Anecdote, fiction and statistics: the three poles of empirical methodology

2 July 2004

Michael Wood
University of Portsmouth
Richmond Building, Portland Street
Portsmouth, PO1 3DE,
England

michael.wood@port.ac.uk or michaelwoodslg@gmail.com
http://woodm.myweb.port.ac.uk/

This article clarifies the role and value of three approaches to empirical research – anecdotes (derived from case studies or small samples of data), fiction (often not regarded as a serious approach to research) and statistics. The conclusion is that all three have an important part to play. Many conventional stereotypes are deeply unhelpful: contrary to the usual assumptions, science is usually dependent on anecdote and fiction, qualitative research is often statistical in spirit, and inquiry is most unlikely to lead to useful change without drawing on anecdote and fiction.

Keywords: research philosophy, qualitative research, quantitative research

Contents

INTRODUCTION

Research in management, and many other areas, encompasses a variety of differing methods, differing views about what is acceptable or useful, and differing guiding philosophies. However, anything claiming the title "research" must relate to reality in some way. By reality, I mean the world of people, events and organizations.

There are very different views about the nature of this reality and our knowledge of it: some would stress "objective" knowledge as the only legitimate goal, whereas others would say that different people may, legitimately, have different versions of the "truth" about reality. My argument in this article does not depend on the view taken here; all I am assuming is that there is something reasonable stable about which it is meaningful to try to find some sort of truth. I will use the term “reality” for this something that is being researched.

Typically, the researcher interacts with reality to obtain data. These data may take various forms: a transcript of an interviewee's response to a question, the ages of a group of people, the salary of a manager, etc. These are the "facts" on which the research is based. The researcher then analyses the data to derive conclusions. The conclusions may suggest searching for further data, which may result in the conclusions being modified, and so on.

In practice this is not, of course, as straightforward as this might imply. Philosophy of science teaches us that empirical assertions are dependent on the theoretical perspective from which they are viewed. Data is never given as the Latin origin of the word implies; it needs interpreting, and where the interpreting finishes and the analysis starts is inevitably hazy. The transcript of an interviewee is an interpretation of the "raw" data that consists of a pattern of soundwaves (although a physicist might challenge the rawness of even this data). And what the researcher is actually likely to use is not the transcript as such, but rather the meaning of the words when interpreted as ordinary language. Similarly, the manager's salary is likely to need a certain amount of interpretation if it is to be a useful piece of information. Is the source accurate, does it include perks, is it before or after tax, and so on. And with the ages of the group of people, we may need a statistical summary (mean and standard deviation perhaps) in order to handle the information. In each case, the *approach to reality* is via data that needs collecting and then interpreting.

This article starts from a categorisation of three broad categories of such approaches to reality - which I will call anecdote, fiction and statistics.

Anecdotes are descriptions of particular cases or events: for example qualitative researchers often collect anecdotes from interview data. Fiction, of course, is similar to anecdote except that it is invented, and so at first sight seems irrelevant to researching the real world - this is however, as many would acknowledge, and I will argue below - an unhelpfully restrictive view. Statistical assertions are, of course, widespread in management research. The purpose of the article is to explore the role and value of these three.

Anecdote, fiction and statistics are the building blocks of empirical methodologies - ie "combination[s] of techniques used to enquire into a specific situation" (Easterby-Smith et al, 2002). For example, Popper's falsificationism (Popper, 1980) recommends an initial period of hypothesis generation - which may be inspired by anecdotes, or by fiction or by statistical results - and then a testing phase which is essentially a search for anecdotes (see below). And there are many recipes for various kinds of research based on statistics. My focus here is on the building blocks - anecdote, fiction and statistics - not on the recipe for the whole methodology.

As all empirical assertions depend on the perspectives, or theories, which are used to interpret them, they may be flawed if these theories or perspectives are flawed, and different perspectives may lead to different empirical assertions. If a good quality system is defined in terms of having a number of specific monitoring processes in place, then the researcher will record as a fact that an organization with these processes has a good quality system, irrespective of the effectiveness of the system from a wider perspective. Such anecdotes may be the basis of broader statistics that, again, will be dependent on the way quality is defined. A good research method will acknowledge this type of difficulty, and recommend counter-measures - perhaps the use of a devil's advocate, or a systematic method for analysing values such as "good". The present article, however, focuses on the nature of anecdotes, fictions and statistics, and not on this particular issue.

I will take anecdote, fiction and statistics in turn, and consider their nature and role in research, how they relate to some of the stereotypes that dominate methodological discussions, and how they might fit together in a typical research project.

ANECDOTE

By *anecdotes* I mean descriptions of *particular* cases or events. They are the basis of case study research, any research where the detail of unique cases is reported or analyzed, journalism in the sense of reporting the news, and "creative nonfiction" (Akin, 2000). Anecdotes have a bad reputation in many parts of the research community, and the word "anecdotal" is often used in a derogatory sense, because they may be selected to prove whatever the narrator wants to prove, and because they do not appear to be a suitable basis for generalising to a wider context.

There are, however, circumstances in which anecdotes are of obvious value. Take the case of a management creed promoted by the gurus but which does not usually work in practice. Suppose that an example of successful practice has been discovered. This would be an example of what is possible which may be worth studying and perhaps emulating. The example may be chosen because it shows the creed in a positive light, and there may be no basis for statistical generalisations, but it still *illustrates an interesting possibility*: what can happen, not necessarily what will happen with any specifiable probability. As well as explaining, or providing hints about, *how* the creed has worked, the detail in such an illustrative case may also be helpful for bringing concepts to life, for making them "stick" (see the discussion of fiction below). There are many other contexts in which such illustrative research may be useful for similar reasons: e.g. an analysis of cases of people with natural immunity to a disease may be invaluable for the management of that disease.

To be studied these possibilities have to be discovered. This may entail studying large populations, not to arrive at statistical generalisations, but to find illustrative examples from whose study useful progress may be made. Decision procedures for guiding such *prospecting research* have been derived (Wood and Christy, 1999; Christy and Wood, 1999; Wood and Christy, 2001); these are, of course, entirely distinct from statistical decision procedures.

Anecdotes cannot, of course, be used in any rigorous sense for demonstrating the validity or invalidity of statistical laws. They are also of limited value for supporting universal laws which assert that something is *always* the case. They may, however, be useful for demonstrating that such universal laws are false. The assertion, for example, that a particular management creed always works, can be falsified by a single anecdote about its failure. This is the basis of Popper's description of how science works (Popper, 1980). It is however, of limited interest in management because non-trivial universal laws (as opposed to the statistical patterns discussed below) are extremely rare. The main value of anecdote in management is that it can illustrate what is *possible*.

A good example (ie an anecdote) of the consequences of ignoring this role for anecdote is provided by the results of an experiment on the test for athletes to see if they have been taking the drug nandrolone (Mackay, 2000). This demonstrated, on the basis of a very small sample, that a positive test result *could* be caused by ingesting legitimate food supplements. This result was dismissed by a spokesman for the International Amateur Athletic Federation as having no scientific validity, largely because of the small sample used. However, it is this argument that has no validity: the experiment demonstrates the *possibility*, which is all that is being claimed. This episode illustrates well the dangers of the blinkers imposed by the crude statistical paradigm, which tends to dismiss research which demonstrates that something is possible as "mere" anecdote.

In the natural sciences anecdotes often play an even stronger role than this. Contrary to Popper's ideas, Einstein's theory of general relativity was widely regarded as being confirmed by observations made during the total eclipse of 1919. The general law was considered confirmed by an anecdote about what happened at a single event. There was no sample of eclipses; just the single event was viewed as sufficient.

Some management research does claim that similar generalisations can be made from a single case: for example, Newton et al (2003: 152) argue that their case study organization was “typical”, and so their findings are “generalizeable” to other “similar” situations. The difficulty here, of course, is that “similar” in a management context is a fuzzy concept, and the noise created by uncontrollable variables means that generalizations are rarely as secure as they are in physics.

The relativity example also illustrates one of the problems with anecdotes: what is ostensibly the same event may be interpreted in very different ways. According to Hawking (1988)

> "later examination of the photographs taken on that expedition [to view the eclipse] showed the errors were as great as the effect they were trying to measure. Their measurement had been sheer luck, or a case of knowing the result they wanted to get, not an uncommon occurrence in science." (p. 32)

This emphasises the importance of a critical attitude to anecdotes, but it does not detract from the value of anecdotes for demonstrating what is possible.

FICTION

Fiction is invented: made up without any necessary concern for truth. Fiction is typically contrasted with facts that are true, and with the theories of science - which are usually accorded a type of "provisionally true" status.

If we take the view that all knowledge is a human construction - i.e. it is invented - and that it is difficult, verging on impossible, to define uncontroversial methods of discerning truth, then the distinctions between fiction and fact, and fiction and theory, look very hazy indeed. The differences are perhaps most usefully viewed as a matter of *intention*: facts and theories are intended to be true whereas fictions are not. On occasions the distinction is blurred: creative nonfiction (Akin, 2000) can perhaps be regarded as anecdote which acknowledges its fictional aspect. (There is also, of course, fiction posing as fact; here I am concerned just with fiction intended as fiction.)

Anecdotes and statistics are empirical assertions about reality in a reasonably straightforward sense. Fictions are not intended as empirical assertions about reality, but they can be regarded as assertions about a *possible or hypothetical reality*, which may be useful for understanding "real" reality. In what circumstances, then, is it a valuable research ploy to make fictional "empirical" assertions?

There are a surprising variety of such circumstances - although this is perhaps less surprising if we reflect briefly on the role of fiction in everyday life. Many books, TV programmes and films tell fictional stories. All cultures throughout human history have had important myths and stories. We often find fictional stories more interesting than accounts of real events. And apart from story-telling, we need to consider hypothetical futures - most of which will never come to pass - in order to plan for the future. From this perspective it is hardly surprising if fiction has an important role to play in research.

Many people claim they can learn more about life - real life, one assumes - from fictional works, than from science. For example, the novelist, Julian Barnes, is quoted in the (UK) Guardian newspaper (29 July 2000):

> "When I read non-fiction I am often aware that it is merely a masquerade of the truth. When you read the great and beautiful liars of fiction you feel that this is what life is. This is true even though it is all made up."

In a similar vein, the "parables of leadership" recounted by Kim and Mauborgne (1992) were intended to "capture the unseen space of leadership" - which they feel is difficult to achieve by more direct approaches.

Fiction is also used in sciences such as physics. It is not, however, called fiction; instead, fabricated stories are called thought experiments, or examples, or, if aimed at beginners, problems or exercises. The conclusions are not fictions, but the "data" on which they are based are.

Some of the great theories of physics derive from thought experiments. Einstein's special relativity (one of whose conclusions is that $E=mc^2$) was prompted by imagining someone travelling at the speed of light (Bernstein, 1973: 39). This would be impossible in practice, but is possible in a thought experiment. Such thought experiments can do more than provide inspiration: they can provide a convincing means of establishing the truth of a conjecture. To take a more accessible example, it is possible to demonstrate that heavy objects *must* fall at the same speed as light objects (ignoring the effect of air resistance) by the following thought experiment:

> Imagine two 1 kilogram weights falling freely. They will obviously fall at the same speed. Now imagine the two weights are joined by a piece of string. This will obviously make no difference to the speed of fall, but the two weights can now be regarded as one 2 kilogram weight, which obviously falls at the same speed as the two 1 kilogram weights. The argument can obviously be extended to demonstrate that any two weights will fall at the same speed.

In this particular case, it would be quite easily to perform these experiments for real, but if you accept the assertions described as "obvious", as being obvious, then there is no need. A thought experiment proves the point. We can deduce real truths from fictitious stories. (This is not to deny that the term "obvious" deserves some consideration, and that there might be situations where one can be misled by what appears obvious but is in fact false. Reality may be a helpful check on thought experiments.)

The truths derived from conventional stories are of a similar type: in this sort of situation, it is plausible that this sort of thing will happen because ... and the fiction will then delve into details of characters' state of mind and so on. Note that the word "plausible" indicates that this argument will rarely be as compelling as the story about the two weights. However, this does not invalidate its usefulness.

The end result is that readers end up with a deeper understanding of the situations depicted in the stories, whether they be social situations or physical events or whatever. In the case of physical science, the final form of this understanding is typically a mathematical model; in the case of a novel it would typically be less articulated. Taylor (2000) refers to the audience for his plays

> get[ting] the idea in the gut. The ideas may then bubble up to their heads. This is important to me because I believe that ideas that are communicated intellectually generally don't stick. The idea needs to be in someone's gut and their head for it to stick.

(p 305)

Novels and plays, then, work by communicating ideas implicitly or tacitly - which may be helpful for them to "stick". Taylor (2000) refers to this as "aesthetic theorizing". Sometimes, presumably, it may be very difficult to develop or communicate the ideas by any other means. And on some occasions, the ideas may fail to bubble up to the head, and then we are just left with the aesthetic or the tacit theory.

The "examples" in mathematics and science textbooks represent the application of general principles to particular circumstances. The stories behind these particular circumstances are necessary to make the ideas "stick", as anyone who has tried to understand a piece of mathematical reasoning without working through examples will appreciate.

For these reasons, stories of various types are often important for theorizing. Factual anecdotes may play a similar role, but fiction has a number of advantages over fact:

1 Fictions generally require less research. It may be easier to make things up than to find out real facts. I have used fictional examples in this article when real examples would have added little to the argument. I invented the assertion below that men get paid 30% more than women: this is a trivial example where little would have been gained by using a genuine anecdote or statistic.

   On the other hand, this is not to deny that many writers of fiction devote considerable effort to researching their topic so that they get the background right. The details may be invented, but the background context, and the way the situation "works" are, in essence, true. Einstein's use of thought experiments obviously depended on a very deep intuitive understanding of nature.

2 Fictions can be designed to test particular parts of a theory. A fictional story - a hypothetical situation - can be made up to put a theory to the most severe test imaginable - thus satisfying Popper's requirement that one should really try hard to find stringent tests for a theory.

3 Fictional stories can substituted for anecdotes that could not be researched and told for practical or ethical reasons, or because of the sensitivities of the people or organizations involved. Journalists may fictionalise details of individuals to protect their identities; management researchers may write up an analysis of a fictional organization for similar reasons.

4 Fictions can explore circumstances that have never happened. These may be possibilities which are worth encouraging, or they may be hypothetical circumstances whose feasibility we wish to explore.

   Writers of fantasy and science fiction take this principle of exploring unrealized possibilities to a deeper level. They explore what happens if the background rules change. Another genre worth noting here are utopias and dystopias: fictional explorations of how societies should, or should not, be managed (see Carey (1999) for a survey of such visions over the last four millennia).

5 Fictions can also be designed for impact and "stickiness" - they can be designed to make the points in a way that "sticks".

On the other hand, obviously, the advantage of factual anecdotes is that they have a built in reality check. Accounts of the moon based on NASA data have more credibility than science fiction. But the fictional accounts do still have their place - for the reasons explained above.

These arguments apply to physical science, social science, and to management. However, in management the rules tend to be less firm and well understood; we can play out a fictional situation using Newton's laws and be pretty sure the answer will be right, but with a social or management situation can we have the same level of certainty that our imaginary playing out of the situation will correspond to what would happen in reality? Probably not.

This means that the role of fiction in management is likely to lie in establishing possibilities rather than general or statistical laws: in demonstrating what might happen, rather than what will happen. Prospecting research (Christy and Wood, 1999) involves looking for empirical illustrations of possibilities, but the fictional mode extends this by including fictional possibilities - things that might happen in the future although they may not have done so far.

The value of this goes beyond exploring general principles. The key concern of management is how to best manage the future: a crucial aspect of this is envisaging a wide range of possibilities - some of which may be worth striving for, whereas others may prompt avoiding action. Initially, some of these possibilities have to be invented; they start their life as fictions.

The fictional stories useful to management research are of a variety of types ranging from fully fledged stories to trivial examples. There are also simulations and role plays, which are designed to explore the consequences of hypothetical, or fictional, scenarios. A computer simulation of a variety of new production systems, for example, may enable problems to be foreseen and the best system to be selected without the necessity to perform a real experiment. What-if models on spreadsheets, and mathematical optimisation models, are all designed to compare a number of hypothetical scenarios with a view to choosing the best in some specified sense. All these techniques will benefit from a high level of creativity in generating new scenarios to test out.

There is another context in which hypothetical scenarios are compared, which leads us into the next section of this article. Statistical inference can be viewed as a means of deciding which of a number of *possible worlds* is closest to the world we live in. All but (at most) one of these must be fictional.

STATISTICS

This is the pattern for much of the research reported in management journals. A sample of instances of whatever is of interest is observed and used as the basis for statistical generalisations. On average men get paid 30% more than women; performance related pay usually fails to improve productivity; standards of numeracy are lower in the UK than in most other countries; and so on. (These are fictional examples, but they are adequate for my purposes, as discussed above.)

The essential feature of these statistical assertions is that they are based on the *frequency* of particular values of measurements in the sample (Wood and Christy, 1999). To find out how much more men earn we need to know the frequency of earnings in each range so that we can work out averages. Similarly, it is not enough to have anecdotes of performance related pay failing to improve productivity; we need to know how *often* it happens so that we can see if the word "usually" is justified.

The instances on which the statistics are based are, in effect, factual anecdotes, but the statistical researcher is not interested in the uniqueness of each case; all that is reported and analyzed is the statistical summary of the sample, which, of course, is presumed to tell us about the underlying population. One of the key assumptions of statistical analysis is that cases are exchangeable (Draper et al, 1993); the sense in which they are unique is considered irrelevant. This style of research is what many people mean when they talk of research in management or similar areas: generalisations based on real data, thus reflecting the truth about the way the world "really" works.

In natural science, non-probabilistic laws - $E=mc^2$ on every occasion, not merely on average or on 80% of occasions - are feasible and so they are the goal; in management, everyone acknowledges that this is not often possible, so assertions qualified by provisos like "usually" or "on average" have to do instead. Statistical assertions like these may be all that is feasible.

However, we should not forget that quantitative conclusions do *not* have to depend on statistical data. $E=mc^2$ is quantitative but not statistical. And there are examples in management: the well known formula for the economic order quantity in the theory of stock control (as described in most introductory texts on management science) depends on a number of straightforward assumptions and differential calculus, but *not* on a sample of data analysed statistically. (It is an optimisation model based on comparing a range of hypothetical order quantities - ie a range of fictional scenarios.) On the other hand, in management, examples like this are relatively rare: most quantitative conclusions - e.g. any regression model - are derived from samples of data analysed statistically.

Most statistical research in management is observational; the situation is observed without any experimentation or manipulation of the situation. Occasionally, an experimental approach is taken - something is changed and the effects are assessed and compared with a control.

Observational (non-experimental) statistical research can be very useful - eg epidemiological research is clearly of potential benefit for the management of our health in the future; statistical research on gender differences in various contexts also has obvious uses. There are two important presuppositions:

1. The future situation will resemble the past - from which the sample of data is inevitably drawn - in relevant respects.
2. The features of interest can be meaningfully and usefully summarised so that they can be compared or aggregated across cases.

To illustrate the problems when these presuppositions do not hold, consider research on companies whose main business is based on the world wide web. Statistical survey results based on the past - as they inevitably must be - are of limited use because the business and technological context is changing so rapidly. In particular, the most interesting types of web business may not have been invented yet, so no amount of diligent statistical research can possibly uncover anything about them. The ideology of this mode of research - its unspoken assumptions - are conservative, in the sense that it cannot describe new possibilities, only old ones. If statistical research is to be useful, it is necessary to research a level at which the situation is not likely to change - e.g. the biological level, or a behavioural, organisational or economic level at which web-based companies are similar to other businesses. Only then is the conservatism inherent in the statistical approach reasonable.

The second presupposition - which applies equally to experimental statistical research - is that the features of interest can be summarised as a number, a series of numbers, or a category, or in some other way that can be processed by statistics. If every case is seen as an individual, so that population summaries or comparisons cannot be made, then the statistical approach is not possible. The detail of the way an individual company works, or of an individual human being, may be too subtle to be captured in a form which can be processed statistically. Forcing it into a statistical straightjacket may result in a very shallow level of analysis devoid of any real insight. On the other hand it is possible to take this argument too far; the surprisingly common assertion that no aspects of human subjectivity can be captured statistically is clearly dubious given the size of the attitude scaling industry.

As another illustration, consider the first regression model (Model 1) described by Dissanaike (1999). This model is based on a sample of share prices of large companies over a number of years - it is firmly based on statistical data. The model provides a prediction of the return which investors would receive from investing in one of the securities for a period of four years, from the return they would have received if they had invested in it in the previous four years. The regression coefficient is -0.112, which means that, on average, a security with a level of returns 10% above the mean for the last four years, would produce an expected return 1.12% below the mean over the next four years. On the other hand, if the returns over the previous four years were below the mean, the expected return over the next four years would be slightly above the mean. Needless to say, these are averages over a large number of companies and time periods; the R^2 value quoted (0.0413) suggests that this prediction is extremely unreliable. However, the negative regression coefficient does show that there is a very weak tendency for stocks which have done well over the last four years to do badly over the next four years, and vice versa. This provides support for the hypothesis of "investor overreaction".

What it does not show, of course, is that the fortunes of every stock will change, or that every investor overreacts. It just demonstrates a very slight tendency for the overreactors to outweigh the underreactors and the non-reactors, but gives no further clues about the underlying causes. It is an empirical assertion, but, in many respects, rather remote from reality, although of obvious interest to investors.

It is also a rather fragile conclusion because it depends on the stability of investor psychology. If, say, this overreaction hypothesis became well known, and substantial numbers of investors start to buy stocks which have done badly in the past, the price of these stocks will be driven up and the statistical pattern would no longer be valid. The first of the two presuppositions above may not then be justified.

The obvious approach for exploring the reasons for investor overreaction, and the reasons why some investors do not overreact, would be to seek out some typical investors, and to study some of their decisions, and the detailed reasoning behind them. Such anecdotes may lead to further insights, which may be worth testing statistically.

Statistical Inference

There is a further level of statistical analysis, described by statisticians as statistical inference. This is the process of attaching probabilistic measures of certainty or uncertainty to statistical conclusions. How sure can we be that the results are reliable? Would we get the same answer if we tried again with a different sample? These questions are, of course, important.

Statistical inference is a complex, controversial and confusing area. There is no one approach to which all statisticians would subscribe. The main division is between the Bayesian school and the classical school, but each has its subdivisions (see Sibson, 2000 for a *very* brief overview). In management, the approach which has cornered the market is null hypothesis significance testing - despite very strong arguments in favour of confidence intervals as a substitute (Gardner and Altman, 1986; Kirk, 1996; Wood, 2003) - so the discussion here will initially focus on this.

A null hypothesis is a hypothesis of no difference or no relationship, set up specifically to try to demonstrate that it is wrong. The statistical analysis then estimates the probability of obtaining results "like" the actual data, on the assumption that this null hypothesis is true. This is the significance level (*p* value); conventionally values of less than 5% are taken as indicating that the null hypothesis is untenable, and, by implication, the alternative hypothesis is supported.

The null hypothesis is a fiction. It is one possible world. The researcher hopes it will turn out to be false, thus suggesting that another possible world is closer to the real world. Statistical inference is about choosing the most plausible of a number of possible worlds - most of which are, inevitably, fictional. One of the defining characteristics of statistics is that these possible worlds need to be defined in explicit detail, so that they can be used as the basis for calculating probabilities.

In practice, null hypothesis significance tests are frequently misunderstood and misinterpreted (Gardner and Altman, 1986; Kirk, 1996; Lindsay, 1995; Morrison and Henkel, 1970; Wood, 2003), and are often used inappropriately. Part of the reason for this may be difficulties in getting to grips with the idea that the argument depends on this fictional world. To take one illustration (more or less at random), McGoldrick and Greenland (1992) found that the mean rating for "helpful/friendly staff" of a sample of bank customers was 6.495 (on a 1 to 9 scale), whereas the equivalent figure from a sample of building society customers was 6.978. The significance level cited for this is 0.000 - ie it is highly significant. The fact that the size of the difference (less than half a point on a 1 to 9 scale) is trivial is not mentioned, but the implication is that because it is statistically significant it must be important - which, of course, is not so.

The null hypothesis under test here is the assumption that the mean rating from *all* bank customers is *exactly* the same as the mean rating from *all* building society customers. The statistics assumes that there are potentially an infinite number of possible customers of each type of institution, and the mean ratings from each of these infinite groups will be identical. In practice, with a finite sample, there is likely to be some difference due to sampling error; the statistical test takes account of this by flagging if the difference observed is too large to be accounted for by sampling error. This has happened here: the significance level (0.000) indicates that there is a real difference.

The sample size here (not given in the article) is obviously very large because this is the only way such a small difference could be flagged as significant. When interpreting significance tests it is important to remember that *any difference, however small, can be picked up as significant by taking a sufficiently large sample.*

The difficulty is that small differences are often of no real interest. In this case, the inevitable differences in the clientele of the two types of institutions mean that it is extremely unlikely that the two mean ratings would be *exactly* the same, even if the quality of service was identical. To be worth testing, null hypotheses have to have some plausibility: this one does not for this reason. And on a practical level, even if small differences did exist, they would not be important in this context.

This means that the significance test is a waste of time if correctly interpreted, and misleading if viewed as indicating an important difference. Perhaps statisticians need to deploy some of the skills of the novelist, and tell the stories of these null hypotheses more vividly so that they are recognised for the silly tales they often are?

This is just one example, but such misconceptions are very common. Failure to distinguish between statistical significance and real importance, and the testing of silly hypotheses, are depressingly common faults, even in research published in reputable journals. Another example is a paper in an accounting journal which included a test of the null hypothesis that respondents to a question were equally likely to answer Yes, No or Don't know. There is no reason to expect this hypothesis to be true, and every reason to expect it to be false, so there is no point in subjecting it to a formal statistical test.

The regression model in Dissanaike (1999) discussed above is also backed up by a significance level: the value cited being 5%, indicating that the null hypothesis of no relationship between past and future returns is not very plausible. Again, the size of the effect found (see discussion above) is small, but small effects *are* of potential interest to investors in large portfolios of shares, so this null hypothesis is of some interest.

An alternative approach to both analyses would be to concentrate on measuring the size of the difference, or the regression coefficient, and indicate the likely extent of sampling error by means of confidence intervals (Gardner and Altman, 1986; Wood, 2003); this is common practice in medical studies, but is rarely done in management. For example, a 95% confidence interval for the difference between the banks' and the building societies' mean rating might be cited as 0.3 to 0.7 points on the 1 to 9 scale: this is more straightforward to interpret, and provides more information than the p value. This approach enables researchers, and readers of research reports, to compare the *size* of different effects, instead of simply being told the strength of the evidence for the *existence* of the effects.

Problems of Understanding how to Use and Interpret Statistics

It is conventional to separate the discussion of the logic and value of research methods and concepts, from the discussion of the educational problem of ensuring that they are adequately understood. In the case of statistical methods, inappropriate use and interpretation are so widespread that there is little point in assuming that researchers know what they are doing, and readers know how to interpret results. We have seen some of the problems of null hypothesis testing, and raised the possibility that other approaches may be more sensible. Even simple statistics, such as averages, proportions and correlations are often misleading for the reasons described by Huff (1973). In management research, a major difficulty is sampling: non-response bias, for example, may mean that results from many surveys should be treated with much more caution than they normally are.

The other side of this coin is that research based on anecdote or fiction may lack prestige, but the fact that it is so much more user-friendly may mean its impact is greater just because people can understand it easily and it "sticks" in the mind. As well as being a research tool, case studies - true and fictional ones - are widely used for educational purposes. (The distinction between their use for educational and research purposes may, on occasions, be blurred.)

In many circumstances (but by no means all) statistical concepts and methods are vital, so there is a serious educational problem here. The conventional approach is to increase the amount and the effectiveness of training in statistics (on which there is an enormous literature and several specialist journals), but there is an argument that trying to make the subject matter more "palatable" may be a more effective strategy (Wood, 2002). Part of the problem is that statistics is a mathematical discipline and many people feel an antipathy towards mathematics. It is possible to develop statistics in a way which is not dependent on any mathematics more advanced than arithmetic by using non-parametric concepts and simulation methods (Wood, 2003), although the concepts involved are by no means trivial.

THE STEREOTYPING OF METHODOLOGY

Discussions of management research tend to distinguish two broad styles of research. For example, Easterby-Smith et al (2002) contrast positivism and social constructionism; although they point out this is a stereotype, it is a stereotype which dominates many researchers' thinking. In the positivist corner (Easterby-Smith et al, 2002: 30), we have "statistical probability" and concepts that "can be measured", implying a quantitative approach; these are both absent in the list of features of social constructionism. Hence, to adopt a positivist approach, statistical techniques and quantitative methods are necessary. Alternative terms for the first (positivist) style of research are "scientific" or "quantitative", and the second style is sometimes referred to as "qualitative" or “phenomenological”.

It is obvious that terms such as "scientific" and "quantitative", and "social constructionist" and "qualitative", do not mean the same thing. But their use reveals the stereotype of two types of researcher. Either you are a positivist researcher and you use statistics and numbers and do not use qualitative data and do not believe that reality is socially constructed, or you are in the other qualitative camp and you use neither statistics nor numbers.

At first sight, the three categories explored in this paper seem to reinforce and extend this stereotype. Positivist research is based on statistics, qualitative research is based on anecdotes, and fiction corresponds to another category not normally dignified by the title research - the idea that insights gained from novels and TV plays are of value for understanding management. However, this simple correspondence is far from accurate.

In practice, qualitative researchers often study a sample of cases and arrive at generalisations such as "most X are Y". This is a *statistical* conclusion. The features studied may be subtle, and the research may require detailed investigation of individual cases, and the conclusion may be, from a statistical viewpoint, very simple, but it is still a statistical conclusion in that it concerns how frequently different things happen. Qualitative research is typically partly based on anecdotes, and partly based on statistics. As Miles and Huberman (1994: 40) put it in their book on *Qualitative data analysis* "... numbers and words are *both* needed if we are to understand the world".

This means that some of the methods and precautions of statistics may be appropriate for some qualitative research. Statistical conclusions are largely meaningless if based on inappropriate samples, and if the samples are appropriate but small, confidence intervals or hypothesis tests may to a useful way of assessing the extent of sampling error, and answering questions about whether other samples would yield similar results.

We have seen above that natural science - the supposed source of the idea of positivism and using scientific method in management research - makes use of fiction and anecdotes, and that many of its conclusions are not at all statistical. Science typically involves carefully contrived experiments to test hypotheses or measure parameters. The application of statistical methods to random samples of people or events is not part of the standard approach. Statistics may have a role on the fringes - analysing errors for example - but it is not central to the way most science works. Science is about imagining possibilities, and searching out extreme or otherwise interesting circumstances to test ideas. Statistics is only necessary when there is no powerful theory; when all we can do is try to find out what will *probably* happen, or what will happen on *average*. It is the last resort, not the pinnacle of the scientific method. There is thus no justification for the often automatic identification of science with statistics.

Similarly, quantitative results in management often do stem from statistical data, but this is not always the case (see the discussion of the economic order quantities above). "Quantitative" and "statistical" have different meanings.

The categories presented here - anecdote, fiction and statistics - cut across most of these stereotypes. Any crude stereotyping of types of research is unhelpful, and may restrict the variety of approaches researchers use. The example in the next section shows how a typical research project needs to use anecdote, fiction *and* statistics.

AN EXAMPLE

I have a colleague, a professor of management, who is currently engaged in a project for a large public sector organization. She is looking at a key service process in this organization: currently there are many delays, errors and anomalies in this process, and her task is to identify these, assess their extent and help the organization to improve the system from the point of view of the main stakeholders. Unfortunately, I am not at liberty to divulge the name of the organisation or the business it is in, so I have fictionalised the case study and transported it into another, very different business, in line with the third of the reasons for using fiction above. The story I will use is from a university setting: the process by which assignments for students are set and marked. This is a much more mundane process than the real process under investigation, but the two processes are sufficiently similar to enable me to use the university story to explain how the professor's research uses anecdote, fiction and statistics.

The (fictional) process involves lecturers setting the assignment, then the students completing the task and handing it in by the deadline, after which the completed assignments are marked and feedback comments written by the academics, some are second marked and sent to the external examiner as a check on standards of marking, and finally the marks are recorded in the database and the scripts, feedback comments and marks are returned to students. The process is beset with problems: assignments get lost, marks are wrongly entered, the second marking fails to happen, delays are frequent with students sometimes waiting months for work to be returned. (The professor's job is to look at the administrative processes: it is not part of her brief to look into the reliability or validity of the marks awarded - which some would say is an even greater problem.)

The detail of the case and the professor's findings are not relevant here. What is relevant is the way she has approached the reality of this situation. Her initial phase was a phase of qualitative data gathering - talking to key stakeholders and collecting *anecdotes* about problems, good practices and so. She also asked many of her interviewees to draw flowcharts of the process as they saw it. No attempt was made to gather statistical data, but simply to catalogue the variety of ways in which the system worked and failed to work. Much of the data gathered in this phase illustrated the importance of recognising that different stakeholders had very different interests, and that simplistic models which ignored this were unlikely to be helpful.

One anecdote which sticks in my mind from the professor's account to me concerns the production of a quality manual by the academic heading the quality committee at the university. He had spent months researching and writing this manual which spells out the procedures to be followed in processing assignments, but the professor failed to find a person who had even opened it. This raised many obvious questions. The professor's search for anecdotes such as this was not haphazard: she needed to be systematic in her selection of stakeholders to interview to reduce the danger of missing important points.

She then went on to explore some fictions. These were of two types. The first involved imagining alternative ways of running the system and then running a what-if analysis to see the impact it would have - usually in the imagination, but occasionally using a spreadsheet. One suggestion, for example, was that only a limited sample of assignments - chosen at random - should be marked. This had clear advantages in terms of the workload for academic and administrative staff, and equally clear problems in terms of the validity of the final marks and incentives for students. These factors needed to be balanced against each other: to some extent this was possible in the imagination, but the research team decided that a small-scale trial (or experiment), analysed statistically, was necessary to check whether the idea would work as expected.

The second type of fiction involved imagining how the process - the actual one or an imagined alternative - would react to unusual situations - perhaps students trying to manipulate the marking system. There are obvious advantages in being able play these scenarios out in the imagination. Both types of fiction required creative input from people familiar with the system: several brainstorming sessions were run for this purpose.

The professor's final level of analysis was the statistical level. There is obviously no guarantee that things that seem to work in the imagination will work reliably and consistently in practice. Similarly, anecdotes can reveal possibilities such as documentation remaining unread, but cannot give any reliable information about how widespread this problem is. Statistical analysis of data from carefully designed samples was necessary to ascertain the overall pattern and to establish levels of confidence in conclusions. The professor used a modified version of the SERVQUAL instrument (Zeithaml et al, 1990) for much of this work.

This research project illustrates the importance of the three approaches to reality which are the subject of this article. Each had a vital role to play. Anecdotes alert us to possibilities which may be important and may teach us how a situation works in some detail. Without such an anecdotal base we may miss much of importance. Fictions allow us to consider new possibilities and so are necessary for genuine innovations. History may be necessary for improving the future, but it needs to be spiced up with some imagination if we are avoid repeating the mistakes of the past. And statistics are necessary to ensure that we have a grasp of the pattern in the whole population and are not being misled by memorable incidents or stories which are not common in the real world.

This case study itself is midway between anecdote and fiction: most of the details above have analogues in the real case, and the real researchers did use a modified version of SERVQUAL. On the other hand, it is just a single example, and cannot be used to support claims about how frequently anecdote, fiction and statistics should be used in this kind of management research.

CONCLUSIONS

Management researchers can approach reality in three ways: by looking at true stories, by making up fictional stories, and by the statistical analysis of samples of data. I have tried to demonstrate that these modes are far more interdependent than they may appear and that many common assumptions are misguided. To summarise some of the points made above:

* Anecdotes are useful for demonstrating that something is possible, analyzing how it is possible, and for bringing general ideas to life so they are more likely to "stick".
* Fiction can also achieve these ends. On occasions, it may have the advantage over anecdote of being more flexible, easier to derive - especially where sensitivities and confidentialities make factual anecdotes impossible to obtain - and for communicating via the "gut". Fictions are also indispensable for investigating possibilities for the future that have not yet happened. And mathematical optimisation methods, computer simulations, and statistical inferences also depend on a comparison of a variety of fictional and real scenarios.
* Statistical concepts and methods are important to see the general pattern in a whole population or process. Anecdote and fiction are obviously inappropriate for this purpose. However, it is important to remember that statistical approaches are only part of the toolkit of researchers - even, perhaps especially, those who describe their approach as scientific. Unfortunately statistical methods are notorious for being difficult to use and interpret in a meaningful manner - which is obviously vital if their value and limitations are to be properly appreciated.

All three approaches have a role to play in a typical research project such as the one described in the previous section. Research restricted to one of these three approaches is likely to be seriously impoverished: we typically need all three. In practice, however, some researchers tend to favour statistical research, others focus on anecdotes from case studies, and others take the view that the art of story telling is most likely to lead to illumination. Each group may dismiss the other approaches out of hand, and the first two may join in dismissing fiction because it lacks an empirical basis. This, however, ignores the fact that it may not be helpful to focus exclusively on what is, and ignore what might be, or what should be.

Each approach - anecdote, fiction and statistics - requires particular skills. The skills required by statistics are obvious (although often in short supply). Anecdotes require the skills of qualitative research and also an appreciation of the dangers of drawing anecdotes from a limited sample. Fiction requires creativity: the ability to imagine alternative realities - which may be stories in the ordinary sense, innovative new production systems, or examples to illustrate a potential problem with an innovative new system.

References

Akin, G. (2000, March). Learning about work from Joe Cool. Journal of Management Inquiry, 9(1), 57-61.

Bernstein, J. Einstein. London: Fontana; 1973.

Carey, J. (ed). (1999). The Faber Book of Utopias. London: Faber and Faber.

Christy, R., & Wood, M. (1999). Researching possibilities in marketing. Qualitative Market Research, 2(3), 189-196.

Dissanaike, G. (1999). Long term stock price reversals in the UK: evidence from regression tests. British Accounting Review, 31, 373-385.

Draper, D., Hodges, J. S., Mallows, C. L., & Pregibon, D. (1993). Exchangeability and data analysis. J R Statist Soc A, 156(1), 9- 37.

Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). Management Research: an introduction. London: Sage.

Gardner, M., & Altman, D. G. (1986, March 15). Confidence intervals rather than P values: estimation rather than hypothesis testing. British Medical Journal, 292, 746-750.

Hawking, S. W. (1988). A brief history of time: from the big bang to black holes. London: Bantam Press.

Huff, D. (1973). How to lie with statistics. Harmondsworth: Penguin.

Kim, W. C., & Mauborgne, R. A. (1992, July). Parables of leadership. Harvard Business Review, 123-128.

Kirk, R. E. (1996). Practical significance: a concept whose time has come. Educational and Psychological Measurement, 56(5), 746- 759.

Lindsay, R. M. (1995). Reconsidering the status of tests of significance: an alternative criterion of adequacy. Accounting, Organizations and Society, 20(1), 35-53.

McGoldrick, P. M., & Greenland, S. J. (1992). Competition between banks and building societies. British Journal of Management, 3, 169-172.

Mackay, D. (2000, August 28). Richardson vitriolic ... : The Observer.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis (2nd edition). London: Sage.

Morrison, D. E., & Henkel, R. E. (eds) (1970). The significance test controversy. London: Butterworths.

Newton, J., Graham, J., McLoughlin, K., Moore, A. Receptivity to change in a general medical practice. British Journal of Management. 2003; 14: 143-153.

Popper, K. R. (1980). The logic of scientific discovery. London: Hutchinson.

Sibson, R. (2000). Entry on statistics in Bullock, A., & Trombley, S. The new Fontana dictionary of modern thought. London: HarperCollins.

Taylor, S. S. (2000). Aesthetic knowledge in academia: capitalist pigs at the Academy of Management. Journal of Management Inquiry, 9(3), 304-328.

Wood, M., & Christy, R. (1999). Sampling for possibilities. Quality & Quantity, 33, 185-202.

Wood, M., & Christy, R. (2001). Prospecting research: knowing when to stop. Marketing Letters, 12(4), 299-313.

Wood, M. (2002). Maths should not be hard: the case for making academic knowledge more palatable. Higher Education Review, 34(3), 3-19.

Wood, M. (2003). Making sense of statistics: a non-mathematical approach. Basingstoke: Palgrave.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). Delivering service quality: balancing customer perceptions and expectations. New York: Free Press.