Statistical Science
2000, Vol. 15, No. 2, 132–152

Statistical Thinking and Statistical Practice: Themes Gleaned from Professional Statisticians

Maxine Pfannkuch and Chris J. Wild

Abstract. Advancing computer technology is allowing us to downplay instruction in mechanical procedures and shift emphasis towards teaching the "art" of statistics. This paper is based upon interviews with six professional statisticians about statistical thinking and statistical practice. It presents themes emerging from their professional experience, emphasizing dimensions that were surprising to them and were not part of their statistical training. Emerging themes included components of statistical thinking, pointers to good statistical practices and the subtleties of interacting with the thinking of others, particularly coworkers and clients. The main purpose of the research is to uncover basic elements of applied statistical practice and statistical thinking for the use of teachers of statistics.

Key words and phrases: Consulting environment, psychology of data, statistical empirical enquiry, practitioners' experiences, characteristics of thinking, applied statistics.

1. INTRODUCTION

1.1 Background

Advancing technology is progressively shifting the balance of statistical learning from the operation of mechanical procedures towards the "art" of statistics. How can we teach that art? One response is to use practical experience through case studies and project work. But even to employ these methods effectively, we need to know how to choose good activities and to know what particular aspects of those experiences to draw students' attentions to. This will require applied statisticians becoming involved in characterizing how they think and solve problems—learning to articulate important aspects of statistical problem solving that "everyone knows" but seldom articulates—to make the implicit more explicit. Many teachers of statistics have much less practical experience than practitioners might like and cannot easily recognize and extract these commonalities for themselves. Such work is needed to underpin even the first course.

Maxine Pfannkuch is Senior Lecturer and Chris Wild is Professor, Department of Statistics, The University of Auckland, Private Bag 92019, Auckland, New Zealand (e-mail: wild@stat.auckland.ac.nz).

Because first courses reach such large numbers of students, they have the potential for enormous impact, but Moore (1998) suggests that these courses are missing an opportunity to equip students with the ability to reason with data. In the words of Mallows (1998, page 3), "We have almost no theory to help us understand how to think about applied statistics. As a first step,...we could look for communalities." We need to start developing such theory as a means of improving the way our discipline is taught. The present paper describes early research in this direction. We have interviewed working applied statisticians from diverse backgrounds and have encapsulated common themes in thinking and practice. Through identifying communalities, we hope to begin to understand how to incorporate this knowledge into teaching programs and how to develop such thinking in students. Using this research and other studies, we have worked toward developing a coherent framework for statistical thinking in empirical enquiry (Wild and Pfannkuch, 1999).

Our paper is based upon interviews with six professional statisticians. We deliberately chose individuals working in different application areas and different environments. Each statistician read and vetted the final paper. We have a great deal of consensus. None of the experiences and lessons reported by any one statistician rang false to any of the others. Although there was some structure to the interviews in the form of open-ended probes aimed at eliciting thoughts and experiences on the themes of the title, the course of each interview was largely shaped by the interviewee and the stories he or she wanted to tell. Those stories tended to be of experiences that had surprised them, things that had not been part of their standard training as a statistician.

The interviews were each of approximately ninety minutes duration and were conducted by the first author. The interviews were audio-taped and transcribed. [Among the more colorful transcription errors were: “box trots” (box plots), “designer shoes” (design issues), “evasive measurements,” “argument by antidote,” “effective interest” and “soul justification.”] The transcripts were analyzed independently by each of the authors. We wanted to build up a tree structure of categories suggested by the data. To do this we used NUD*IST (Richards and Richards, 1995), a software tool not yet widely used by statisticians, which enabled such analysis of qualitative data (see Appendix). We wish to emphasize that this is exploratory research aimed at uncovering higher-level thinking skills used by statisticians and uncovering characteristics of the environments in which they work. The discussion is complemented by references to the literature on statistical consulting including Tweedie (1998), Hoadley and Kettenring (1990) and Boen and Zahn (1982).

1.2 The Statisticians

Our statisticians included two females and four males ranging in age from their mid-twenties to late forties. Four had worked in more than one country. We will give them descriptive names based on their areas of application.

Biology. Biology was initially trained as a zoologist and subsequently became interested in statistics. A professor in the biological sciences teaching biometry to biology students, Biology has provided statistical consultancy for many years to researchers in the biological sciences and their students and has done a great deal of collaborative research as well as working on his own biological and statistical research. He regards himself primarily as a scientist.

Brain. Brain is a statistics professor who has done collaborative research and consulting for many years. However, most of his stories relate to his most important research project of recent years, working with a group of medical researchers on brain mapping.

General. Between two periods of employment as a statistics professor, General worked for ten years as a statistical scientist in a government research agency. He has extensive experience in collaborative research in agriculture, forensics and market research, but consults much more widely than this for both government agencies and private companies.

Market. Market began work as a statistician for a market-research company where she made such an impact that she was soon moved into more central marketing and management roles.

Medicine. Since completing her Ph.D. in statistics, Medicine has worked as a statistician assisting medical researchers in a hospital environment.

Quality. Quality began his career as a statistician assisting researchers in a medical school. About 15 years ago he resigned to become a private statistical consultant. He developed interests in quality improvement in organizations so that now his chief professional focus is quality management, although he still does some medical and pharmaceutical consulting and some university teaching.

1.3 Organization

Basic to our discussion is the concept of a “practical problem” (“real-world problem,” or a “whole problem”)—one in which both the source and the solution lie outside statistics, but where statistical investigations supply some of the understanding needed to arrive at a solution. From the context of this practical problem, one may extract one or more “statistical problems” aimed at reaching some particular learning goals. We often use the notion of “the system” in the rather vague sense of the environment in which the practical problem is grounded. It may be a biological system, the mechanisms giving rise to a disease, or an administrative system in an organization. The word “system” suggests the presence of processes, components that change or can be changed, and interrelationships. In contrast, the connotations of the word “population” are much more static. The word “transformation” is sometimes used in a narrow technical sense of a simple reexpression (e.g., a log transformation), and sometimes in a more general sense. We have tried to flag technical usage with “(reexpression)” where the distinction matters. Context knowledge refers to knowledge about the system, knowledge about the phenomenon being studied and knowledge about the subject matter from which the data are generated. Where possible, quotations that encapsulate viewpoints have been used in preference to paraphrasing.

Our material, for organizational purposes, is based around PPDAC (Problem, Plan, Data, Analysis, Conclusions), a generic depiction of the statistical empirical problem-solving cycle. However, a number of themes could not be categorized in this way, either because they were interwoven throughout the cycle, or because they were related more to the background in which PPDAC is embedded.

The statisticians were introduced and their backgrounds discussed in Section 1.2. Section 2 discusses the realities of the environments in which statisticians work and the constraints these impose on the ways in which they approach the problem. Particular attention is paid to interactions with clients, the effects of the client's expectations, knowledge and psychology and the way these influence statistical planning, analysis and reporting. Section 3 discusses elements of statistical thinking. The remainder of the paper, Section 4 to Section 6, is built around the empirical enquiry cycle. We conclude with a discussion in Section 7.

2. ENVIRONMENTAL REALITIES

2.1 Defining the Role of the Statistician

Our statisticians differ in terms of the types of people with whom they tend to interact. General, Biology, Medicine and Brain usually interact with researchers, people with some investigative sophistication and statistical intuition, if not formal statistical training. Market and Quality deal with commercial end-users, people much more like the general public in this regard. Our statisticians also differ in where they tend to be brought into the investigative cycle. Often they are consulted at the analysis stage of an investigation and sometimes at the design stage (although not often enough). In addition, Market, Quality and Biology often deal with "whole problems" and this is their preferred working mode. For Market and Quality, this means that they are very much involved in problem formulation (or isolation) and their end product is a recommendation for practical action rather than a statistical report providing background information only. Biology is intimately involved with "the science," as is General in some areas, particularly his forensic work (cf. Kirk, 1991, page 29; Hunter, 1981, page 72).

Market believes that taking a very broad view of the statistician's role is essential to be successful in her industry. People who do not do so "... really end up being marginalized, so they sit around doing t-tests and crunching away at computers, and no one really understands what they're doing and no one really wants to talk to them or value them and I think that's a terrible shame.... The image of a statistician out in the big commercial world is pretty negative. It's more a case of other people defining your job for you and maybe thinking that as a statistician you don't have anything relevant to contribute. You've got to be active in redefining it" (cf. Hoadley and Kettenring, 1990, page 246; Box, 1990, page 251; MacKay, 1990, page 263; Snee, 1990a, page 267).

All of our statisticians believe that whole-problem understanding is necessary to fulfil even the purely technical roles effectively (cf. Tweedie, 1998, page 2). All believe that they have skills that are useful at other stages of the investigative process. However, the work they are offered is determined by the perceptions of others. Nonstatisticians tend to see the statistician's specialties as analysis and a small subset of research design. Although Market and Quality think of themselves as statisticians by virtue of their training and early experience, many clients see Market as a market researcher and Quality as a quality-management professional—labels that one might expect to promise big picture expertise. Biology is a trained biologist. The three interviewees unequivocally operating under the "statistician" label generally work in more technical roles. Medicine has often experienced the very narrow "get me a p-value" view of statistics (with clients "obsessed with p-values by the cubic foot") in which the statistician is viewed as a technician and, in some cases, statistics is viewed as a tool for proving things one already knows. Quality also finds that "Clients see 'statistics' very narrowly—as something to do with number crunching rather than extracting meaning from data." (cf. Hoadley and Kettenring, 1990, page 245).

We conclude this paragraph with an astounding observation from Brain. For some of the people he has worked with, "There is a belief that statisticians have nothing to offer them because they [the brain researchers] have such strange data [three-dimensional images]." He later concluded that this had been an unfortunate side effect of these individuals' instruction in elementary statistics, where they developed the impression that statistics can only deal with a small class of simple and regular data structures (cf. Moore, 1998; Hoadley and Kettenring, 1990, page 244; Box, 1990, page 251; Moore, 1990a, page 266; Tribus, 1990, page 271).

2.2 Constraints

Almost without exception, our statisticians do their applied statistical work on problems owned by someone else. The relationship is most cut-and-dried in a commercial setting where the client is paying for specified services, but in almost all settings, our statisticians are in the position of having to satisfy a "client" (or client group). Major decisions are made by, or must be cleared with, the client. Market states, "What usually happens is that the client defines the territory." This may involve limits: "They have other territories that they consider they know already and they don't actually want you touching or thinking about.... If they think they know something, they don't want someone coming in and upsetting their whole idea about how things work."

In addition, our statisticians work under resource constraints on money, equipment and time. They seldom have the luxury of solving a problem to their complete satisfaction. It is virtually always a matter of doing the best they can with the limited resources available. General says, "In terms of actually coping with people coming into my office every day, I need a broad, shallow approach." Constraints on time, money and the availability of information affect sample sizes, study designs, and the adequacy of solutions. General continues, "But it's a toss-up between giving them what they want, which you know is really only 60 percent of the answer, or really shifting their way of thinking and giving them 80 percent of the answer.... Everybody is short of time and money and if 60 percent of the answer is good enough [then it could be considered to be a practically adequate solution]." Or the achievable solution may not even be practically adequate ("It's not even near enough; it's the best we can do") (cf. Hoeting, 1998, page 12; Hahn, 1990, page 258; Boen and Zahn, 1982, Chapter 2).

2.3 Gaining and Maintaining Client Trust

Apart from purely technical areas, the statistician operates as an adviser who can only influence events by persuading the client of the desirability of a course of action. Even in technical areas, the utility of authoritative pronouncements ("This is my area of expertise and I know best") is severely limited. Statisticians' qualifications might get them a job, or a place on a project, but do not automatically buy them credibility. Brain elucidates, "You know they couldn't just take the word of some mathematician who comes into their lab and tells them that they are doing it wrong." Our statisticians emphasized the need for the statistician to gradually build up the clients' trust in their judgment (cf. Tweedie, 1998, page 1; Kirk, 1991, page 28).

An important consideration in "building trust" is not taking clients too far from territory in which they feel secure. Indeed, the building of trust can be thought of as enlarging the client's security zone. General stressed "the usual practice...what has been done in the field before..." as an important element in client security, be it for measurement decisions, study design, analysis or even the presentation of results. We will refer to this throughout as the *first-in-the-field effect*, an effect which gives the work of the first people working in any field undue influence. The statistician must try to recognize the client's level of statistical knowledge and work from that level. Working cooperatively alongside the client, communicating effectively about what is being done as the project proceeds and developing the client's statistical knowledge over a prolonged period of time, all help build the client's trust and reliance on the statistician's judgment and prepare the ground for the use of more sophisticated designs and analyses in future work. The statistician must supply solutions that address the client's objectives and are understandable to the client (cf. Kirk, 1991, page 32; Hooper, 1990, page 260; Hunter, 1990, page 261; Snee, 1990a, page 267; Wangen, 1990, page 273).

The statistician also needs to become aware of the consequences of the research for the client, so that a clear recognition of the underlying priorities can help guide the statistician's decision making. Brain spoke of the desirability of being on the spot (e.g., in the laboratory where the research is taking place) as a way of becoming fully integrated into a research team and thus being consulted at all stages. The importance of having materials prepared for publication be careful, accurate and justifiable was remarked upon several times (Market: "A lot of the research that I've done for government is very, very politicized, so it matters that you get it right.") Problems due to differences between members of a client group were also raised by Market: "If you have lots of different stake holders, then the territory is different and often you can't please all of them. And sometimes our clients use researchers almost as a way of resolving disputes.... Whatever we say is going to make one of them unhappy and then, you know, you get a lot of flack from them." Brain talked about the extraordinary importance of reputation to scientists working in expensive areas such as brain imaging. It affects their ability to obtain the large sums of grant money needed to run their experiments and maintain the livelihoods of their laboratory staff. As a consequence, he learned to be more careful in his criticism of methods at public gatherings.

2.4 Managing the Client's Expectations

Both General and Market talked at length about managing client expectations, about getting clients to accept the limitations of their data sources, to accept that the outcome might not meet their expectations, or that the problem may simply not be solvable within the practical constraints. General elaborates, "It's very hard to give a client negative results. It's not that they are biased, it's just that they have a natural expectation and the hope that their product and their idea is better than in fact it has turned out to be . . . . Most people are disappointed by their experiments." For one set of clients, "It's taken them a long time to get to the realization that their data process just can't generate the data in that form and that they will have to settle for something simpler and to get them to accept it." General finds that clients accept his findings when the results are disappointing, but perhaps only ostensibly. "Basically every client says 'yes' ultimately. They might not ever come back to you again." The litmus test is a client's returning with further projects.

2.5 Ethical Problems

Fulfilling the priorities of clients can lead statisticians into situations where ethical warning bells begin to ring. In Medicine's environment, it is not uncommon for doctors trying to maintain a research profile to submit conference abstracts saying that they will present new data on some question of interest before they have actually done the research. Medicine is then often in the position of working under intense time pressure for clients who were looking at previously collected clinical data hoping to find something interesting in it and being desperate to find "significant results." Market tells of the conflict between telling clients that they would just be wasting their money doing the desired research (e.g., because she expects excessive nonresponse) and realizing that doing so would mean "You've just talked yourself out of $200,000."

3. ELEMENTS OF STATISTICAL THINKING

Here, we approach some general ideas about the elements of statistical thinking, the basis of which is to produce an improved context-matter understanding (cf. Box, 1990, page 251). We see four elements as foundations of statistical thinking. The first element is the taking account of variation. The second is "transnumeration," a fundamentally statistical process that we will expand on later. These two elements are what makes something inherently statistical in the modern sense. Third is constructing and reasoning from models, with statistics having its own distinctive class of models. Fourth is the integration or synthesis of problem context-matter and statistical understandings. Sufficient statistical knowledge and sufficient context knowledge must underpin these four elements to allow such thinking to take place. Some supporting factors or personal attributes (e.g., imagination, logic, scepticism, curiosity), which we have categorized as dispositions, are discussed.

3.1 Foundations

Variation. It is General who best expresses this idea. "Basically what distinguishes statistical thinking from anything else is that you accept that variation exists." To Quality, variation is ubiquitous ("Statistics is the science of variation,") but more than that, it is informative and the key to improvement ("And variation, of course, contains all the information about what's going on") (cf. Moore, 1992, page 426; Snee, 1990b, page 116; Provost and Norman, 1990, page 43; Moore, 1990b, page 135). The basis of the way Quality approaches the world consists of noticing variation in the output of processes, wondering about causes of that variation, investigating possible causes and using suitable identified causes to change the pattern of variation. To Biology, the fact of variation in biological systems is inescapable. His world view is oriented towards isolating sources of variability in order to understand or explain the reasons for physical and behavioral differences. Talking about the implications of variation, General states, "It's the recognition that, because things are not going to be the same next time, there is no one answer; that everything is a summary or everything is a model." Do his clients accept that variation exists? "Well, it varies from client to client. Anybody dealing with agriculture, or horticulture, or biology, or market research or medicine knows very well that variation exists. Industry and commerce are less tolerant." Medicine qualifies this, at least for her area. Although there is general awareness of variation at some levels, (e.g., patient to patient), there is much less consciousness of variation at other levels (e.g., variation in repeat measurements on the same individual, or measurer-to-measurer variation) (cf. Hawkins, 1996, page 2; Armstrong, 1990, page 249).

From his long involvement with quality improvement, Quality has developed the mental habit of noticing variation and wondering "Why?" to the extent that it carries over into daily life. Examples he related included varying amounts of water in a dehumidifier from morning to morning, moss growing on some parts of the pavement but not on others, and a cup of coffee at a restaurant sometimes arriving partially spilled into the saucer. Biology states, "What most of the stuff I do boils down to, if you look at it, is 'Why are these animals not all the same?' " We hazard that this very basic element of statistical thinking, "noticing variation and wondering why," is actually at the root of much, if not most, scientific research. Many of Quality's problems arise from pressing problems affecting a company, for example, customer complaints. "Why were the problems that are pressing this week not sufficiently pressing to do anything about last week?" Perhaps this also comes down to noticing undesirable changes (variation) and wondering "Why?" with a view to reversing them. Of course, "wondering why" is just the start. To quote Quality again, statistical thinking is also "... about knowing that the only way we get any information about the world is by taking samples of data in one form or another .... It's about saying how would I find out about that?"

Transnumeration. Fundamental to a statistical approach to understanding the world is the forming and transforming of data representations of aspects of a system to arrive at a better understanding of that system. We have given the name transnumeration to this idea. Our definition is "a numeracy transformation for facilitating understanding." It is not mere translation, in the sense of substituting one thing for another. It is informed by contextual and statistical knowledge and it is driven by the desire for a better understanding. It occurs in three phases: when there is a quantitative description of the real system, when data are transformed in the statistical system and when data representations are formed that help communicate to others what the data is saying about the real system. We now consider these phases in more detail.

A statistician will look at a system from the perspective of capturing data from it. Quality, when commenting on the spilt coffee, moss and dehumidifier, stated "All those things [phenomena and contributory factors] are measurable in some sense. So you start thinking 'How could you capture some of those things?' " This theme is echoed by Biology as "What characteristics of the system must I measure in order to try and answer that question?" At a more basic level, transnumeration occurs when data on attitudes, for example, are captured by forming ordinal categories. Market: "There were a lot of questions relating to their attitudes towards recycling ... rating on a scale from 1 to 10." The way of thinking is focussed on obtaining data (through measurement or classification) that captures meaningful elements of the real system.

Transnumeration also occurs every time we find a new way of looking at the data that conveys new meaning to us. We may look through many graphical representations to find several really informative ones. Quality spoke about how changing a representation could deliver new information about a system. Data may be transformed via reexpressions, aggregation or stratification and reclassifications in a search for new insights. In reference to his students, Quality remarked "I try to get them to do lots of stratification and to look for differences ... and to aggregate where it makes sense." In Biology's words, "The data may suggest all sorts of things if you just look at it right." The data must be transformed or looked at in a new way to reveal new features. Market commented on how this type of thinking allowed an unsuspected insight into the data: "It took us to look at the data, look at the numbers and to think about them .... We looked at what *sorts* of things were being recycled in the different suburbs and the way they were being measured." Another form of transnumeration occurs when a variety of statistical models is used in order to find salient and relevant features in the data (cf. Hawkins, 1997, page 144).

At the end of the process, transnumeration happens yet again when we discover data representations that help convey our new understandings about the real system to others. Brain related how the neuropsychologists communicated statistical summaries to one another: "They've developed an awful lot of tools for visualization by themselves that they can directly relate to and they superimpose the statistical images [e.g., color coded p-values] on to these background images [anatomy of the brain] so they can interpret what's going on." He then went on to say he had used the technique of color coding for a residual plot so that the information could be conveyed instantly to his audience (cf. Hoadley and Kettenring, 1990, page 247).

Transnumeration is a dynamic process of changing representations to engender understanding. The new field of data mining or knowledge discovery in databases (MacKinnon and Glick, 1999) is closely related to this transnumeration type of thinking.

Building and reasoning from models. Whether formally or informally, statisticians do much of their reasoning from data in terms of mathematical constructs called statistical models (cf. Tweedie and Hall, 1998, page 19). Less consciously, a host of other less formal mental models act as precursors and in support of the "statistical models." For example, the statisticians formed visual models of the interconnections of the system they were studying. Talk of models pervades Biology's transcript, accompanied by repeated references to "checking the mapping," that is, ensuring that the abstractions and simplifications used in forming models retain the essential elements of biological meaning: "I tend to go with the models only insofar as the model is a relevant abstraction (and a useful abstraction) of the reality, so parts of the model for which you are going to use description actually do relate to recognizable features of the reality." Moreover, in statistics as in any modelling process, "You are making abstractions; you are losing information. Your sole justification is the assertion (that others must believe) that losing that information doesn't invalidate your results."

Biology has a scientist's view of statistics as the fitting of models, formal analyses and the "measuring of evidence." "The only reason you are doing statistics is because you can't go directly. You know there are a lot of scientists who say that a good experiment doesn't need statistics and they're absolutely right. It's a first-rate one because the inference is direct. Nobody in their right mind wants to use an indirect inference process, it's obviously weaker than direct so, yes, statistics is second best. It's a good second best, but it's second best." Brain said that statistics played an important role in determining which parts of the brain were activated because the signal was very weak compared to the noise. If a better scanner were available that "... cut the noise by a factor of ten ... then you would actually see what the activation was without having to apply statistics at all." But he then mused that there would always be a need for statistics because the cognitive experiments on the brain would become more subtle if the noise were reduced. However, statistical thinking is more fundamental than statistics itself according to Biology: "Statistical thinking makes us aware of the difficulty of separating effects and provides us a way of thinking about how effects could be separated. Incidentally, statistics also provides us with a tool to do it, but the logic underlying the statistical methods is far more important and statistical thinking can be applied even if you don't do a significance test. If you design a good experiment, keeping in mind there's no statistical technique involved in the design of a simple (relatively simple) experiment, all the decisions can be made on common sense; the statistical thinking is in the awareness of sources of variability ... the separation; it's all pure logic, common sense, except it's not very common. The conclusions in a good experiment probably won't need statistical significance testing anyway. The confidence intervals and the standard errors, they're just refinements letting people know how reliable the results are, so they're descriptions of level of evidence to some extent, but very often the primary conclusion requires no statistics at all. Yet statistical thinking went into the whole process (cf. Trosset, 1998, page 23).

Context knowledge, statistical knowledge and synthesis. Statistical investigation is carried out because people deem their context knowledge insufficient for their desired uses, be it as a basis for decision making and action or simply for understanding. The statistician uses both context knowledge and statistical knowledge to formulate a plan, to collect data, to fill essential gaps in context-matter understanding and to extract information from this data. Then he or she draws on both context-matter and statistical knowledge in order to synthesize the new information with current context matter understanding to obtain improved context-matter understanding. This is the use of statistical thinking in which the authors are interested, the key element being the integration of context-matter and statistical perceptions. As Quality said, "Good statistics is not so much analytic as synthetic. The synthesis of information from (usually) a wide variety of sources to tell us something about the wider system is a key element of statistical thinking."

To Biology, R. A. Fisher was the quintessential statistical thinker—the science and the statistics were perfectly integrated within a single individual with the scientific goals being paramount and the statistics existing to serve them. "What I would like to see is every scientist a statistician." [General annotated this with "... and every statistician a scientist." (cf. Hoadley and Kettenring, 1990, page 247).] Biology's view of Fisher may surprise some statisticians: "Fisher was first and foremost a biologist as far as I'm concerned. His major input to the world was in genetics. He was also a genius in statistics and he sat on that interface and could see both ways. As a result, he revolutionized statistics because he brought a scientist's perspective to it." Although this may be the ideal, Biology recognizes that in the absence of a Fisher-like colossus who can bestride two fields, the perfect integration of context and statistics within an individual must be approximated using communication between individuals.

Biology thinks in terms of the statistical system modelling the real system and states that the mapping between these two systems must go both ways ("It's this integration that's often actually ignored by people.") From his stance as a biologist he believes that "Statistical knowledge can't be used adequately unless the [statistical] knowledge is actually an integrated part of your context knowledge." Market deals with an initial lack of context knowledge by spending time "knowing what customer's problems are." Furthermore, she questions whether a statistician "... can actually do a good job if they're only involved in pieces of something." All of the statisticians built up some context knowledge of the situation before they started the design or analysis. It was regarded as essential to integrate this context knowledge with their statistical knowledge in the carrying out of the investigation. During the process of the investigation there was usually close collaboration with their clients to further enhance this integration (cf. Tweedie, 1998, page 2; Cobb and Moore, 1997, page 801; Snee, 1990a, page 268).

3.2 Dispositions

Imagination. Biology sees imagination as one of "... the roots of statistical thinking," particularly in the ability "... to imagine what is happening in the system," "... the ability to come up with the alternative explanations for a phenomenon and confounding variables," and "... to identify factors that may be of importance within the analysis." Biology's primary aim is to reach a position in which we have an explanation for a phenomenon "... and there are no other plausible alternative explanations at the current level of understanding." Much of the work in arriving at such a position consists of generating explanations and devising (by means of experimentation, observation and analysis) ways of narrowing the field by ruling candidate explanations out of contention. "If you lack the imagination either to see possible confounding explanations for the results or dynamics that will affect what technique you ought to use ... then you're not going to come up with appropriate things. So imagination's at the foot of the ramp."

Scepticism and critical thinking. The statisticians critically assess and appraise both their own work and that of others. Quality stated, "So I think that critical thinking is a really important part of statistical thinking." General continues, "It goes hand in hand with scepticism ... you don't have particular preconceptions." Biology: "Every step requires intelligence and assessment .... I like disagreeing with people. As soon as they come up with an explanation I immediately try and see an alternative explanation." Quality encourages a general scepticism: "Yeah. I try to teach them to be fairly cynical about that sort of thing." [Believe] "... half of what you see, a third of what you read and a quarter of what you hear, or something like that."

Other factors. A number of other factors were suggested with little amplification, for example, *logic, commonsense* and *a sense of number*. Both General and Brain refer to *an openness* and the need for *curiosity* (cf. Quality's "wondering why"). To this, Brain adds the need *to get involved*. Market goes further, "I feel in my job that you're motivated to do a good job by knowing who you're doing the job for ... and how much those problems actually mean to them." Biology states, "I'm bloody minded and I worry at a problem" (implying *persistence*). He emphasises taking the time to think about a problem rather than rushing into busy work, "Apologies to Marx, but work is the opiate of the thinking classes. Most people who ought to be spending time thinking about what they're doing would much rather be out there digging up the animals." The reference is to biologists, but the statistical parallels are obvious (cf. Hunter, 1981, page 75).

4. PROBLEM

4.1 Introduction

From this point on, we use the statistical cycle PPDAC (Problem, Plan, Data, Analysis, Conclusions) as our organizational framework (cf. MacKay and Oldford, 1994). PPDAC is applicable to any inquiry cycle. After all, as Quality points out, "Most of the data we get isn't in numerical form. It's impressions, fleeting thoughts and so on." However, our concern is with PPDAC cycles that are recognizably statistical because they employ a gathering of data and an analysis that is recognizably statistical (see Chatfield, 1991, for discussion on guidelines for avoiding pitfalls in a statistical investigation). We begin at the beginning, with the problem.

We need to distinguish between the problem that initiates PPDAC and the problem presented to the statistician. The latter problem may just be a specialized subproblem of the former, most commonly, a part of the analysis. But even then, the statistician must go back to the beginning and form an understanding of the Problem-Plan-Data steps. This is an unavoidable prerequisite for any proper analysis of the data.

There is typically a great deal of work required to get from initial inklings about the nature of a problem to a set of questions that can feasibly be answered by collecting statistical data. Just how much of this work the statistician participates in depends on when she or he enters the cycle and how vague or specific the problem is. At one end of the spectrum, General (recounting his experiences with commercial clients) talks of problems of the form, "Oh my God, we've got this problem. What can we find out about it?" Quality also often experiences such situations, situations in which the existence of a problem is clear because the undesirable effects are obvious (e.g., substandard goods or large numbers of customer complaints), but there are only vague ideas about what is causing the problem. Getting from the existence of a problem to questions to be answered by gathering data constitutes a very substantial part of the work. Market also singles out her commercial clients as tending to come in without having really thought through what exactly it is that they need to know. She notes how the process of actually sitting down and trying to write survey questions helps clients understand their problems more clearly. The process requires a great deal of client-researcher interaction and sometimes she has the frustration of several false starts (cf. Tweedie, 1998, page 2).

In the middle part of the spectrum are clients with quite specific problems, though still not expressed in technical statistical terms. Market's government and social research clients are of this kind. General finds, "Often people have a specific question or a specific problem but they don't know how to formulate it, or they don't even know how to start gathering data or what sort of data to gather. Or they may have data which is gathered for other purposes and they don't know how to apply it. But they always know what they want to investigate and they often think they know what the answer is." The researcher-clients come for specialized technical assistance because of gaps in their expertise regarding parts of the design or analysis. Brain's work for brain research is at the other end of the spectrum and involves very specific analysis problems.

4.2 Grasping the Dynamics of the System

Biology: "The first step is not to find the problem, the first step is to find the context of the problem" (cf. Hahn, 1990, page 258). The statisticians begin by forming some understanding of the dynamics of the system being studied. The most common source of information about the system used by our statisticians is other people. They ask questions, they "interrogate." Interrogation is an imperfect means of extracting information. Biology laments, "They never know what they ought to tell me and they never tell me the things I need to know. As a result I always say that the first requirement for a good consultant is telepathy!" Medicine echoes the frustrations of interrogation: "I find myself often having to be quite firm with people . . . and sometimes they find it hard that I don't understand what they're trying to get at on the clinical side, but to me it's that important to be able to get a handle on the problem." (cf. Trosset, 1998, page 23; MacKay, 1990, page 263; Hunter, 1981, page 73). Generally, the client is interrogated, but Quality and Market sometimes dig deeper to people working within the system (e.g., shop-floor workers), people experiencing products of the system or having some other special knowledge. These people's perceptions about the problem can be quite different from those presented by the client. Quality advises his students, "It doesn't matter if the staff can barely speak English, if it looks as though they can't add 1 and 1. Study their perspectives; they are going to be informative." Industry projects force Quality's students to experience this process as well. Market tries to follow ". . . certain industries to try and work out what trends are occurring and how that might later on impact on some research you might do for them." The more specialized the context-field that the statistician works in, the more feasible it becomes to develop into a context-matter expert (cf. Broman, Speed and Tigges, 1998, page 8).

How do they "grasp the dynamics of the problem"? Quality has a range of tools based around process models for analyzing systems. These tools, and their use, are commonly taught in the quality area. For the rest, it just seems to be an intuitive combining of information and imagination to form a mental picture or model. Biology recounts, "I'm trying to find out from the person enough information so that I can construct in my mind a model of the dynamic system . . . and I keep on sort of going round and round again. 'Well what about this? What about this?' 'I'm sorry, I don't understand that' and it can be extremely irritating [to the client]." Unable to interrogate animals in the system, Biology inserts himself into the picture and imagines the dynamics. Referring to a study of crabs, he says, "I think 'crab' and 'What affects me as a crab?' "

4.3 Defining the Problem

The problems presented to statisticians may be at the planning or analysis stage. However, in order to carry out the work, they must form some understanding of the system from which the data are to be created or have been generated. Once the system is sufficiently understood, the next stage is to clarify and define their clients' problems. As Quality states about his students and clients: "Invariably the problem they define initially is this wide. They get a lot of feedback from me . . . to help them narrow it down to something that is attackable and achievable. That is always a difficult step." Biology also mentioned that his students spent a long time establishing the question to be addressed and whether ". . . in the context of their system it was actually relevant or valid or useful." In this definition phase there is a constant seeking of alternative explanations or hypotheses for the phenomenon under study with decisions being made as to what aspects should be concentrated on and what the actual question is. There are also constant mappings to the statistical system to test whether it can answer such a question and to the real system to test whether the question has validity.

We conclude this subsection with an example of good practice from Quality. The best of his students doing projects in industry are "... good at talking to people and finding out what really matters as a means of cutting through the stuff which is irrelevant. And saying, 'Well, it seems that these are the key measures which we should be concentrating on.' "

4.4 Factors Affecting Perceptions of the Problem

A number of factors, including level of context knowledge, affect conceptions of a problem. Statisticians get much of their context information from clients. Along with valid information and perceptions, statisticians can tend to take on the mistaken preconceptions of their clients. For example, Market told about investigating a city's recycling scheme, where everyone involved proceeded from an unexamined community assumption that all neighborhoods recycled similar items (see Section 5.1 for further detail).

It is not only context-matter knowledge that affects how we conceive real-world problems. Statistical concepts also play a part. Thus, the distinction between special-cause and common-cause variation plays a very important part in the way that Quality thinks about problems. This distinction addresses how to decide between two basic strategies for finding causes of a problem: whether to look for "unusual happenings" as potential causes, or to study relationships between variables that capture important aspects of the system. "The Pareto Principle is one of the powerful principles. You know that most of the variation is caused by a very small number, of causes." General perceives problems from the stance of, "You've got to be able to abstract from what the person's saying ... something, some sort of problem framework. And then rapidly connect it to something that you've done before, that you can recognize from before." Statistical knowledge and statistical experience give us much more than just tools for analysis. They enrich the body of mental models at our disposal for conceiving the very nature of a problem.

5. FROM PLAN TO DATA

For much of the company and institutional data he sees, Quality says, "the reality is the data is mostly useless and you've got to start again. Depressing message, really." Similar experiences led almost all of our statisticians to tell disaster stories and to stress the importance of careful planning for investigative success. In this section, we distinguish between units (entities), characteristics (properties) of those units and variables (attempts to capture properties with some form of "measurement"). Section 5.1 discusses how we get from characteristics to variables, Section 5.2 discusses the psychology of data and Sections 5.3 to 5.5 deal with data production issues.

5.1 Measurement

Biology introduces this topic: "Once you've grasped what the system you're studying is and what the question is that you wish to ask about this system, then you can ask, 'What characteristics of the system must I measure in order to try and answer that question?' " A useful distinction in thinking about measurement is the distinction between characteristic and variable. The "characteristic" is an intuitively held idea about a property of the system or unit under consideration, whereas a variable is an attempt to measure the characteristic. Biology elaborates, "I can intuitively define it [a characteristic], size, shape ... intelligence is a very good one ... and these are the ones I give my students because, intuitively, we know what they mean. Actually, they are appallingly difficult to measure and people can come very unstuck trying to do it." Quality continues, "The things we are interested in are always fuzzy things. Like if we are talking about a car door [unit] for example, it should be easy to shut [characteristic], a soft idea. But it should stay open on a hill not swing back, banging and shutting. And those ideas we then try to get substitute quality characteristics for them. Which are things like the stiffness of the spring [variable] measured in whatever units they measure the stiffness of springs in .... You know the customer wants prompt service. What's prompt service? It sort of depends on what type of operation it is. Like XX fast food company. They have defined prompt service as being within two minutes at this time of day, within three and a half minutes at this time of day .... Or the XXXX Hospital has defined a response to phone calls within 40 rings as being prompt response—which doesn't correspond to what we think of as a prompt response!" In the end, the measurements used must relate well to characteristics that are important to the customer if they are to be useful. Otherwise they could actually mislead the investigators. In the words of Biology, it all comes down to the quality of "the mapping" between characteristic and variable. And as Quality explains, "the softer the notion the harder the measurement" (in the sense of it being harder to come up with variables that map to the characteristic well). The mappings are seldom perfect. Quality refers to "... that thing John Tukey said, 'The more you know about what's wrong with the figure the more useful it becomes.' " Ultimately, the question remains as to whether, with the variables one can construct, "... it is going to be even remotely possible to answer the question."

Market's investigation into the recycling habits of people in different suburbs, mentioned earlier, provides a cautionary tale about measurement. The amount of recycling done was measured by weight. When suburbs were compared, the results conformed with prior expectations. Very late in the process, however, the researchers came to the shocking realization that people in different suburbs had different consumption patterns. Those in higher socioeconomic suburbs tended to use heavy items like wine bottles more often, and those in lower socioeconomic suburbs used more plastic drink bottles and cans. Comparing the suburbs' propensity to recycle by comparing weights of recycled materials was thus misleading.

Medicine and Biology pointed to the arbitrariness of much classification. Medicine used the subjective classification of patterns in ECG images by cardiologists as an example. She had different cardiologists applying the same classification system to a set of images. Their classifications were often quite different. Another subjective rating system she described was self-rating of pain levels by patients on a scale of 1 to 10. "Now that's not a good measurement to [use to] compare with other people, but if they [patients] can on subsequent occasions see where their pain level is, then they [cardiologists] can relate one person's pain level to the same person's pain level at a different time if there has been some intervention or something." Related issues Medicine raised included intermeasurer variation even in much less subjective measurement scales, time-to-time variation in measurements such as blood pressures, and study-center to study-center variation. These issues have implications for designing a data collection process (see Section 5.4) and in analysis. Medicine also talked about measurements that are invasive, in the sense of involving pain, discomfort or excessive inconvenience to the patient. Determining whether invasive measurements were really necessary or whether simpler alternatives could be used has been an important part of some studies in which she has been involved. "So you're almost ranking the variables, not in importance as such, but in ease of taking—measuring just to find out if it is worth actually going to the effort of taking an invasive or interventional measurement."

We return to the fact that measurement decisions affect data analysis. An easy "measurement" to take may be a hard one to analyze. General elaborates, "If you're looking at the number of fungal spots on a leaf, it's an impossible burden to actually count the number and so people say, 'Oh well, if there are none we'll score it as a zero, and if there's one or two, it'll be a one, two to ten it will be a three . . . .' And so you get a simple score like that and then what do you do with it? That's the problem."

5.2 The Psychology of Data

Much of this section deals with the ways in which people (particularly clients) often relate to numbers and statistical data. These have important implications for planning, analysis and interpretation.

The first-in-the-field effect. A strong factor in the way that data are collected in a given situation, or the way a client will want the statistician to approach a problem is (quoting General), "... what's been done in this field before. The usual practice. 'We've always done it this way,' or 'The last person who looked at this kind of thing did it this way' ... which may be completely erroneous or may be completely inapplicable." It is not uncommon for a client to come with both a problem and also with a published paper and say, "We want to do something like this." Brain cautions, "Sometimes papers get published that do this kind of thing and they get to appear in prestigious journals and they take on a sort of an authority all their own. And people will repeat it and they'll say 'Well this was published in *Science* so it must be correct and now you are telling me it's not correct?' You know it's difficult to argue against that once it's been published in a prestigious journal. But you do get serious errors." This idea of early work taking on an authority of its own whether or not it is warranted, we like to call the *first-in-the-field effect*. It influences every decision in the process, right through to presentation. General continues, "Whoever drew the first graph in that area or that field, all the graphs then tend to look like the first one. If it was a good one or a bad one they all look like the first one." Boen and Zahn (1982, pages 45–47) discuss problems stemming from the use of new techniques.

Measure everything in sight. General developed this theme in detail with particular reference to explanatory variables. “Most people measure everything in sight. ‘We’re only going to come this way once.’ I mean I’m only ever going to have this sack of potatoes once. If I don’t measure it now, I will never measure it.” Medicine echoes this point. “When you are not sure what is important, it makes sense to err on the side of taking too much.” However, General continues, “. . . then you get into the problem that they then believe that absolutely everything is valuable.” There is a consequent tendency to want to use all the data they have, even though that may not be appropriate. General continues, “If all of the variables are telling you the same story, you might just as well measure one,” but, “. . . in a sense the customer is always right, so if the customer wants 20 variables analyzed . . . .”

Measured variables are sacred. General again: “They have difficulty in translating the variables into other units or working with transformed [reexpressed] variables. Somehow if you’ve measured something it’s got a sanctity even though it’s maybe causing lots of downstream problems to the analysis.” People are “. . . resistant to looking at it in a very different way.” “Because we measured it on this scale we have to stick with it on this scale.” Why? “Well again, you see it’s ownership and the people have invested the time.” Perhaps “ownership” is not quite the right word. It seems plausible that the physical act of measuring something in a particular way confers on it more of a concrete reality. It is natural to try to think in terms of quantities you feel you understand.

General told of having to forgo presentation of results for a 2-factor experiment with a nice additive structure in the factors and go back to cell means because it was these means that the client understood, being most directly related to the measurements the client had taken (“They want the whole picture.”). Medicine touches on the variability of a single measurement and the fact that it is often overlooked. “It must be right because that’s what the instrument said. So you say to them, ‘Well, if you did it ten times you’d probably tend to get ten different answers.’ And then if you said that, ‘Oh! of course. Yes, that makes sense.’ ” This was not something the medical researcher would have taken into account. The idea had to be planted.

Brain has a further story illustrating an almost emotional attachment to data. The data involved a measurement replicated twice under two sets of conditions. The suggested analysis involved taking the difference of the within-pair averages to form a single variable, thus in a sense reducing the “number of observations” by a factor of 4. In Brain’s work every observation is enormously expensive, and thus, enormously precious. The researchers “. . . were sort of worried that I had, you know, sort of stolen . . .” some of their data, and “It took me a long time to convince them that what I had done was correct.”

Transformations. Biology has a scientist’s suspicion of the statistician’s readiness to reach for a transformation (in the technical sense of reexpression). He talks of “. . . an awful lot of statisticians who, for example, automatically assume that you should transform a variable. It changes the model. If it changes the model, it changes what the parameters are mapping to and may not be a good idea, but you better check because there is nothing worse than transforming for some statistical reason and the model now no longer having any relevance because there’s no mapping back to the reality.” This is not referring to a difficulty in interpreting nonlinear scales, but rather an emphasis on subject-matter understanding being in the driving seat. “If your biological dynamic is a multiplicative one, there is no point in doing an additive analysis even if you do have a homogeneous error structure. You’ve got to do it in a log scale whether you like it or not. Otherwise, for example, interaction terms don’t mean diddly . . . . The mapping is frightfully important.”

General is careful to “map back” and highlights some of the subtleties. “Even though life may be a lot easier if you take logs, because people don’t understand logs, there is a resistance to analyzing data in log terms. . . . But you can help them along the way. You can plot stuff on a log scale and you actually put 10 instead of 2.23026. The fact that the tick marks are unequal distances apart makes them less upset than having funny numbers on the scale . . . . Routinely, people gather data in terms of proportions and percentages. They can accept when you show them that you can’t average percentages when the percentages are varying over a wide scale [because of grossly different variances]. So you transform the percentages, analyze them on a transformed scale and then back transform them—the transformed scale has got no intuitive sense. Even though they’ve accepted why you’ve done that step and that you’re only doing this back transformation to make them understand the results, they still say, ‘Oh, why can’t we take a simple mean of the percentages?’ or else they take a simple mean for themselves and get a different answer from you. Why? The scale on which the measurements have been made has got a primacy.”

Attitudes to data among less technical clients. Market talked at length about the complex, and even contradictory, ways in which the general public relate to statistical data: a mix of distrust, a fragile trust, and a desire to have something solid to hold on to. Market has reported the results of both qualitative and quantitative research to clients, often the same clients. She has been amazed at the difference in the reactions. "I can stand up and say, 'I talked to five people and this is the recommendation I drew. You should launch this ad and it will be ten times more successful than that ad based on these ten respondents.' And they'll actually not challenge me at all ... [The qualitative results are] basically my opinion after I've talked to a couple of people .... In a space of an hour and a half they say a lot of things. I might have asked them a lot of different questions in different ways and I can be selective about the quote I pick to put forward in my report to illustrate a point." [We *do* appreciate the irony. In our defence, this paper has been vetted by the interviewees!] But they seem to have more faith and more trust in something like that, than if I go out and talk to a thousand people, analyze the data and give them a number and put a margin of error around it .... I find when I present numbers, my clients sit in the room interrupting, 'But did you do this? Did you think about it that way? ... [some types of customer] really want to attack you and pick at the numbers." The difference in reactions "... absolutely astounds me ... it's led me to believe that people just have a deep mistrust of numbers" (cf. Boen and Zahn, 1982, pages 60–64; Roberts, 1978, pages 48–49).

Why the distrust? Basically it stems from the common perception that you can prove anything with statistics. On the other hand, however, "The reason why numbers are used to persuade is because people in some sense really do believe in the power of numbers. You know if you just say, 60 percent of this did that, someone actually thinks now that's meaningful. But then on the other hand, they know that lots of numbers are meaningless and could be made to say different things, so they don't trust them. A funny balance between trusting in numbers, but not quite trusting in numbers." Another place where Market sees the fragility of the trust in numerical data is in her client's reaction to data cleaning. "I'm all for data cleaning, but I think people don't like that because they want numbers to be immutable objects that we can rely on and trust, and they secretly worry that they're not. So as soon as you start playing around with them, they get very, very uncomfortable because it goes against their idea of what a number should be."

The statisticians mentioned other facets of the way nontechnical people tend to approach data. Quality explained that people have "... this tendency to examine the extremes [and] will say, 'What went so right this time? Why can't we do that all the time?' Well you can't because what you are doing then is what you are doing all the time. It's just a part of the natural variation." General attributed the propensity of clients to focus on detailed aspects of the data to the fact that it is "... difficult to assimilate big pictures. If there's a lot of data, there's a lot of conclusions. People will naturally hone in on some particular aspect that they're interested in." When the conclusions were not what they expected they will query the findings and, according to General, "... may say 'Oh, no, that's not right. I know so and so happens.' And then you may have to try and show them that their belief is not supported by the data." (cf. Snee, 1990a, page 269).

5.3 Design and Anticipation

Although most of the clients come to our statisticians for their expertise in design and analysis, our interviews did not probe technical aspects of these areas. Sample-size problems were broached by Biology, who continued his discussion of measurement ("whether you can get relevant information") with, "Whether you can get enough [relevant information] is a separate question again." But when it came to formal sample-size calculations and power calculations, Biology was scathing. "That whole business of estimating sample size is a bad joke." His objection is based upon the unreliability of the estimates of variance that are required by such calculations. Variances are notoriously hard to estimate and the idea that one can get a reasonable estimate from a pilot study he finds laughable. The resulting estimates, he told us in an earlier conversation, "... are not so much in the right 'ballpark' as in the right 'National Park' " (e.g., Yellowstone). Conservatism points to the use of the lower confidence limit of a variance estimate. However, "There is only one thing that usually determines sample size and that is how many you can afford." He later softened this a little to "Sample size tends to be on the basis of logistical constraints, intuition and whatever other people have found" (cf. Boen and Zahn, 1982, pages 119–122).

Anticipating problems and finding ways to minimize or work around them is an important part of all planning. For Market, planning for data collection and analysis begins with, "How can we design the data collection so that it will minimize all the sorts of error that we could possibly get?" Medicine continues. "So you've got to really think about all sorts of possibilities that could occur while you're stating the problem, while you're trying to obtain information to answer the problem and also when you're answering the problem. You've got to think about unexpected things happening. And also you've got to be aware that people that you might be working with might not have an appreciation of what you're doing in terms of the statistical answer to the problem, so you've got to try and convey to them what you are doing and why you're doing it so that they understand. Biology sees experimental design itself in this light. "Experimental design is part of that, it's just anticipating problems with the levels of evidence so you maximize the quality of evidence you can get and the ease of communicating the results to other people" (cf. Taylor, 1998, page 14).

Market's stories concerned psychological dimensions of the people being studied ("I think psychology is all important."). One story involved a study of forms of income and spending patterns. The client wanted very detailed information. Market anticipated problems with the level of detail (e.g., 94 different sources of income) that one could reasonably expect people to remember. It might be reasonable to expect most people to remember things like employment earnings, but income from interest payments on bank accounts? A prime purpose of the survey was to gauge the extent of abuse of government benefit payments. "Who's going to tell us the truth? I mean if they're actually receiving the benefit that they're not entitled to and we sit them down and say 'Right, we want to hear every single detail about your income and outgoings.' We clearly shouldn't expect to get a right answer. I mean, would you?" Market is trying to anticipate how respondents are likely to react to the questions. She feeds such concerns to the client (cf. earlier comments about managing client expectations; Boen and Zahn, 1982, pages 48–50).

One psychological technique Market uses in questionnaire design is to mix up items so that positive support for something like recycling sometimes corresponds to the right-hand end and sometimes to the left-hand end of a rating scale. "If someone really did [have a positive attitude to] environmental causes and they ended up giving high ratings all the time or saying yes, yes, yes, yes . . . they tend to get a bit uncomfortable with that and think, 'Goodness I've given a lot of high ratings, maybe I should go for a low rating.' So somehow the respondents want to vary their own ratings for their own responses so that it's better to have questions phrased the opposite way around and sometimes they can agree and sometimes they can disagree. It certainly makes them think more strongly about the question you're asking them." It also helps identify lazy respondents who are just marking the same box the whole way down a questionnaire, or to bring to light deeper problems (as we will see).

In a recycling scheme study, Market's team found the data ". . . collected from the XXX people [an ethnic group] just didn't make sense as a whole. They agreed strongly with two statements that might actually be complete opposites . . . [she also described cross-checks with other available information revealing that claimed recycling levels were far too high to be credible] . . . And we realized that our whole approach had been completely flawed because of not being fully aware of certain cultural issues. We were ringing up these people who possibly are new to the country and saying, 'Hi, we're ringing on behalf of the City Council. We want to talk to you about how often do you put out your recycling bin.' I think what happened was that maybe there's more of a deference to authority . . . the average pakeha [Caucasian New Zealander] doesn't really mind telling callers to get stuffed, or say, 'No, I don't take part in your silly scheme.' But a lot of XXX people actually felt quite intimidated by our approach and clearly felt pressured into thinking that they had to agree to everything we'd said." The researchers concern with the data was so great that "We basically had to discard parts of it that didn't clearly make sense." The main message of this story: "You've got to have a lot of knowledge, not necessarily about the subject area thing, what actual product or service you're trying to survey, but about people . . . . It's just a bit harder when you also throw in the cultural differences." Planning to cross-check the data for accuracy against other sources turned out to be critical in this investigation.

Biology talks about the benefits of talking to a lot of people when planning. "You talk to people as much as you can because if you don't, if you can't explain to people what you're trying to do, then you don't understand yourself and at the very least you ought to talk to your colleagues, make sure what you're trying to do, what other people think you ought to be doing. And you improve your understanding as you go along . . . . You're banging other people's heads against the possibility that there's alternative explanations . . ." And the best means of checking the workability of data collection plans? Biology: "The pilot study is to ensure that it's going to work and the mechanics do happen."

But despite all the anticipation and planning, unexpected things often happen that introduce complications. General explains, "You may start out an experiment with the best will in the world. And one cage of bumble bees may die. One of her colonies was very long-lived. And when the original experiment was set up, there was somebody coming in and measuring how much they ate every day. Well, if it goes on too long it starts to run into Christmas and you can't expect somebody to come in and measure the colony every day. So the colony may live for three or four days without anybody making any measurements on it. Even though the experiment is 'well designed,' the data may not turn out to be particularly well designed."

5.4 Data Production

The importance of attention to the data production process was strongly emphasized by several statisticians. They recalled many instances of routinely collected data from institutions, organizations and businesses where many people were involved in the collection of the data and the people were not trained in good consistent procedures. (General: "Basically there was a whole army of service personnel out there collecting the primary data and that was really the problem.") The result was convoluted and haphazard data gathering and record-keeping systems that added unnecessary noise to the data. Quality elaborates. "They're not regarding it as a process which should be managed with a specific objective, that is, collecting or producing data which is actually reliable or accurate. So the variation in the actual measurement process is probably greater than the variation in the signal—which makes it useless .... This whole business of gathering data is not looked at as having a great deal of importance .... You're not feeding it back to the people that you gather it for and so of course they [the gatherers] have absolutely no stake in having good data." As General says, "It's better to measure a small amount of data well than a large amount of data badly." They consider it essential that people involved in collection and record-keeping have an incentive to keep accurate records, that systems be designed around the capabilities of the people who work in them and be designed for the people who will be using the data.

Part of designing robust data collection systems involves knowledge of the problems involved in data collection, record-keeping and storage. Market believes strongly in holistic involvement. She talks of the importance of having "... time to spend with the field force. You go out and do your door knocking .... I think it's important to be involved in things like that so you understand everybody else outside of your discipline so it's like going out and mixing and mingling with people that aren't statisticians, if you like, so you can quickly understand the context, the place of statistics and where it fits in."

5.5 Criticizing and Cleaning Data

Quality described his process improvement project students: "The good ones find out very early that the data they've got doesn't reflect the reality. They need to dive down a bit more deeply. The poor ones accept the stuff at face value and try to analyze it as it goes." (cf. Trosset, 1998, page 24). Our statisticians are suspicious of their data, especially in the early stages. They are on the lookout for variables which do not map well to the characteristics they are supposed to measure, and for implausibilities in the data itself. Medicine: "You get things like date of birth keyed in the wrong way or you might get the date of birth mixed up with the date of the ECG recordings so you get a negative age and, you know, people just tap away ... ."

When done systematically, this is called data cleaning. It is a stage that statisticians plan for, even to the extent of providing automatic data-cleaning devices in the data collection phase (cf. Taylor, 1998, page 15). In large data sets this is a time-consuming (but vital) process aimed at checking individual data values (cf. Hoeting, 1998, page 12). As Market described it, cleaning has three phases. First comes identification which focuses on looking for impossible values and implausible values (outliers), not only in single variables or responses to single item, but for related variables and related items (checks on consistency between related responses). More complicated identification requires skills in statistical analysis. Second comes checking identified problem values for validity against other information sources where possible. Data points which are found to be invalid and unable to be corrected become missing values. Third comes deciding what action to take about points whose validity is still suspect. This opens up a murky area and one our interviews did not probe. Market did talk of imputation for missing values, as is fairly common in survey research. As far as PPDAC goes, imputation can be classified as part of analysis (one of several approaches to missing values). Cleaning the data doesn't completely erase nagging doubts. Medicine says, "Even if it makes reasonable sense, it doesn't necessarily mean it's going to be right, though." Although the data checking and cleaning process begins before analysis starts, it continues through analysis when we notice features of the data (e.g., as a by-product of model criticism) that seem suspect.

We note that with both General's story and the several Quality told about fatally flawed data-collection processes, the story did not stop at "this data is no good." The consulting job then transformed into one of working together with the client to improve the data collection process so that future data would be useful (cf. Kadafar and Morris, 1998, page 26).

6. FROM ANALYSIS TO CONCLUSIONS

6.1 Planned and Unplanned Analyses

Analysis is the subject most emphasized in applied statistics teaching. We have not attempted to weave hard-to-connect technical snippets from the interviewees' discussions into our fabric and have concentrated more on the statistician's interactions with the world in which they work.

Data exploration and hypothesis generation. General and Biology spoke about "... getting a feel for the data" or "fooling around with the data" prior to any formal analyses. Quality considered that looking at the data and asking "What's going on here?" and "Is this common cause variation or is this special cause variation?" as very important. "I think graphical techniques are a much more fundamental part of statistics than the statistical tests and the mathematical statistical theory which we have that underlies all these things. I think that to a large extent what we are trying to do is put a sound basis to what are usually reasonably sound judgements, personal judgements about what's going on here. So I think a lot of statistical thinking really relies on good understanding of how to use graphics." He remarked that his students often did not realize that a histogram said something about what was going on inside the process, that there were clues to be read.

Therefore, the reading and interpreting of graphs requires making connections between features seen in data and "what must be going on in the system" (context knowledge) in order that reasons can be put forward as to why the data looks like this or so that the data can be split in another way or the data can be explored graphically. As Quality eloquently says, "What always struck me when I became involved in quality management was that I had got this degree in statistics. I'd done a whole lot of consulting over the years in all sorts of mostly biomedical, biological and medical areas. And I hadn't really made the connection from the summary statistics of the information through to the understanding of what was going on. What must be going on to generate these patterns?"

According to Biology there is "... absolutely no point in doing a piece of work if you haven't got a question." Analyses cover a spectrum that ranges between situations where questions are narrowly focussed and the analysis is tightly prespecified (as in Phase 3 clinical trials) to situations with ill-defined questions and only the vaguest prior idea of a form of analysis. General says of the latter, "There's just this great mass of data. There's this undigested problem and they want help with summarizing and squeezing down." With problems of this last type, analysis begins with data exploration. With prespecified analyses, exploration is still necessary for data cleaning and model criticism. Then there is the large range of problems in the middle of the spectrum in which one knows what basic modelling tools will be used, but a lot of exploratory work is necessary for model building as well as model criticism. A facet Biology considers important is that "A main objective of a scientific data analysis is to see if any of the competing alternative explanations are clearly inconsistent with the data and then these can be abandoned. Then you are left with (ideally) one explanation for the phenomenon." Exploration of the data to answer primary questions is still highly directed. Biology continues. "After addressing the [primary] question, it is bloody stupid to throw the data away. That data may suggest all sorts of things if you just look at it right." Unexpected features seen in the data trigger new ideas and help one generate new hypotheses, the life's blood of further research. "I'm a firm believer that today's hypothesis is tomorrow's grant application." Unfortunately, "Most people can't afford the time, but in a sense it's a pity because there's an awful lot of data out there with an awful lot of hypotheses just waiting to be generated." Multivariate data provide particularly rich hunting grounds.

Exploration largely consists of looking for patterns and exceptions amidst the variation, and of "extracting signal." Biology gives a vivid evocation of "real or random." "... with the proviso that the human being is hard-wired to see pattern even if it isn't there. It's a survivor trait. It lets us see the tiger in the reeds and the downside of that is that our children see tigers in the shadows on the wall." Separating the real from the random leads us to consider ideas on measuring evidence and characterizing uncertainty (see "significance" to follow). Exploration draws heavily on intuition and context knowledge. "You're using all the priors you've got from everywhere—even if it's just that I had a talk with that guy in the pub at the last conference and he mentioned that in his species it did that. I wonder if.... "When it comes to who should be doing this open-ended exploration, Biology becomes controversial: "The constructive person to do this kind of thing is a scientist. If he has command of the tools of data analysis, then he will pick up or discard and throw away far more constructively than any statistician who doesn't know the system." This is clearly not just any scientist. We are back to Biology's idealized scientist–statistician and the importance of context knowledge. "There's a difference between a student looking at the data and a professor who's been studying all his or her life." He goes on to say that it is not so much knowing about this precise system but knowing about many other "similar" systems whose workings might shed light on this one.

Not surprisingly, General's perspective is statistician-centered rather than scientist-centered, but there are tantalizing similarities. He is often working in situations where he has only a smattering of context knowledge. "I'm using knowledge from other problems and other experiences and other data sets." (other data sets from different topics). He thinks he is using statistical knowledge. "I mean the client always knows much more about the area than you do . . . . Context knowledge builds up incrementally. And if you've got a problem on apples and then somebody brings you a problem on kiwi fruit then . . . you may think to yourself, 'Hey, that worked well, why don't I look at that?' And then you're sort of borrowing from a sort of different context." In view of Biology's comments about "similar systems," some of the differences between Biology and General may simply be semantic differences over the word "context," but not all. General says "What I'm looking for is similarities in the data analysis and the way the data process is going." Patterns and exceptions raise questions which the context expert can then try to answer. So we seem to be identifying two dimensions. First, we have context knowledge which enables recognition of what data features mean and context experiences which suggest other questions to ask of the data. Second, we have past experiences with statistical data which also suggest other ways of looking at the data and questions to ask. We might expect the scientist to have more of the former and the statistician to have more of the latter.

Exploration can throw up things that are unexpected or do not make sense in context or theory. Brain, who does not usually do the actual data analysis on the imaging project, found one set of data, ". . . so fascinating I did the analysis myself. I found something around here in this area. I said, 'What about that?' and they said 'Oh, that must just be an artifact.' And yet this thing is statistically significant but it wasn't part of their hypothesis and the funny thing is in the second set of data the same thing showed up again. It's a completely unexpected finding in an area of the brain that's not supposed to be related to this thing at all . . . . It's been replicated in two independent sets of data so it's got to be there." (cf. Kadafar and Morris, 1998, page 25).

Biology sees a larger role for the statistician in answering primary questions. "To a large extent, many experiments, once they've been designed to a certain point, can be taken over by people with very little understanding of the system, even to the point of the immediate interpretation of the results . . . . The final discussion, of course, ought to have the global context, and the design ought to have the global context."

Models and the behavior of data. The conclusions our statisticians draw from data are based upon models of one form or another. They criticize their models using the data in hand "to make sure that your analysis accommodates the reality of the data" (Biology) before using them to draw conclusions. Models which are contradicted by the data sufficiently strongly are not used. Furthermore, structural features of models used for explanation should correspond to meaningful aspects of the context reality. (This is not critical with models used purely for prediction.)

The importance of assumption checking before analysis was described by Brain. For example, if he was treating the data as independent then he would have to assume there was no temporal correlation. This was important for brain measurements as the scans had ". . . to be separated in time by something like fifteen minutes so there is little carryover from one scan to the next scan and so you are pretty safe on that assumption but there are other types of data where the scans are taken every second or so and then you definitely can't make the assumption that they are exchangeable and then you have to be quite serious about how you account for this sort of temporal correlation."

Brain related the following story which illustrates how he found a way of checking his models so that he and his collaborators could be convinced that they were sound. The original (and simplest) problem involved finding the equivalent of t-test thresholds. There were n people, where n is small because the scanning data was extremely expensive to obtain. For each person, baseline images are taken. A stimulus is given and then new images are taken. Locations in the brain showing a change in activity greater than the threshold would be considered to have been activated. This is a situation in which, if only one location were involved, one would use a paired t-test. However, thousands of brain locations were involved, leading to high-dimensional data with very strong spatial autocorrelations. The signals in the data tend to be weak compared to the level of the noise. When Brain encountered the situation, his intuition said that their fairly naively set thresholds were too low. After developing his theory, he arrived at quite different thresholds. Fortuitously, he happened upon a set of experimental conditions which had been repeated twice. By subtracting baselines he got a three-dimensional "pure-error" estimate of spatial variation. He showed the researchers that their thresholds for signalling brain activation were set too low because they were signalling activated areas in the brain when no activation had occurred. "It was only at that point they really trusted that this theory was working." It also gave him more confidence in the applicability of the theory himself, based as it was on fairly strong assumptions (e.g., multivariate normality) and being applied to small sets of data (cf. Tweedie, 1998, page 2; Moore, 1990a, page 265).

6.2 Toward Conclusions

The title here is informative. The path from analysis to conclusions is not linear. We begin to form conclusions while in the process of doing analysis, and the emerging conclusions can have a large effect on the future course of an analysis. Market summarizes the conclusions phase: "I think one of the key roles as a statistician is to take the output of the analysis stage and actually relate that back to the original objectives and questions, and then to communicate the results in such a way that they are understandable to the client." The issues here are those of mapping back to the context realm, deciding what can be reasonably concluded about the context reality, and the communication of these new understandings. Other individuals are more likely to be convinced by the conclusions if the statistician can show that these emerge from the data.

Does it "make sense?" There is a continual checking throughout an investigation back to the context reality. Checking "Does this make sense in context terms?" occurs while making modelling decisions and then again in forming conclusions. What "makes sense" in context terms is a two-edged sword, however. As Biology states, "Today's dogma can be tomorrow's bad joke." Competing explanations must be carefully considered in the interpretation of the data, "If in the writing up you don't acknowledge that fact [the plausible alternative explanations], then you are likely to end up with egg on your face, because other people will take great delight in rubbing your nose in it." We recall General's story at the end of Section 5.2 of conflicts between a client's prior beliefs about a system and the data, where he had to "show them that their belief is not supported by the data." In this case the data was used to provide a check on assumptions about the system itself, not just about a statistical model. Recall also Brain's story (Section 6.1) of the location in the brain that "should not" have been activated by a particular stimulus but clearly was. In Market's recycling story, she checked against context reality and threw away some of the data relating to a substantial subgroup.

Design issues. Biology drew attention to how design is linked to generalizability of conclusions. "The whole concept of inference is determined in the first instance by the inference space you specify, and that's determined by the population to which you wish to generalize. So, unless you can identify what a population is of sampling units, you can't actually determine what generality your conclusion has. The inference space is at the root of many of the best arguments going on in biology at the moment. People are making inferential claims relating to populations or spaces from which they have not actually sampled." Medicine talks of the difficulties in generalizing from results on study groups, which have particular profiles and correspond to specific subpopulations, to making treatment decisions for patients in general. There can be a tension between a pressing practical need to generalize results beyond the study population and the knowledge that we can be led badly astray by doing so. Statisticians constantly remind clients of the latter point.

Significance and statistical justification. Biology worried about the way significance is often treated in his field, especially the equating of lack of statistical significance with lack of biological significance. Lack of statistical significance often simply means that you did not look hard enough or did not collect enough data. Medicine echoes this point. The difference between statistical significance and scientific significance is an issue that seems to be attracting more attention in statistics teaching. The scientific significance of an effect depends upon its consequences and that in turn tends to depend upon its size. Statistical significance concerns the evidence for the existence of an effect and says nothing at all about its size. One can go further, however, as Biology points out. Those situations where we are most concerned about statistical significance (or lack thereof) may also be situations in which the data contains insufficient information to say anything definitive about scientific significance because the confidence limits extend from trivial effects to radical effects, or even between radical effects in opposite directions.

Medicine also voices other concerns about the treatment of "nonsignificance." Whereas Brain might use slightly lower thresholds to suggest promising areas of the brain for further investigation, some of Medicine's clients push this to extremes where even nonsignificant differences that are small compared with standard errors can be seized upon as "suggestive of an effect." They cling to their beliefs despite a nonsignificant result, justifying this by the fact that the sample size was small. On the other hand, however, when they did obtain a significant result (which they then believed was convincing), they worried that other people might point out that "the sample size may be too small, may be unrepresentative."

6.3 Communicating Conclusions

The question of how to allow *this data* to speak to *this client* is at the forefront of the consciousness of each of our statisticians. Medicine says, "Multiple regression might not mean very much to people but if you actually demonstrate things graphically then that tells a lot more .... I tend to try and produce diagrams which are meaningful to the person that I'm working with, they are meaningful to their field." Brain's color-coded test results superimposed on a three-dimensional image of the brain (see Section 3.1) are a vivid example. They link statistical results to problem context in a way that permits the results to be immediately absorbed in context terms. Other statisticians also told of attempts to produce graphics with this sort of immediacy of communication (cf. Hahn, 1990, page 259).

Some of Market's commercial clients have no interest in the statistical underpinnings of her conclusions. These clients simply want "the bottom line," "They want informed recommendations and [to do this] you have to understand other parts of their business .... So we hide the statistics from them; it's something that's under the bonnet [hood] if you like. It's the engine driving the whole process, but it's not actually what they're interested in. They are interested in the red shiny sports car on the outside that allows them to go somewhere .... [Often] we would write a report for them that simply answers their questions .... It wouldn't be obvious to them necessarily that we've gone away and done a whole bunch of statistics." The technical details are something these clients pay her for and trust her to deal with. Many, however, want justifications once they have heard conclusions. The latter "makes them interested in then hearing the justification." Statistical appendices in a report provide the client with some assurance that the conclusions are backed by analysis, even if they do not bother to read and critique any details of that analysis. Market's insights reported in Section 5.2 about the contradictions of people's reactions to statistical information is relevant here.

Communication of results can be thought of (adapting words of Market) as *translation into the language of the client* (or readership of a report) (cf. Kadafar and Morris, 1998, page 27). This tends to put a premium on reporting in terms of very simple plots and summaries. The "language" of large numbers of clients does not include confidence intervals and p-values. Other clients, Medicine finds, are more than happy to receive these, despite not really understanding them, because they are commonly used in the field. Brain reports a great deal of difficulty with getting across the basic distinction between a parameter and an estimate. We discussed the difficulties caused by transformations (reexpressions) for communication with a client earlier in Section 5.2. Our statisticians work very hard at determining the "right" technical level for communication. With many of Market's clients, this does not extend beyond bar graphs, percentages and a vague notion of margin of error.

In summary the report and visuals are tailored to the needs and statistical understanding of the clients. The visual communications are used to facilitate the mapping from the statistical system to the real system through the use of a coding that is meaningful and easily interpretable by the clients.

7. DISCUSSION

A number of themes pervade this paper. We took as our starting point a conception of a whole (or real-world) problem relating to some system and statistics being applied to contribute an additional understanding of the system. We moved on to the realities of the environments in which statisticians work and the constraints these impose on the ways in which they approach the problem. Particular attention was paid to interactions with clients, and the effects of the client's expectations, knowledge and psychology and the way these influence planning, analysis and reporting. We talked of the way in which others tend to confine statisticians to a narrow technical role, unaware of the contribution they can make to the whole process, and of the need for the statistician to take positive steps to avoid this. Foundational elements underpinning statistical thinking were discussed, including the taking account of variation, transnumeration, constructing and reasoning from models, and the integration and synthesis of problem context-matter and statistical understandings. This led into a discussion of personal, dispositional elements that affected the statistician's problem solving (e.g., imagination, scepticism, curiosity).

The remainder of the paper was built around the statistical empirical enquiry cycle. There was no attempt to be comprehensive about the cycle since it was merely used as a convenient structure for organizing the data from our interviews. The problem-to-plan progression described is necessarily cleaner than the reality. In reality, there is no precise point where isolating the questions to be asked stops and thinking about how we should obtain data to answer them starts. We can constantly shuttle between these modes as we edge towards a mature conception of the questions. Thoughts about how we measure things, and even analyze the resulting data, can be intricately entwined with the ways in which we conceive problems and frame questions.

We highlighted the importance of whole-problem understanding, of the distillation and encapsulation of complexity leading to the isolation of relevant statistical subproblems, of determining the relevance of available data to problems, and of obtaining new data that do address the problem. Interrogation of others was seen to be a crucial vehicle for gaining context understanding and for communicating, adjusting and refining the understandings of statistician and client throughout the whole statistical process. We addressed measurement and design issues, the limitations of data and the difficulties in obtaining relevant data. We gave particular emphasis to seldom-discussed psychological aspects (e.g., "measure everything in sight," "the sanctity of the measured variable," and other general attitudes to statistical data). An important part of planning is the anticipation of possible problems and planning one's way around them, be it with a measure, the units to be measured, the data collection process, the analysis or the conclusions. Throughout the investigation process there was a continual checking for validity and positing of other explanations for a phenomenon. We proceeded to the need for well-thought-out processes for data production and management that apply basic tenets of quality assurance. This was followed by the progression through analysis to conclusions.

Underlying all of this is the building of statistical models to capture relevant aspects of the context reality, the continual checking of mappings between context reality and statistical system, between models and data, and of the judging of the reasonableness of solutions in the conclusions phase. Features in data are connected to context experiences and other experiences with data, both of which point to questions to ask that affect the course of an analysis. We noted how statistical knowledge and experience confer more than methods of analyzing data. They provide new ways of conceiving the nature of the context-matter problem. Reporting revolves around the central question of how best to allow *these data* to speak to *these clients* (or audience).

The practising statisticians that we have interviewed have revealed a number of dimensions to success as an applied statistician that are not covered in any standard statistics course. Many of these dimensions relate to the "up-front" elements like understanding the dynamics of a system, problem formulation, measurement and nontechnical aspects of the planning of studies. A surprisingly strong thread through much of the discussion concerns understanding the thinking of other people, be they clients, coworkers or study subjects. The challenge now is to use communalities, such as those identified in this paper, in the development of teaching that incorporates more of the "art" of statistics.

We leave the reader with this intriguing extract from Market: "I think of it as statistics, but I think a lot of other people wouldn't necessarily think of it as statistics. There isn't a formula for it and it very much involves common sense."

APPENDIX

The software tool, NUD*IST (Richards and Richards, 1995), provided the means for each statistical issue being discussed to be extracted from the transcripts and sent to one or more relevant categories. An in-depth analysis of each category, which contained the extracts from the transcripts of all the statisticians, enabled common themes to be synthesized. The resultant understandings and interpretation were checked, corroborated or refuted by the subjects.

Because we were attempting to understand and interpret the complex thought processes of statisticians interacting with their environment, a qualitative research method was the most appropriate to employ. We attempted to capture a partial reality of a whole situation through engagement with practitioners and through that engagement, to develop a sense of meaning and understanding. The strategy of using independent analyses from two researchers, and using the interviewees and other statisticians to vet the analysis and interpretation of the data, was instigated to add more credibility to the soundness of the methodology. A triangulation method involving other studies was used to further validate our interpretation and understandings of the common elements of statistical thinking processes employed by these statisticians. (See Denzin and Lincoln, 1994, and Miles and Huberman, 1994, for literature references on qualitative research.)

ACKNOWLEDGMENTS

The authors thank the Editor and the referees for their careful reading and constructive comments, which helped improve the paper.

REFERENCES

ARMSTRONG, R. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 249–250.

BOEN, J. and ZAHN, D. (1982). *The Human Side of Statistical Consulting*. Lifetime Learning, Belmont, CA.

BOX, G. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 251–252.

BROMAN, K. SPEED, T. and TIGGES, M. (1998). Estimation of antigen-responsive T cell frequencies in PBMC from human subjects. *Statist. Sci.* **13** 4–8.

CHATFIELD, C. (1991). Avoiding statistical pitfalls. *Statist. Sci.* **6** 240–268.

COBB, G. and MOORE, D. (1997). Mathematics, statistics and teaching. *Ameri. Math. Monthly* **104** 801–823.

DENZIN, N. and LINCOLN, Y. (eds.) (1994). *Handbook of Qualitative Research.* Sage Publications, Thousand Oaks, CA.

HAHN, G. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 257–258.

HAWKINS, A. (1996). Can a mathematically-educated person be statistically illiterate? Paper presented at the 1996 Nuffield Conference.

HAWKINS, A. (1997). Commentary on "New pedagogy and new content: the case of statistics." *Internat. Statist. Rev.* **65** 141–146.

HOADLEY, A. and KETTENRING, J. (1990). Communications between statisticians and engineers/physical scientists. *Technometrics* **32** 243–274.

HOETING, J. (1998). Sandbars in the Colorado River: an environmental consulting project. *Statist. Sci.* **13** 9–13.

HOOPER, J. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 259–260.

HUNTER, J. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 261.

HUNTER, W. (1981). The practice of statistics: the real world is an idea whose time has come. *Amer. Statist.* **35** 72–76.

KAFADAR, K. and MORRIS, M. (1998). Commentary on "Consulting: real problems, real interactions, real outcomes." *Statist. Sci.* **13** 25–29.

KIRK, R. (1991). Statistical consulting in a university: dealing with people and other challenges. *Amer. Statist.* **45** 28–34.

MACKAY, J. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 263–264.

MACKAY, J. and OLDFORD, W. (1994). Stat 231 Course Notes Fall 1994. Univ. Waterloo.

MACKINNON, M. and GLICK, N. (1999). Data mining and knowledge discovery in databases: an overview. *Austral. New Zealand J. Statist.* **41** 255–275.

MALLOWS, C. (1998). 1997 Fisher memorial lecture: the zeroth problem. *Amer. Statist.* **52** 1–9.

MILES, M. and HUBERMAN, A. M. (1994). *Qualitative Data Analysis.* Sage Publications, Thousand Oaks, CA.

MOORE, D. (1990a). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 265–266.

MOORE, D. (1990b). Uncertainty. In *On the shoulders of giants: new approaches to numeracy* L. Steen (ed.) 95–137. National Academy Press, Washington, D.C.

MOORE, D. (1992). Statistics for all: Why? What and how? In *Proceedings of the Third International Conference on Teaching Statistics* (D. Vere-Jones, ed.) **1** 423–428. Internat. Statist. Inst., Voorburg.

MOORE, D. (1998). Statistics among the liberal arts. *J. Amer. Statist. Assoc.* **93** 1253–1259.

PROVOST, L. and NORMAN, C. (1990). Variation through the ages. *Quality Progress* December 1990 39–44.

ROBERTS, H. (1978). Statisticians can matter. *Amer. Statist.* **32** 45–57.

RICHARDS, T. and RICHARDS, L. (1995). NUD·IST. Qualitative Solutions and Research, La Trobe Univ., Bundoora, Australia.

SNEE, R. (1990a). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 267–269.

SNEE, R. (1990b). Statistical thinking and its contribution to quality. *Amer. Statist.* **44** 116–121.

TAYLOR, S. (1998). Setting up computer-assisted personal interviewing in the Australian longitudinal study of ageing. *Statist. Sci.* **13** 14–18.

TRIBUS, T. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 271–272.

TROSSET, M. (1998). Commentary on "Consulting: real problems, real interactions, real outcomes." *Statist. Sci.* **13** 23–24.

TWEEDIE, R. (1998). Consulting: real problems, real interactions, real outcomes. *Statist. Sci.* **13** 1–3.

TWEEDIE, R. and HALL, N. (1998). Queueing at the tax office. *Statist. Sci.* **13** 18–23.

WANGEN, L. (1990). Commentary on "Communications between statisticians and engineers/physical scientists." *Technometrics* **32** 273–274.

WILD, C. J. and PFANNKUCH, M. (1999). Statistical thinking in empirical enquiry (with discussion). *Internat. Statist. Rev.* **67** 223–265.