwise, we conclude that the sampler is not mixing well, in the direction specified by $T(X)$. When two Markov chains are compared for the same target distribution, one may omit the "benchmark" cusum path plot.

Now we are ready to illustrate the use of the cusum path plot in the Ising model example in Gelman and Rubin (1992a) and in the prostate cancer example from the article by Besag, Green, Higdon and Mengersen. Note that we know that the mixing speed is slow in the Ising example, and Besag, Green, Higdon and Mengersen have concluded that there seems no significant multimodality problem in the prostate cancer example.

For the Ising model, professor Andrew Gelman kindly provided the two runs which appeared in Gelman and Rubin (1992a). For $n_0 = 1{,}000$ and $n = 2{,}000$, the sequential and cusum path plots are in Figures 1–3. Each of the cusum plots shows clearly that the mixing is slow, while each of the sequential plots suggests that things have stabilized.

For the prostate cancer example, the authors kindly offered the simulation data presented in their paper. For $n_0 = 2{,}000$ and $n = 7{,}000$, we monitored the 49 log-odds ratios ξ_{ij} and the corresponding reconstructed z_{ij}. The cusum path plots for all 98 parameters compare well with the benchmark plots, indicating good mixing behaviors, consistent with the claims of Besag, Green, Higdon and Mengersen. In this note, I include only the sequential and cumsum plots for two of them: $\xi_{7,1}$ and $z_{7,1}$ (Figures 4 and 5). The cusum plots display comparable paths of the data and the benchmark paths, in terms of smoothness and exclusion size. As the authors note in Section 4.2, fast mixing arises because of the block updates and a large sampling interval or gap. Note that, since the θ's, ϕ's and ψ's are themselves unidentifiable, it would be necessary to monitor them via appropriate contrasts. It is interesting to point out the effect on the cusum plots when single component updates are used and in addition the sampling interval is reduced from 50 to 10. Figure 6 shows the results for a burn-in of 20,000 cycles and data collection over a further 25,000 cycles. It is clear that the cusum plots bring out the mixing properties more explicitly than the sequential plots, and in order to obtain valid inference based on MCMC methods, extreme care is needed with convergence diagnostics.

In conclusion, MCMC users have to explore sufficiently the convergence issue before trusting the estimates that the Markov chain gives. Among other diagnostic tools such as sequential plot and autocorrelation plot, the cusum path plot is a simple and an effective device to monitor the local mixing speed of a Markov chain.

ACKNOWLEDGMENTS

This research was supported in part by NSF Grant DMS-93-22817 and Grant DAAH04-94-G-0232 from the Army Research Office.

Rejoinder

Julian Besag, Peter Green, David Higdon and Kerrie Mengersen

We thank the discussants for their contributions and insights, and for raising numerous interesting points. We shall respond to these as best we can, although obviously there are many questions for which, as yet, only partial solutions exist. We shall also try to rectify some misunderstandings that have arisen as a result of possible ambiguities in the paper. Our response is organized primarily by topic, rather than by discussant.

"ON BEING BAYESIAN"

Separation of Concerns

We have pondered Geyer's call for a separation of concerns, particularly between philosophy and computational technology, and we agree that the aim is an attractive one, but have come to a different conclusion, because in this case there are interactions that are too strong to be discounted. For example, the agricultural experiment in Section 5 of the paper is concerned with ranking and selection in comparing 75 varieties of spring barley. We contend that here it is a point of philosophy that the Bayesian paradigm provides an approach that is more useful than (indeed, we would say vastly superior to) any non-Bayesian approach. However, even in quite straightforward formulations, it is exceedingly difficult to implement a fully Bayesian analysis without MCMC. The simultaneous credible regions in the paper provide another example,