THE PHILOSOPHY OF EXPLORATORY DATA ANALYSIS*

I. J. GOOD†

Statistics Department
Virginia Polytechnic Institute and State University

This paper attempts to define Exploratory Data Analysis (EDA) more precisely than usual, and to produce the beginnings of a philosophy of this topical and somewhat novel branch of statistics.

A *data set* is, roughly speaking, a collection of k-tuples for some k. In both descriptive statistics and in EDA, these k-tuples, or functions of them, are represented in a manner matched to human and computer abilities with a view to finding patterns that are not "kinkera". A *kinkus* is a pattern that has a negligible probability of being even partly potentially explicable. A potentially explicable pattern is one for which there probably exists a hypothesis of adequate "explicativity", which is another technical probabilistic concept. A pattern can be judged to be probably potentially explicable even if we cannot find an explanation. The theory of probability understood here is one of partially ordered (interval-valued), subjective (personal) probabilities. Among other topics relevant to a philosophy of EDA are the "reduction" of data; Francis Bacon's philosophy of science; the automatic formulation of hypotheses; successive deepening of hypotheses; neurophysiology; and rationality of type II.

1. Introduction. Both data analysis (EDA) and confirmatory data analysis (CDA) have existed, under any reasonable definition, for more than a century, but in recent years the distinction between them has been recognized much more consciously by statisticians, partly because of the influence of Tukey (1977).

EDA is concerned with observational data more than with data obtained by means of a formal design of experiments. When data are obtained informally, we are not surprised if the methods for handling them are also often informal, and perhaps EDA is more an art, or even a bag of tricks, than a science. If this is so, it might be difficult or impossible to find a reasonably comprehensive philosophy of EDA. As Cochran (1972) says, in his article on observational studies, "we can claim only to be groping toward the truth".

EDA is an extension of descriptive and graphical statistics so it seems pertinent to quote David Cox (1978, p.5) also. He says "There is a major need for a theory of graphical methods", and goes on to say "Of course, theory is not to be taken as meaning mathematical theory!" Leamer (1978)

*Received October 1982; revised January 1983.

†I am grateful to John W. Pratt for some useful criticisms. This work was supported in part by N.I.H. Grant R01-GM18770.

Philosophy of Science, 50 (1983) pp. 283–295.

uses much mathematics, but he too makes only modest claims about the progress achieved in his interesting book.

It will be argued in this paper that the theory of EDA will at first involve philosophy, psychology, and even neurophysiology. If the theory becomes well developed it might eventually become mathematical also, because all hard sciences are partly mathematical. In the present paper there are only some glimmerings of mathematics.

Let's think of descriptive statistics as including graphical statistics. The purpose of descriptive statistics is mainly to present the salient features of data in an intelligible form. On the other hand, Tukey (1977, p. 1) describes EDA as detective work. In this respect, it goes somewhat further than descriptive statistics, although the distinction is more quantitative than qualitative. The exploratory data analyst knows that he wants to formulate hypotheses, whereas the descriptive statistician is less conscious of it.

There is not a sharp distinction between observation and experiment. For example, astronomy is usually regarded as observational, yet an astronomer can design an experiment in which he plans in detail where to point his telescope. Similarly, when there is a large mass of data, an exploratory data analyst can apply his techniques to random subsets of the data, in an experimental spirit. When this is done, the distinction between EDA and CDA also becomes fuzzy.

In some circumstances, there is not a superabundance of data, and additional data might be unavailable or expensive. This happens sometimes in medical research (e.g. Good 1978), in cryptanalysis, in astronomy (e.g. Good 1969, Efron 1971, Nieto 1972), and in judging the usefulness of a racing or stock market tipster. Perhaps one useful classification of EDA would depend on whether there is a superabundance of data.

To clarify the meaning of EDA still more, it may be useful to say what it is not. It is not primarily confirmatory though EDA and CDA can be combined. Confirmatory statistics consists of experimental design, significance testing, estimation, and prediction, whereas EDA is concerned mainly with the *encouragement* of hypothesis formulation. The philosophy of confirmatory statistics has been much discussed since 1950 (e.g., Good 1980), and I have mentioned it only for clarification.

One aim of the present paper is to try to define EDA more precisely and thereby to point towards a philosophy of the topic.

A referee states that this work "challenges the old positivist dichotomy between 'logic of discovery' and 'logic of justification'".

An earlier version of this paper appeared in Good (1981).

2. What is a Datum? Philosophers of science are usually more interested in defining a science than the scientists are. Scientists prefer to *do* science instead of defining it. For example, *Van Nostrand's Scientific Encyclopedia* has no entry for "science". Because I am groping towards a philosophy of EDA, I shall first try to define "data".

I think, at least as a first approximation, that a body of data can be defined as a collection of vectors or k-tuples that all relate to a single topic. The components of even a single vector can be of different kinds: they might be continuous or discrete; they might represent measurements or counts or ranks or names. Each component, or set of components, might have a distribution. If a component is an integer, it might be a label for various sub-populations.

Let's take an $r \times s$ contingency table as an example. It can be regarded as a set of 3-tuples or triples $(1,1,n_{11})$, $(1,2,n_{12})$, . . ., (r,s,n_{rs}) where (i,j,n_{ij}) means that there were n_{ij} items in category (i,j), but we don't usually *think* of a contingency table as a set of triples.

The description of a data set as a collection of k-tuples does not cover all cases. There might also be relationships between the k-tuples. If a triple (a,b,c) is paired with another triple (d,e,f), then we could say that $\{(a,b,c), (d,e,f)\}$ is a datum or an element of our data set. Or several k-tuples might be related by belonging to a subset of the data set. Moreover, there might be relationships between the subsets, and the subsets might have a hierarchical structure known in advance of any analysis. Thus a datum might be illustrated in a "hypergraph" as in Figure 1, with some complexity. The representation of Figure 1 in a form suitable for input to a modern computer would have less appeal to human intuition.

3. What is Descriptive Statistics? The example of a contingency table tells us something about the nature of descriptive statistics. It is all right to think of a contingency table as a collection of triples if we want to feed it into a computer, but for presentation to a human it is better to write the table in the usual manner because the tabular format enables us to see more clearly what's going on in each row and column. In short, descriptive statistics takes advantage of our spacial perception of tables and, more obviously, of diagrams. Components of k-tuples that actually denote spacial and temporal dimensions are especially often represented spacially.

But descriptive statistics is not just concerned with presenting data to the eye, it is also concerned with the so-called "reduction of the data", such as when averages are computed. Perhaps the earliest example of data reduction in all civilizations and tribes was counting. Instead of describing all the animals on a farm separately, they could be described as say 17 swine and 43 kine. To put the matter somewhat paradoxically, one of the aims of descriptive statistics is to *suppress* data. The uninformative features are suppressed so as to make the important features stand out

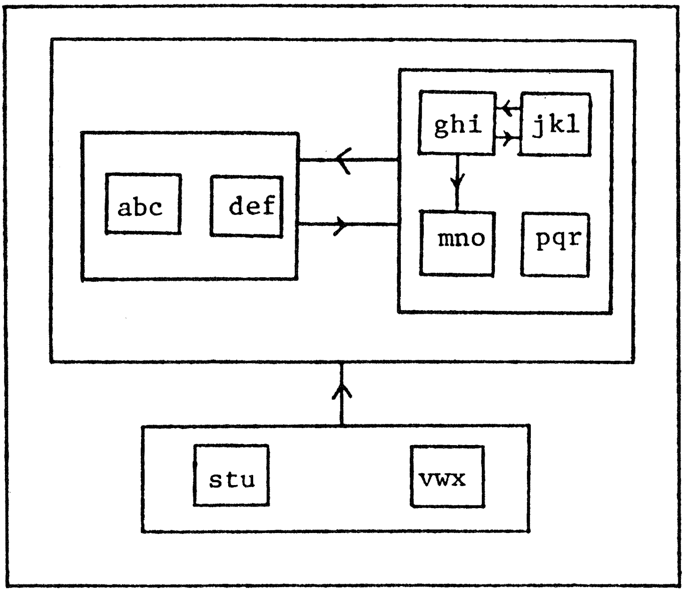

Figure 1. A representation by means of a hypergraph of a fourth-order element of a data set. It is of the fourth order in the sense that the hierarchy has four levels. The arrows denote some relationship such as causal influence.

more clearly. We need insufficient statistics so to speak. An example mentioned by a referee relates to classical statistical mechanics. Even if one knew the position and momentum of each molecule in a gas, it would be necessary to discard nearly all this information to make useful statements about the gas as a whole. Otherwise, one would not be able to see the gas for the molecules.

This feature of descriptive statistics is a feature of all communication. We must suppress some of the truth to communicate the truth. We must tell the truth if it is not misleading, the whole truth that is of sufficient importance, and nothing but the truth apart from some figures of speech.

Communication can be from A to B and it can also be from A to A. That is, when you do descriptive statistics, it might be for your own use. This is even more true for EDA. A detective makes a case for his own use before he presents it to a court.

Let us list some of the techniques of descriptive statistics and then ask what they have in common. Here is such a list: frequency counts, histograms, frequency polygons, scatter diagrams, means, modes, medians, geometric means, quantiles, ranges, standard deviations, graphs (possibly with logarithmic or other transformations), circular pie diagrams, correlation coefficients, contingency tables, measures of association in contingency tables, bar charts, paired bar charts, a variety of monochrome statistical maps, and colored statistical maps.

One quality these techniques have in common is ease of comprehension, by adults in our culture, with very little learning. Adults in primitive cultures, and young children, have difficulty even in counting. Another quality of the techniques is that, for the most of them, there is a deliberate sacrifice of information for the sake of simplicity, as just mentioned. Also, the amount of information that should be sacrificed depends on the application; for example, histograms of various bin widths might be preferred in various circumstances, just as maps should have more or less detail. A third quality is that the techniques that present data to the eye often contain more detail than numerical summaries. The reason is of course that we mammals can usually extract the salient features from visual information more rapidly than from numerical information. See also Kruskal (1982).

In short, the techniques of descriptive statistics are designed to match the salient features of the data set to human cognitive abilities.

Within about ten years, most applied statisticians in America will perhaps work synergistically (not "symbiotically") with computers, as a few cryptanalysts did as early as 1943 when sitting in front of the cryptanalytic machine known as the Colossus. Every few minutes, the analyst would have a new program run, controlled by a plugboard, the choice of program being dependent on the output of the previous program. The process was analogous to that of a doctor questioning a patient, the current question depending on the previous reply as well as on the doctor's judgment.

Synergy between statisticians and computers will not make obsolete any of the standard methods of descriptive statistics, but it will lead to new methods. It is often beneficial to look at data from more than one point of view, and this can be more cost effective with an interactive computer than without. It will even become practicable, if desired, to represent time by time instead of by space; that is, a time series could be represented by a time series on a visual or auditory display, or both. If the display were auditory, then salient features might emerge as "musical" phenomena. I doubt if the temporal representation would usually be better than representing time by space, but it might supplement it.

4. From Descriptive Statistics to EDA. Francis Bacon (1561–1626), in his book, *Novum Organum*, proposed that the basic method of science was the systematic collection and tabulation of observations, and that this would almost automatically lead to the discovery of important scientific truths, provided that certain forms of fallacious reasoning were avoided. William Harvey, the discoverer of the circulation of the blood, who was

Bacon's physician, said sarcastically that Bacon wrote philosophy (science) like a Lord Chancellor. Russell (1946, p. 566) states that Bacon underestimated the difficulty of forming good theories or hypotheses, and Russell made a remark that is less true today than it was when he said it, namely "So far, no method has been found which would make it possible to invent hypotheses by rule". In spite of this, Bacon was highly influential in science, partly because there was too much tendency at the time for people to try to do science by pure thought alone.

I mentioned Bacon because I think exploratory data analysis has the Baconian flavor. EDA tries to go beyond descriptive statistics in the direction of true science, that is, in suggesting scientific hypotheses. In this respect, it is somewhat more judgmental and Bayesian (in the subjectivistic sense) than orthodox confirmatory statistics.

Whereas many elementary non-Bayesian textbooks, written with a pair of scissors and a pot of glue, say that hypotheses should be formulated before looking at the data, the Baconian, the exploratory data analyst, the clinician, the cryptanalyst, the scientist, and the detective, all tend to reverse this advice. Their Rule 1 is "Look at the data", or look at the patient, or look at the cipher messages, etc.

5. Towards a Theory of EDA. Statistics as a whole is more concerned with superficial structure than with deep structure. To discover deep structure in a science usually requires much familiarity with that science. Even an exploratory data analyst cannot expect to obtain truly deep results in a science with which he is unfamiliar unless he cooperates with a scientific specialist. But whether or not the analyst interacts closely with a scientific specialist, the theory of EDA will need to deal with the following two questions:

How should we represent a collection of k-tuples to match the cognitive powers of the analyst so that he can (i) see patterns in the data, and (ii) formulate sensible hypotheses about the data? Part (i) deals more with the aims of descriptive statistics, and part (ii) with the additional aims of EDA. A theory of EDA should therefore ultimately incorporate a theory of cognitive psychology. Meanwhile, when choosing between methods of displaying features of data, we have to rely on introspection, common sense, and on *experimental* psychology. We shall soon return briefly to these aspects.

What criteria should be used, or are implicitly used, for calling a feature of the data "salient"? The main criterion is that the feature should have an appreciable prior probability of being potentially incorporated in a useful hypothesis for explaining the data. Therefore, a philosophy of data analysis needs to be concerned with probabilities of hypotheses, with the nature of explanation, and with the nature of scientific discovery.

One aspect of scientific discovery is "successive deepening" just as in chess analysis (de Groot 1965). A hypothesis is formulated, and, if it explains enough, it is judged to be probably approximately correct. The next stage is to try to improve it. The form that this approach often takes in EDA is to examine residuals for patterns, or to treat them as if they were original data, an approach that has been emphasized by R. J. Freund, Vail, & Clunie-Ross (1961), Anscombe (1967), Tukey (1977), and Daniel (1978). Another aspect of scientific discovery is the use of "divergent thinking", or the avoidance of mental ruts. This too is incorporated in the technique of EDA of looking at data from several points of view just as maps can show political, meteorological, agricultural, or geological information. As the economist Ronald Coase said "If you torture the data long enough it will confess". (Perhaps he should have said that *they* will confess.)

Much EDA is a matter of looking for patterns in data sets. It is believed that the right hemisphere of the brain is the part that recognizes patterns, whereas the left hemisphere is more analytic (see Ornstein 1973, Section II). So perhaps EDA should especially emphasize the right hemisphere, by using graphics (and music?), and CDA the left by using formulae.

Russell's remark that there is no known way to formulate hypotheses automatically might now-a-days be regarded as an argument for more EDA, although there are at least two exceptions to his remark. The first is the use of maximum entropy for hypothesis formulation (Good 1963) (instead of for selecting priors—Jaynes 1957), and another was described in a session on induction at a conference on machine intelligence a few years ago (Michalski & Negri 1977; Good 1977b) and some of the ideas date back at least to 1959 (Good 1959, p. 25). But in 1980 Russell's remark is still not very misleading.

As mentioned five paragraphs ago, one way to decide whether a technique in descriptive statistics is likely to be useful is by introspection, or common sense, that is, by judging whether it is likely to be of wide appeal, perhaps in terms of its appeal to oneself. For example, if we wish to represent the geographical distribution of a 2-tuple, we might regard it as sensible to represent one component by shades of color and the other component by any standard method for 1-tuples. We might then have spots of various sizes and in various shades of blue. This seems to me to be clearly better than representing each of the two components by shades of two different colors, and then mixing the colors, though I think this latter method would not be bad if the two colors were primaries such as red and blue. There is more on this matter later, under the heading "Multivariate data".

Instead of using introspection to choose effective representations, one can use the methods of experimental psychology as in Wainer & Francolini (1980). Such experiments make use of ordinary statistical methods, so that CDA can be used for research on EDA. So far, I believe introspection has been more important for EDA than experimental psychology has been, but this will not necessarily always be so.

One of the devices for making discoveries is to sleep on a problem. There is no reason why this device should not be used for EDA. Similarly, any other psychological technique for aiding discovery and creativity can be incorporated into EDA (see, for example, Ghiselin 1952, and Hadamard 1949). That such techniques are not peculiar to EDA is not a good reason for ignoring them.

Non-Bayesian statistics avoids the discussion of the probabilities of hypotheses, yet in my opinion the probabilities of hypotheses are fundamental to the philosophy of data analysis. I became especially convinced of this as a cryptanalyst during World War II. As mentioned earlier, Rule 1 for a cryptanalyst is "Look at the data", just as it is for any other exploratory data analyst. A cryptanalyst makes frequency counts and carries out various other standard tests, but is always on the lookout for "interesting" patterns. An interesting pattern, by definition, is one that has a non-negligible subjective or logical probability of being potentially explicable, at least in part. *It is possible to judge that a pattern has an underlying explanation even if we are unable to find it.* It is no good for a pattern to be highly significant statistically if it is, in the terminology of an anonymous cryptanalyst, a "kinkus" (plural: "kinkera") or a "mere coincidence". A kinkus is a pattern with an extremely small prior probability of being potentially explicable, given the particular context. The theory of probability that is most relevant here does not emphasize "sharp" probabilities, but is a theory of partially ordered subjective probabilities, or equivalently of upper and lower subjective probabilities, of the kind emphasized, for example, by Good (1950) and in many later publications. Keynes (1921) emphasized partially-ordered *logical* probabilities.

Let us consider a simple cryptanalytic example that could arise from a standard test, namely a letter frequency count. Suppose we found that the letter E had a slightly lower cipher frequency count than any other letter in a batch of messages originating in Germany. This would be interesting because E is the *most* frequent plain-language letter in German.

We might find it curious that the most frequent plain language letter is the least frequent cipher letter, a property that could hardly be regarded as a kinkus. Could it be that no letter can encipher as itself? Ah yes, that is a property of the Enigma cryptographic machine.

In ordinary statistics, we talk far more about hypotheses than we do about patterns, but in EDA the two concepts are or should be intimately related. When we think we have found a probably potentially explicable pattern, we like to think of one or more hypotheses to explain it, but if we regard the pattern as a kinkus we ignore it. The Titius-Bode law in astronomy is an example of a pattern that some astronomers have taken seriously enough to look for an explanation, whereas others have regarded it as just a fluke or kinkus.

In at least one sense, a good explanation of a pattern or event E is a hypotheses H such that $P(E|H,G) >> P(E|G)$ while $P(H|G)$ is not too small, where G denotes the context. (The vertical stroke means "given". This is the standard international notation in books on statistics and on mathematical probability.) I shall take G for granted and omit it from the notation. Very reasonable desiderata lead to a quasi-numerical or semi-quantitative expression for the power of a hypothesis to explain the event. The expression, called "explicativity", is defined as $\eta(E:H) = \log P(E|H) - \log P(E) + \gamma \log P(H)$, where $0 < \gamma < 1$ and $\gamma = \frac{1}{2}$ seems a reasonable value. See Good (1977a) for the reason for adopting this definition and for some of its implications for CDA. It might be thought that such a formula would be irrelevant for EDA because EDA downplays formal probability models. I interpret this downplaying to mean that, in EDA, partially-ordered subjective probabilities are more relevant than the sharp probabilities used in many forms of CDA. Thus a formula for explicativity might be philosophically fundamental even for EDA so long as the probabilities are only partially ordered. A kinkus is a pattern corresponding to which we believe there is no hypotheses H of reasonable explicativity.

It seems to me then that the aims of EDA are:

(i) *The presentation of the data*. To represent the k-tuples, or the trees, or hypergraphs, that constitute the data, in a manner that is matched to our information-handling capabilities, especially to our vision (and perhaps to our hearing), and to the right hemispheres of our brains.

(ii) *Pattern recognition*. To find patterns that are not kinkera.

(iii) *Hypothesis formulation*. To find hypotheses of reasonable explicativities to explain the patterns.

(iv) *To look for hypotheses of greater explicativities* such as by treating the residuals as new data. If the hypotheses are improved *too* much, the analyst would be doing ordinary science rather than EDA. Thus EDA is intermediate between descriptive statistics and hard science.

(v) *Type II rationality*. To maximize expected utility allowing for the guessed costs and delays of computation and thinking. This principle is always used, but is especially relevant to EDA because of the large variety of possible ways of examining data.

Before stating the conclusions, let us consider a few miscellaneous elaborations of the foregoing discussion.

6. Miscellaneous Comments. *Neurophysiology.* We tend to take for granted that it is useful to present information in diagrams. Clearly one reason is that our eyes can handle a great deal of two-dimensional information, but neurophysiologists might be able to give more detailed reasons. For example, Hubel and Weisel showed that individual neurons of the cat's visual cortex are activated by short straight lines in the visual field, different neurons corresponding to different directions of the lines. A straight line is the simplest curve and has the highest chance of being explicable, and by Natural Selection neural "equipment" has evolved that is useful for EDA because animals have been doing ur-EDA for hundreds of millions of years. Moreover, other neurons are specialized for the detection of patterns that we would regard as almost as simple as line segments. The visual system has presumably evolved so as to tend to ignore minor kinkera. It might be possible to discover in much more detail what is visually simple by experiments in neurophysiology as well as by experimental psychology. A theory of graphics might be largely a theory of the eye-brain system that will be closely related to the theory and practice of artificial intelligence.

Multivariate Data. By defining a data set as a collection of k-tuples, and sometimes as a collection of hypergraphs, we have already implied that we want to represent multivariate data. Among the methods for representing multivariate data is Herman Chernoff's suggestion of using faces with various attributes. I think his idea would be improved if he used faces that looked like good cartoons instead of like childish drawings. In the Orient, oriental faces should be used because all Westerners look alike. The point of the facial method is that faces are the first things that a baby can recognize visually. (Auditory recognition of the mother's voice appears to occur prenatally: see *Science 80*, 1980 Sept./Oct., p. 8.) In other words, Chernoff was trying to match the presentation of data to the statistician's natural powers of perception and to his experience. A disadvantage of the method is the difficulty of translating back from the faces to their meanings.

If the dimensionality of our data is too large, we can try looking at selected subsets of the components of our k-tuples, and we can try other methods for reducing the dimensionality without losing too much information. This is the aim of the method of principal components, of factor analysis, and of multidimensional scaling. In the past, the outputs of these techniques have often been represented in only two dimensions, but more elaborate techniques are becoming popular, and with colored graphics I believe we could achieve a clear representation of at least seven dimensions. For example, a cluster might be detected as a local group in two dimensions of arrows of nearly equal length, thickness, and direction and of nearly the same shade and intensity of green, where the green occurs as a mixture of blue and yellow. The effectiveness of directionality is explicable by the Hubel-Weisel observation. I conjecture that the weakest part of this representation would be the joint use of shade and intensity of greenness and that the effectiveness of color schemes would vary more with the beholder than would the other more spacial attributes. Perhaps an eighth dimension could be achieved holographically.

Simplicity. At first I tried to formulate a philosophy of EDA based on simplicity, but I believe prior probabilities, given the context, are somewhat more fundamental. We like simple hypotheses because they help communication with others and with ourselves, and also because there is a strong relationship between simplicity and prior probability, as Harold Jeffreys emphasized. A simple hypothesis is often more probable initially, *as a good approximation*, than a fixed complicated hypothesis with given values for its parameters (see Good 1950, p. 60), though there are exceptions, such as the logically false statement that $0 = 1$. Similarly, a simple pattern is often a non-kinkus whereas a complicated pattern is often a kinkus. Perhaps complexity theory would lead to a measure of kinkocity, but kinkocity might depend more on the context than does complexity.

Simplicity cannot be defined simply for it is in the brain of the beholder. Briefly, it is brevity. Brevity depends on the "language", where here, in the spirit of Artificial Intelligence, I use "language" to include pictorial information. The appropriate language in some field of knowledge or activity is one that, *on the average*, enables one to communicate and to think about a topic most efficiently. This implies (i) that the concepts in the language should be adequately clear, and (ii) that the important facts can be stated or presented concisely, with just enough redundancy to facilitate the functions of error detection and correction. Having thus defined an appropriate language, then one can define simplicity as brevity, as a first approximation.

In a very general sense, a simple hypothesis for explaining data leads to a smoothing of the data. An example is when small high-order interactions are equated to zero and the original data are then adjusted by an inverse transform. This example was mentioned, in an early paper on fast Fourier transforms, by Good (1958, p. 361) where smoothing was defined as the adjustment of observations "with the intention of making the adjusted values closer than the original ones were to the expectations given the true, but unknown, hypothesis" (see Whittaker & Robinson 1944/67, p. 303*n*.). Tukey's term for the smoothed observations is "the smooth", a convenient usage, but an ugly one at least until we grow accustomed to it. (Golf players talk about "the rough" so why not "the smooth"?)

The value of smoothing, as carried out by the exploratory data analyst, enables one to avoid the explicit formulation of a truly scientific hypothesis while nevertheless "explaining" the data to some extent. Smoothing removes irrelevancies and makes it easier for the scientist to formulate good hypotheses, but the rough can be relevant and bumps, like babies, should not be thrown out with the bathwater.

7. Conclusions. This paper has contained another groping towards a philosophy of data analysis. It has led into some controversial areas of the philosophy of science, such as the theories of probability, explanation, and simplicity. These are concepts that have not yet been separated from personal judgment. If we could formalize judgment in detail, we would stop calling it judgment (Good 1959). That is one reason why we are still groping towards a formal theory. Another reason is that we do not yet have a well rounded theory of the nature of perception and of hypothesis formulation. But the examples and discussion in the present paper, especially the list of aims of EDA at the end of Section 5, might provide some ingredients for bringing home the bacon.

REFERENCES

Anscombe, F. J. (1967), "Topics in the investigation of linear relations fitted by the method of least squares" (with discussion), *Journal of the Royal Statistical Society B 29*: 1–52.

Cochran, W. G. (1972), "Observational studies", in *Statistical Papers in Honor of George W. Snedecor*. Ames, Iowa: Iowa State University Press, pp. 70–90.

Cox, D. R. (1978), "Some remarks on the role of statistics of graphical methods", *Journal of the Royal Statistical Society, Series C (Applied Statistics) 27*: 4–9.

Daniel, Cuthbert (1978), "Patterns in residuals in the two-way layout", *Technometrics 20*: 385–395.

de Groot, A. D. (1965), *Thought and Choice in Chess,* G. W. Baylor (ed.) The Hague & Paris: Mouton. A translation, with additions, of a Dutch version of 1946.

Efron, B. (1971), "Does an observed sequence of numbers follow a simple rule? (Another look at Bode's law.)", *Journal of the American Statistical Association 66*: 552–568 (with discussion).

Freund, R. J., Vail, R. W., & Clunies-Ross, C. W. (1961), "Residual analysis", *Journal of the American Statistical Association 56*: 98–104.

Ghiselin, B. (ed.) (1952), *The Creative Process*. New York: New American Library.

Good, I. J. (1950), *Probability and the Weighing of Evidence*. London: Charles Griffin; New York: Hafners.

Good, I. J. (1958), "The interaction algorithm and practical Fourier analysis", *Journal of the Royal Statistical Society, Series B 20*: 361–372.

Good, I. J. (1959), "Could a machine make probability judgments?" *Computers and Automation 8*: 14–16 and 24–26.

Good, I. J. (1963), "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables", *Annals of Mathematical Statistics 34*: 911–934.

Good, I. J. (1969), "A subjective analysis of Bode's law and an 'objective' test for approximate numerical rationality", *Journal of the American Statistical Association 64*: 23–66 (with discussion).

Good, I. J. (1977a), "Explicativity: a mathematical theory of explanation with statistical applications", *Proceedings of the Royal Society A, 354*. London: 303–330. Also in

Studies in Bayesian Econometrics and Statistics in Honor of Harold Jeffreys, Arnold Zellner (ed.) Amsterdam: North Holland, 1980; and, in part, in *Good Thinking: the Foundations of Probability and its Applications*. Minneapolis: University of Minnesota Press; 1983.

Good, I. J. (1977b), "Summing up of the discussion on inductive inference", *Machine Intelligence 8,* E. W. Elcock and D. Michie (eds). Chichester: Ellis Horwood Ltd. & New York: Wiley, pp. 205–206.

Good, I. J. (1978), "Adenine arabinoside therapy", *Journal of Statistical Computational Simulation 6*: 314–315.

Good, I. J. (1980), "Some logic and history of hypothesis testing", in *Philosophical Foundations of Economics,* J. C. Pitt (ed.). Dordrecht: D. Reidel.

Good, I. J. (1981), "The philosophy of exploratory datum analysis", *American Statistical Association: 1980 Proceedings of the Business and Economic Statistics Section*. Washington, DC: American Statistical Association, pp. 1–7.

Hadamard, J. (1949), *The Psychology of Invention in the Mathematical Field*. Princeton: University Press. Reprinted by Dover Publications, New York, 1954.

Jaynes, E. T. (1957), "Information theory and statistical mechanics", *Physical Reviews 106*: 620–630.

Keynes, J. M. (1921), *A Treatise on Probability*. London: Macmillan.

Kruskal, W. H. (1982), "Criteria for judging statistical graphics", *Utilitas Mathematica 21B*: 283–310.

Leamer, E. E. (1978), *Specification Searches: Ad Hoc Inferences with Nonexperimental Data*. New York: Wiley.

Michalski, R. S. & Negri, P. (1977), "An experiment on inductive learning in chess end games", In *Machine Intelligence 8,* E. W. Elcock & D. Michie (eds.). Chichester: Ellis Horwood & New York: Wylie, pp. 175–192.

Nieto, M. M. (1972), *The Titius-Bode Law of Planetary Distances*. New York: Pergamon Press.

Ornstein, R. E. (1973), *The Nature of Human Consciousness*. San Francisco: Freeman.

Russell, Bertrand (1946), *History of Western Philosophy*. London: George Allen and Unwin.

Tukey, J. W. (1977), *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.

Wainer, H. & Francolini, C. M. (1980), "An empirical inquiry concerning human understanding of two-variable color maps", *American Statistician 34*: 81–93.

Whittaker, E. T. & Robinson, G. (1944), *The Calculus of Observations* (4th edn.). London: Blackie; New York: Dover reprint 1967.