REJOINDER

PETER J. HUBER

Harvard University

The volume, diversity and quality of the discussion is overwhelming. Some six years ago I had felt there was more to PP than what appeared from the few and rather inaccessible papers. To rekindle interest, I then invited Jerry Friedman to give a talk on PP in a data analysis session I was organizing for the 1979 IMS Annual Meeting in Washington, D.C. As it turned out, this invitation triggered the invention of PPR by Friedman and Stuetzle. The ensuing chain reaction makes me feel like the sorcerer's apprentice—but I hope that the current excitement over PP will leave more than a big wet spot behind it!

There is hardly any overlap between the fifteen contributions, and I cannot possibly respond to all points raised in the discussion. I am particularly grateful to *Friedman* for adding comments on the early history and motivation of PP, and on recent, as yet unpublished developments in PPR.

Some contributions to the discussion exemplify how PP stimulates new ideas and helps us view existing approaches in a new light (those of *Bookstein* and *Wahba*, parts of *Diaconis* and *Hastie and Tibshirani*'s). Naturally, each new development raises a few more problems of interpretation and of sampling theory, and not having had time to play around with any of the new ideas, I am not ready to comment on them, apart from a brief remark on *Diaconis*' contribution. He sketches the extension of the PP idea to discrete groups (finite geometries). Here, the mathematics is much cleaner than in the case of continuous PP, and we can hope to have definitive results sooner. Diaconis mentions a suggestion (by Tukey) to reduce this case to the continuous one by mapping discrete data into Euclidean space in various ways. At first, my reaction therefore was rather skeptical, even though the reduction may help with the intuitive interpretation; but as a matter of fact, it may also help in a purely mathematical sense (namely by tapping into the theory of group representations, but I guess I need not remind Persi of this!).

I am happy that some contributions already begin to tighten the many loose ends left dangling by my paper (in particular the theoretical contributions by *Cheng and Wu*, and by *Donoho et al.*). Others explain why certain proposals that had appealed to me do not work (*Buja and Stuetzle*, *Friedman*). Some contain interesting historical footnotes (*Miller*, *Tukey*). Some relate independent, parallel experiences and conclusions (*Jones*). One is concerned with the strategy of PP (*Switzer*), and one with the relations to CT (*Shepp*). Finally, to bring realism into a possibly overheated topic, two contributions provide healthy, cooling restraint (*Cox* and *Gnanadesikan*).

A murky theme hidden beneath the surface of PP is the psychology of visual perception. I had not dared to touch it in the paper, but several discussants now force it into the open. *Buja and Stuetzle*'s comments on how much we can see in a "Grand Tour" are interesting. I believe the real issue here is that our perceptual system is at its best when we watch a *slowly* changing *three-dimensional* scene, and that it tries to fit visual input into such a framework whenever possible. We run into curious optical illusions if we disregard this fact (e.g. if we watch real-time PPR, then the *flat* curve representing the current smooth in the screen plane is perceived as a snake wiggling in 3-space, and is virtually uninterpretable). In order to utilize this human facility in a systematic fashion, we must make sure that the "Grand Tour" locally behaves like an ordinary rotation in 3-space, and that additional dimensions are brought in only slowly. This way, we can get 3 dimensions (or perhaps even 3.5?) for the price of 2. It still may take us several seconds to absorb a three-dimensional scatterplot—we may have to walk around a 90° arc, or so, to see the essential features—but I would estimate that this trick can speed up data inspection by an order of magnitude, and it robs the squint angle argument of some of its force. There is a recent movie by Buja, nicely illustrating these issues. *Diaconis* wonders about the reasoning behind my formulas for the number of projections to be visited. The formulas appear in Tukey and Tukey (1981, page 216); I do not know how they derived them. I had found them independently through crude arguments of the type: for $d = 2$, the number of one-dimensional projections is $A = 180°/(2 \times \text{squint angle})$, or roughly 10 for a squint angle of 10°, and the number increases—crudely—by a factor A for each dimension. If we accept A^{d-1} for the number of one-dimensional projections, then the approximate number A^{2d-4} of two-dimensional projections is found by taking a random direction (there are A^{d-1}), and a second random direction orthogonal to it (there are A^{d-2}), and then dividing the product of the two numbers by A (to correct for rotations in the plane of projection). I was surprised to see my numbers fall between Matthews' bounds.

I agree that the RANDU example is atypical. Though, with statistical data, RANDU-like planes frequently occur as artefacts caused by grouping or rounding, and while such planes are uninteresting, they are also distractingly conspicuous. By the way, the experiments mentioned in Section 5 after (5.17) were unable to find the RANDU-planes, all our projection indices smoothed them away. Our eyes do not smooth; on the contrary, they seem to use some sharpening when they encounter such structures.

The main interest of such examples (and possibly of PP itself!) may be less in data analysis than in human vision: how can we analytically characterize structures that trigger an attention reaction in our visual system? The "usual" projection indices obviously capture only part of the story; they may find clusters, but they ignore texture.

Miller's contribution recounts one of the first successful applications of three-dimensional displays to statistical data. It also is most relevant as a case study in the psychology of vision. In a certain way, his account can be taken as an argument *against* projection pursuit: the "rabbit head" structure of the diabetes data is strikingly visible in three dimensions and it then helps you understand the two-dimensional coordinate-wise projections. Once you have seen the structure in three dimensions, you very clearly see it in two, but it is hard to sort things out on the basis of the two-dimensional projections only. In other words, a two-dimensional PP search for interesting projections would not have helped with this particular data set.

Thus, *Cox* certainly is right when he stresses that the more interesting features do not appear in one-dimensional summaries (cf. also Section 7). Though, I believe what first triggered my interest in PP more than anything else, was that Friedman and Tukey (1974) had been able to find interesting structure by mechanically optimizing over one-dimensional projections.

Gnanadesikan and Kettenring equally rightly stress that PP is not a panacea (there are no panaceas in data analysis), and it would be a mistake to view it as a substitute for principal component type and other methods that do not require numerical optimization. I should worry if my paper has evoked this impression. The purpose of affine invariant projection indices is to discover structure that might have slipped through the mesh provided by the more traditional approaches. It still remains to be seen how successful they will be; before we have some theory assisting us with the automatic tuning of these methods, we cannot even try them out on a wider variety of data sets. By the way, the intuitively appealing description of the Friedman-Tukey projection index as a combination of two separately interpretable factors (scale and clumpiness) is, at least in a certain sense, misleading; the scale factor is needed to counterbalance an inverse scale implicit in the definition of the second factor, so as to make the total index scale invariant.

Miller and *Gnanadesikan and Kettenring* all express some concern about how the analyst should assess the reality of structure he or she believes to have perceived in the data. But possibly the question is wrongly posed; seeing an interesting and unexpected feature in the data—whether it be spurious or not—is the beginning of a new line of investigation, and while it may induce us to abandon the previous line, it does not involve a terminal decision in the sense of decision theory in the old one. The question may be rather: how many false leads do we want to follow up for the sake of a true one? Moreover, visual data analysis here conflicts head-on with the old adage that you should not look at the data before you decide on the test to be used, and for good reason. Our visual system continually performs a kind of informal test of the null-hypothesis that nothing interesting is to be seen, and I guess that evolution of the species and past experience of the individual have adjusted the critical levels in such a way that errors of either kind—false attention reactions or nonreaction to an important stimulus—do not happen with an unreasonable frequency. Since we cannot assign formal P-values to the strength of such stimuli, we cannot perform the proper conditional tests, given that an attention reaction has occurred; but we can be sure that the actual (= conditional) level of a statistical test will be way above the nominal one. Compare Huber (1983) for some additional thoughts on this matter.

One frequently hears the complaint that the directions found by PP, and especially by PPR, are very difficult to interpret. *Hastie and Tibshirani* now demonstrate that in actual data these directions sometimes *are* uninterpretable, by virtue of being undetermined, and that it is possible to check this through straightforward bootstrap methods. This is a very interesting methodological contribution.

It is gratifying to learn that *Jones* has independently reached conclusions similar to ours. I should however comment on Jones' remarks concerning the usefulness of three-dimensional projections. This depends very much (i) on the display equipment you have (unless it is of the top grade, you invariably are disappointed) and (ii) on the kind of data. Whenever we had data sets consisting of points in natural 3-space (e.g. positions of galaxies, earthquakes or atoms, to mention data spanning size differences exceeding 30 orders of magnitude), we found it difficult to live without three-dimensional views. With the "statistical" types of data (where the variables are geometrically unrelated), often the collection of all coordinate-pairwise scatterplots already is good enough (and easier to interpret than more general projections). Though, there are exceptions, compare in particular *Miller*'s contribution, and some of the comments by *Buja and Stuetzle.*

I agree with *Switzer*'s recommendation about the separation of location/scale from the search for other structure (also in the case of density estimation). The only serious exception to this principle, and to affine invariance in general, I would make in cases where part of the data is categorical in nature.

We found that a simple visual-manual version of PP can be a highly effective method to search for clusters, and that these clusters, once found, usually can be best described and interpreted in projections where only very few variables have nonzero coefficients, in agreement with Switzer's suggestions. Although, and this was rather unexpected, we often might not have initially recognized the clusters as such in these projections; they would separate better (and hence be more conspicuous) in slightly tilted views.

I do subscribe to *Shepp*'s philosophy and particularly would like to endorse his last sentence. In order to understand what a technique can and cannot do, we must have some (mathematical or nonmathematical) theoretical understanding, and we must construct synthetical—and find actual—data sets where it works well, moderately well or not at all, and we must know why.

REFERENCE

HUBER, P. J. (1983). Data analysis: in search of an identity. Neyman-Kiefer Conference, June 1983, Berkeley.

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138