procedures, is the practical one that the resulting image, in being non-negative, automatically represents a valid solution. This results from the presence of log p_i in the formula for entropy. With other procedures, the values obtained for some p_i might be unrealistically negative. Even then, however, the complete solution could well be quite adequate in many applications and could well be attained with comparatively little computation.

In most applications of regularization, the regularization functional, exemplified in this method by entropy, S, incorporates some prior belief about the local smoothness of the true image. For example, suppose the pixels are ordered so that p_i and p_j are likely to be similar if i and j are close together; then one might consider, as an alternative to S, a criterion such as $-\sum(\log p_i - \log p_{i+1})^2$. At a higher level, one will know, *a priori*, that Fig. 1 is a smudged photograph of a car. The entropy functional does not permit representation of spatial ideas of this type. Certainly the cross entropy $S = -\sum p_i \log (p_i/m_i)$ allows for some acknowledgement of a prior model m but leaves the question of how to obtain m in any given application. There is the *ad hoc* suggestion of using the data to produce a preliminary m, but this would require using the data twice, first to find m and then to carry out the regularization. This is pragmatically reasonable, if not inevitable, but cannot be regarded as being formally respectable.

My principal point, however, is to dispute the main thrust of Skilling's article, that the maximum entropy method for data analysis, as used in the production of Fig. 1, stands in splendid isolation, on fundamental grounds. This statement is based on the axiomatic work of Shore and Johnson[2] and of Tikochinsky *et al.*[3]. They show that the choice of a probability distribution required to fit certain linear constraints has to be made on the basis of maximum entropy if certain axioms have to be satisfied. Typically, the effective constraints in the image-processing problems are highly nonlinear, and are expressed by the intensities themselves, not their proportions. It is not obvious that the mathematics of Shore and Johnson carry over to such a problem, and I cannot accept that their axioms, particularly those of subclass and system independence, are relevant to the regularization context. Tikochinsky *et al.*, in their discussion of reproducible experiments, specifically require the p_i to be probabilities; the assumption of linear constraints is vital, and even their definition of reproducible experiments does not correspond with the interpretation given by Skilling.

Thus, the maximum entropy method for data analysis is a useful tool, with some practical advantages and disadvantages over other regularization procedures. It is mildly interesting that there are links with the formal work on the principle of maximum entropy but the parallel simply has not been shown to be exact enough to disallow flexibility in choosing a procedure to clarify Fig. 1 of ref. 1. As 'consistency', as defined by Shore and Johnson, seems irrelevant in this context, we need not fear the charge of inconsistency by rejecting entropy.

D. M. TITTERINGTON
Department of Statistics,
University of Glasgow,
Glasgow G12 8QW, UK

1. Skilling, J. *Nature* **309**, 748–749 (1984).
2. Shore, J. E. & Johnson, R. W. *IEEE Trans. Inf. Theory* **IT-26**, 26–37 (1980); **IT-29**, 942–943 (1983).
3. Tikochinsky, Y., Tishby, N. Z. & Levine, R. D. *Phys. Rev. Lett.* **52**, 1357–1360 (1984).

SKILLING REPLIES—I find the arguments for using maximum entropy to be deeply compelling. In fact, I am unable to improve on Titterington's own words: "the maximum entropy method ... stands in splendid isolation, on fundamental grounds". Quite so. However, he takes me to task for not going beyond the description of an image as a set of probabilities/proportions to incorporate "some prior belief about the local smoothness of the true image". In this he has raised a most important point: clearly, the simple entropy formula $S = -\sum_i p_i \log (p_i/m_i)$ does not include any such belief; just as clearly, maximum entropy is losing power by ignoring this, provided of course that it is realistic to expect smoothness, or spikiness, or cars, or whatever in one's images.

I hold that these more complicated situations should ideally be handled by extending a technique which is basically correct rather than by invoking *ad hoc* alternatives. As Titterington's preference for smoothness tends to zero, I would like his formulae to reduce to the simple entropy form which is the compelling solution to that problem. This is because the regularization function, which expresses my preference for different shapes of image, should be independent of the particular form of the data to be measured. Logically, my preferences precede the data, so that I should use the same function whether I have convolution data (as in the car example) or marginal data (as in the theoretical discussion).

Fortunately, maximum entropy can include such preferences in an easy and natural way. The trick is to develop the identification of p with the image. The simplest identification is to let i be a single index ranging over the cells of the image, and p_i the corresponding proportion of flux. However, we can also let i be a composite index ranging over pairs of cells, and p_i the corresponding product of fluxes (itself a proportion). The model m_i can now incorporate pair correlations between cells as well as simple position-dependent information. If we take i to be a highly composite index, the model can encode correspondingly subtle details of prior knowledge. The arguments for using maximum entropy still hold, and we have been exploring this development. True, I do not (yet?) know how to encode 'smoothness' in an absolute as opposed to *ad hoc* fashion. Nevertheless, I hope I have answered constructively Titterington's first major comment.

Concerning the relevance of Shore and Johnson's axioms[1], I do not think that the distinction between linear and nonlinear constraints is important here. The form of the experimental constraints is quite separate from the form of one's prior belief about the observed image, as coded in the regularization formula. Shore and Johnson explicitly state that their mathematics include nonlinear constraints. They use the technical trick of replacing a (convex) nonlinear constraint by that set of linear inequality constraints which defines the hull of the nonlinear constraint. Tikochinsky *et al.*[2] do not state this, having had physical applications in mind, but clearly the same trick can be used.

If the experimental constraints include dimensional information on intensities, the maximum entropy image will be accompanied by a dimensional number describing its normalization. Defining the regularization in terms of proportions rather than intensities merely requires the resulting image to have the same shape whether the data refer to microwatts or megawatts. The point about the subclass and system independence axioms is not that they hold for an arbitrary collection of data (such as a blurred photograph), but that they are compelling for particular collections (such as marginals). If one particular type of data forces me to use entropy, then I will consistently use entropy for other types also.

I admit that the use of entropy in data analysis imposes an interpretative gloss on the Shore and Johnson[1] and Tikochinsky *et al.*[2] papers, but I was writing a News and Views article, and I believe that these authors themselves approve of the application. We have to do something in data analysis: I am prepared to use probabilistic results to help to find sets of proportions. Probabilities—not that we need them anyway—and proportions are isomorphic: probabilities are just proportions in a sample space, and to each set of proportions there corresponds a probability distribution. Finally, I return to Titterington's remarks defending flexibility of choice. Maximum entropy using the composite (multi-cell) identification allows great flexibility in choosing the analytical procedure, but it also gives a framework within which choices can be related quantitatively to specific types of prior knowledge and discussed in a logical as opposed to a pragmatic fashion.

JOHN SKILLING
Department of Applied Mathematics and Theoretical Physics,
University of Cambridge,
Cambridge CB3 9EW, UK

1. Shore, J. E. & Johnson, R. W. *IEEE Trans. Inf. Theory* **IT-26**, 26–37 (1980); **IT-29**, 942–943 (1983).
2. Tikochinsky, Y., Tishby, N. Z. & Levine, R. D. *Phys. Rev. Lett.* **52**, 1357–1360 (1984).