

The Rationale of Research

This is Chapter 8 from *The Design of Research Studies – A Statistical Perspective*¹

John Maindonald,
Statistical Consulting Unit of the Graduate School
Australian National University.

john.maindonald@anu.edu.au

Contents

1. Balancing Scientific Scepticism with Openness to New Ideas	2
2 Data and Theory	4
3 Models	5
4 Regularities (Law-Like Behaviour).....	5
5 Statistical Regularities	6
6 Imaginative Insight.....	8
7 Science as Hypothesis Testing	9
8 Strategies for Managing Complexity.....	9
9 Cause and Effect.....	11
10 Computer Modelling	11
11 Science as a Human Activity.....	12
12 The Study of Human Nature and Abilities	15
References and Further Reading	17

¹ Look for document GS00/2 on the web page:
<http://www.anu.edu.au/graduate/pubs/occasional-papers/>

The aim of science is to seek the simplest explanation of complex facts . . . seek simplicity and distrust it.

[A. N. Whitehead]

Both scepticism and wonder are skills that need honing and practice. Their harmonious marriage within the mind of every schoolchild ought to be a principal goal of public education.

[Sagan 1997, p. 289.]

Any adequate account of the scientific method must allow for the exercise of imaginative insight. It must also place checks on the unconstrained use of the imagination. There must be a mechanism for distinguishing claims that can be substantiated from claims that cannot be substantiated.

It must allow a role both for data and for theory. Any collection of data pre-supposes some notion that these particular data are likely to be interesting and useful. In this sense, science is driven by theory. It is the genius of science that data may challenge and even destroy the theory that guided their collection. This is the means by which science places a check on unbridled exercise of the imagination.

Theory works with models. Our special interest is in statistical models. A good model captures those aspects of a phenomenon that are relevant for the purpose in hand. A model is, inevitably, an incomplete account of the phenomenon. The reward for simplifying by ignoring what is irrelevant for present purposes is that the model is tractable – we can use it to make predictions.

I use the word *science* in a broad sense, not much different from the word *knowledge*. Scientific research is directed to gaining new knowledge.

1. Balancing Scientific Scepticism with Openness to New Ideas

The methods of science stand in strong contrast to belief systems — religious systems, cults of every description, popular prejudices, political ideologies of both the left and right, those claiming magical or other powers of healing, the claims of much commercial advertising, faith healers, promoters of new therapies who resist the rigours of scientific testing, and so on. Scientific claims are open, at least in principle, to rigorous objective testing. Admittedly, science does not in practice always live up to these high ideals.

There is a strong contrast with systems of ideas that resist rigorous testing. These systems readily generate, or more often rehash, ideas that are away from current mainstreams of scientific knowledge. They have rarely shown much interest in rigorous testing. They typically spurn scientific standards, even as an ideal. Standards of evidence are weak.

Theory is a fruitful source of ideas. Ideas may come from methodically working through the implications of current theory. There may be a bold and imaginative extension or adaptation of existing theory. Or the challenge may come from a new theory that questions existing notions. Whatever their source, ideas should never have an automatic claim to credence. They must stand on their merits. There must be reality checks at key points along the way — does it happen as claimed? Occasionally a theoretical insight may seem so compelling that there is no need to check further. Previously inexplicable facts now make perfect sense. Even here one has to proceed with caution, keeping in mind our capacity for mistake and self-deception, and our proneness to jump to conclusions. Scepticism, directed at current assumptions as well as at any new theory, must be the order of the day. There are many case-

histories that demonstrate the need for caution. An example is the claimed link between salt and hypertension that we discussed in Section 3.1.

There are by contrast well-known instances where the scientific community refused to take seriously, on the grounds that there was no mechanism, an idea that had strong empirical support. Or important and significant results may be dismissed out of hand. The examples that follow illustrate, in turn, these two possibilities.

Continental drift

My discussion pretty much follows the account the very readable account in Hallam (1989). Wegener (1880-1930) presented a range of evidence in support of his theory that the present continental land masses had formed from the splitting apart of older continental masses. He pointed out that the Western coast of Europe and Africa fits fairly well the contours of the Eastern seaboard of the Americas. He argued that former land bridges between continents explained important features of the present distribution of fauna and flora. But geologists had a long tradition of mechanistic explanation. Prominent and influential figures denounced Wegener's ideas, creating an intellectual climate where any young and bold spirit who took up these ideas thereby placed their career at risk.

Biologists were more sympathetic. They had rarely been lucky enough to find detailed mechanisms for the phenomena that they studied, and were more willing to live with the idea that an understanding of mechanisms would have to come later. At the same time, they respected the prevailing judgment of geologists that such splitting and moving of land masses was impossible. The opposition to Wegener's ideas remained strong through into the 1950s. The highly respected geophysicist and mathematician Harold Jeffreys (1891-1989) was especially vocal in his opposition to Wegener's ideas.

A further impossible hypothesis has often been associated with hypotheses of continental drift and with other geological hypotheses based on the earth as devoid of strength. . . . In Wegener's theory, for instance . . . the assumption that the earth can be deformed indefinitely by small forces, provided only that they act long enough, is therefore a very dangerous one, and liable to lead to serious error.
[Jeffreys 1926, p.261]

A group of younger researchers who revived Wegener's ideas, still without much idea of the mechanism involved, thereby risked their careers. One of those younger researchers – Edward Irving – took a position at the Australian National University. Australia provided, at that time, more fertile ground for his ideas. Far from leading geologists into serious error, the theory has been the point of departure for huge advances in the understanding of earth history. It is a cornerstone in a unified framework for the interpretation of data from biogeography, geophysics and geology.

Clues to the Functioning of the Immune System

The *bursa of Fabricius* is a small sac at the tail end of the digestive tract in birds. In the 1950s two graduate students, Glick and Chang, discovered that this organ has a vital role in the production of antibodies. Glick, who had been unable to find any effect from the removal of the organ, gave his chickens to Chang for a class demonstration of the production of antibodies. The demonstration failed, a result of the surgical removal of the bursa while the chickens were still very young. A paper that described their finding was rejected by the journal *Science* as “uninteresting”. It finally appeared in the journal *Poultry Science*, where it went unnoticed for many years. After it did finally come to attention, it became in due course the most quoted paper ever to appear in that journal (Clark 1995, p.42.) It marked the beginning of fundamental discoveries regarding the immune system.

There are many reasons why a good idea may be slow to gain acceptance. The forces of conservatism can act just as strongly in scientific communities as in other communities. The

word of one dominating and influential figure may be enough to prevent a hearing. “How dare you challenge my authority?” While it is the force of the argument that should prevail, not the pronouncements of elder statesmen, this may not be what happens.

2 Data and Theory

Science is different from many another human enterprise – not of course in its practitioners’ being influenced by the culture they grew up in, nor in sometimes being right and sometimes wrong (which are common to every human activity), but in its passion for framing testable hypotheses, in its search for definitive experiments that confirm or deny ideas, in the vigour of its substantive debate, and in its willingness to abandon ideas that have been found wanting. If we were not aware of our own limitations, if we were not seeking further data, if we were unwilling to perform controlled experiments, if we did not respect the evidence, we would have very little leverage in our search for truth.

[Sagan 1997, *The Demon-Haunted World*, p. 252. Headline Book Publishing, London.]

Data

Data are crucial to science. Up until the 20th century a prevailing view was that science was generalisation from data. The name given to this process of generalisation is *induction*, which contrasts with *deduction* as used in mathematics and logic.

The view of science that emphasised induction and generalisation from data was strongly influenced by Francis Bacon, who in 1620 published a book that argued for a new method of research that, as he claimed, gave ‘True Directions Concerning the Interpretation of Nature’. In Bacon’s ‘improved’ plan of discovery, laws were to be derived from collections of observations. (Silverman 1985.)

Theory

Scientists do not collect any old data. They collect the data that seem most useful. How do they get this sense that some data will be helpful, and other data of little use? For example a study of the effects of passive smoking is likely to look for specific effects, most likely effects that are known to be a result of active smoking. One would not expect to find that passive smokers have an unusually high number of ingrown toenails! So we will not waste effort on gathering data on ingrown toenails. We will examine the occurrence of lung cancer, bronchitis, heart disease, and so on, but not ingrown toenails. There’s no theory to suggest that smoking of any kind might cause ingrown toenails.

For studying the health of children living in some area of New Guinea, one might collect data on age, sex, height and weight. Hair colour and eye colour are unlikely to be of interest, for this purpose. It seems obvious that height and weight are important indicators, but that hair and eye colour are unlikely to be relevant. It is assumed that some measures are useful and some are not. There is an extensive literature that provides guidance on what measures other workers have found useful, which sets out “theory” that anyone who now undertakes collection of data on the health status of one or other human group will want to note². Those who initiated work in this area had to make their own judgments on measures that seemed useful indicators of health status.

Any adequate understanding of science must have regard both to theory and to data. Researchers do not collect any data. Data collection is driven by a judgement of what is worth collecting. It is in this sense that theory drives scientific research. None of the great scientists have followed Bacon’s prescription. Typically they showed unusual insight, aided sometimes by good fortune, in the data that they collected.

² See for example chapters 7 and 8 in Little and Haas (1989).

Data may carry within themselves the power to challenge and perhaps destroy the theory that guided their collection. It is this that gives science its power. Statistical insights and approaches have a key role both in data collection and the extraction of information from data. They assist in the efficient choice of data, in teasing out pattern from the data, and in distinguishing genuine pattern from random variation. The pattern may be as simple as a difference between the means of two treatment groups, or a linear relationship between two variables.

This is a convenient place to introduce the idea of a ‘model’. This is an important idea, both in science generally and in statistics.

3 Models

Consider the formula for the distance that a falling object, starting at rest above the earth’s surface, moves under gravity in some stated time. The formula is:

$$d = \frac{1}{2}gt^2$$

where t is the time in seconds, g (≈ 9.8 m/sec/sec) is the acceleration due to gravity, and d is the distance in metres. Thus a freely falling object will fall 4.9 meters in the first second, 19.6 meters in the first two seconds, and so on. This formula describes the way that objects fall.

Observing the fall of a stone (especially if you happen to be underneath) is a different experience from encountering the formula on a piece of paper. There are important aspects of the fall about which the formula tells us nothing. It gives no indication of the likely damage if the stone were to strike one’s foot. The formula can tell us only about the distance traversed in a given time, and other information that we can deduce from distance information.

Watching the stone fall and making measurements is different from doing calculations using the formula. The results will not be quite identical, if only because of the limits of accuracy of the measurements. The formula is a model, not the real thing. It is not totally accurate – it neglects the effects of air resistance. For the limited purpose of giving information about distance fallen it is, though, a pretty good formula. As Clarke (1968) says: “Models and hypotheses succeed in simplifying complex situations by ignoring information outside their frame and by accurate generalization within it.”

A good model captures those aspects of a phenomenon that are relevant for the purpose in hand. A model is, inevitably, an incomplete account of the phenomenon. The reward for simplifying by ignoring what is irrelevant for present purposes is that the model is tractable – we can use it to make predictions.

There are also non-mathematical models. An engineer may build a scale model of a bridge or a building that is to be constructed. Medical researchers may speak of using some aspect of mouse physiology as a model for human physiology. The hope is that results from experiments in the mouse will give a good idea of what to expect in humans. As those who know the history of such research understand all too well, animal medical models can be misleading. At best, they provide clues that must be tested out in direct investigation with human subjects.

The model captures important features of the object that it represents, enough features to be useful for the purpose in hand. An engineer can use a scale model of a building to show its visual appearance. The scale model might be useful for checking the routing of the plumbing. The model will be almost useless for assessing the acoustics of seminar rooms that are included in the building.

4 Regularities (Law-Like Behaviour)

Mathematical models describe law-like behaviour, i.e. one can use the model to describe or predict. The falling object formula predicts distances.

We take a variety of regularities for granted in our everyday lives. We expect that the sun will rise in the morning and set in the evening. We expect that fire will burn us, and so on. These expectations have nothing to do with logic. They are based on our experience of the world. We take such regularities for granted.

There is no logical reason why what has happened in the past will continue to happen in the future. There is no logical reason why the sun should continue to rise. Fortunately for humans, it does! Indeed, it is impossible to carry on our lives unless we do take such regularities for granted. We speak of law-like behaviour. The process by which we generalise from our experience of the world to rules that tell us what will happen in the future is called induction. Inductive science looks for regularities in phenomena.

The natural sciences look for very wide regularities. They have found a huge range of phenomena, many of them outside of the range of our everyday experience, that exhibit law-like behaviour. There has been more limited success in finding law-like regularities in the biological sciences. In the social sciences there has been very limited success in finding law-like behaviour.

The nature of the social sciences makes law-like behaviour hard to find. The phenomena are more complicated. Consider the complicated processes that are at work to make some people criminals, and some law-abiding citizens. The relatively simple falling object equation is a striking contrast with our incomplete understanding of the 'forces' that work to make some people criminals. Typically there are many effects at work. It is impossible to do experiments or make observations that separate these effects out individually. The processes are almost certainly different for different individuals. While it is possible to say that children who suffer severe neglect or abuse are much more likely to become criminals, this is just one of many different factors that are at work. We cannot explain why criminal behaviour is a much greater problem in some societies than in others.

5 Statistical Regularities

Statistical regularities rely on probabilistic forms of description that have wide application over all areas of science. In studying how buildings respond to a demolition charge, there will be variation from one occasion to another, even for identical buildings and identically placed charges. There will be variation in which parts of the building break first, in what parts remain intact, and in the distance and direction of movement of fragments.

Deterministic models, i.e. models that do not use probabilistic or statistical forms of description, have a place, especially in the physical sciences. Statistical variability is often so small that it can be ignored. In the natural sciences however, statistical variation is ubiquitous and statistical forms of description are generally essential. No two animals or plants or humans are identical.

Statistical models typically have at least two components. One component describes deterministic law-like behaviour. In engineering terms, that is the *signal*. The other component is *noise*, i.e. statistical variation. Here is an example. Different weights of roller are rolled over different parts of a lawn, and the depression noted³. What we find is:

³ Data are from Stewart, K.M., Van Toor, R.F., Crosbie, S.F. 1988. Control of grass grub (Coleoptera: Scarabaeidae) with rollers of different design. N.Z. Journal of Experimental Agriculture 16: 141-150.

	Weight (t)	Depression (mm)	Depression/Weight
1	1.9	2	1.1
2	3.1	1	0.3
3	3.3	5	1.5
4	4.8	5	1.0
5	5.3	20	3.8
6	6.1	20	3.3
7	6.4	23	3.6
8	7.6	10	1.3
9	9.8	30	3.1
10	12.4	25	2.0

Table 3: Depression, and Depression/Weight Ratio, for different weights of lawn roller.

We might expect that depression would be proportional to roller weight. That is the signal part. The values for Depression/Weight make it clear that this is not the whole story. Rather, we have

$$\text{Depression} = b \times \text{Weight} + \text{Noise}$$

Here b is a constant, which we do not know but can try to estimate. The Noise is different for each different part of the lawn. If there were no noise, all the points would lie exactly on a line, and we would know the line exactly. In Fig. 4 the points clearly do not lie on a line. We therefore explain deviations from the line as random “noise”, at least until some more insightful explanation becomes available.

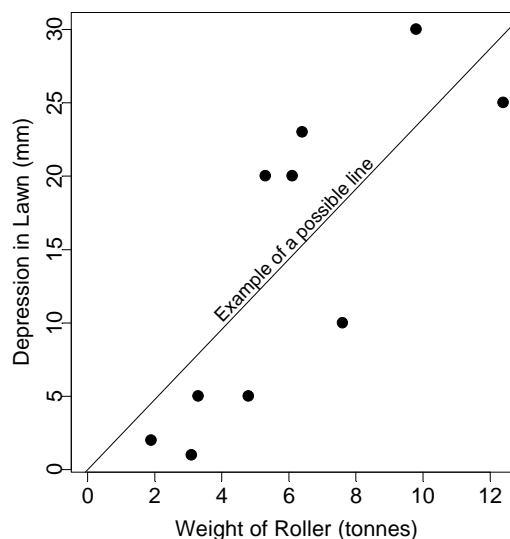


Fig. 4: Lawn Depression, for Various Weights of Roller, showing one possible line. The line is one of many that are consistent with the data.

We need a model for the noise also. We’ll leave the details till later. Anyone who has done a first year course in statistics will expect to hear words such as *normal* and *independently*

distributed used to describe the noise components. For now, let's call it a random term without spelling out the details.

It is a feature of statistical models that they have a signal component and a noise component. In some data the signal is strong and the noise small. In other data noise may dominate the signal. Fig. 5 illustrates the range of possibilities:

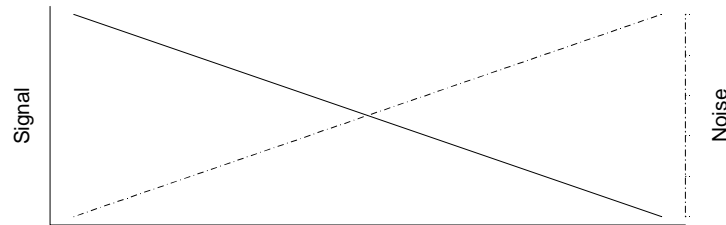


Fig. 5: Different positions along the horizontal axis correspond to different mixes of signal and noise. At the left extreme, there is only signal, while at the right extreme there is nothing except noise. Statistical models lie somewhere between these extremes.

We would prefer to get rid of the noise altogether. That is not a totally silly idea. While we cannot get rid of the noise altogether, we may be able to reduce it. There are several ways in which we might be able to do this:

1. By using more accurate measuring equipment.
2. By improving the design of the data collection.

A skilled experimenter will get as near as is reasonably possible to the extreme left in Fig. 5. That is where every experimenter would like to be.

Question: In the lawn roller experiment, how might one reduce the noise, i.e. reduce the scatter about the line or other response curve?

6 Imaginative Insight

How do radically new theories arise? No doubt generalisation from data, i.e. induction, has a role. At most it can be only part of the explanation. There is a large element of imaginative insight – the recognition that looking at the phenomena in some new way will perhaps simplify the description, or explain former anomalies. Trying to understand imaginative insight may not be much different from investigating the psychology of scientists.

There are however styles of investigation that provide fruitful ground for the exercise of imaginative insight, and styles that are likely to confuse and derail it. Thus a carefully conducted experiment usually provides much better raw material for the exercise of imaginative insight than does unsystematic experimentation and poor design. In the former case anything that is unusual or unexpected will stand out as different and demand investigation, while in the latter case unexpectedly large or small values may have a multiplicity of explanations.

An apple transport trial in which I participated (Maindonald 1986) illustrates how careful design helps highlight anomalous results. The trial had sufficient elements of careful design that those few crates where there was heavy bruising stood out as anomalous. We found that they were unstable, shearing first to one side and then to the other as the truck negotiated bends in the road. Our design had neglected what turned out to be the most important factor affecting apple bruising. Nonetheless, because we had controlled for other factors such as the condition of the apples, the effects of bin instability stood out clearly.

7 Science as Hypothesis Testing

... in learning by experience ... conclusions are always provisional and in the nature of progress reports, interpreting and embodying the evidence so far accrued.
[R. A. Fisher]

Imaginative insight readily creates worlds of its own that may have little connection with reality. There is a place for imaginative drama, fiction, legend and myth, but not as part of science. So there must be severe checks on the exercise of imaginative insight. How do we keep imaginative insight in check, ensuring that what we claim to find is real rather than the product of a fertile imagination. Why should we believe scientific explanations for patterns in the frost, rather than the claim that “the fairies did it”? The difference, according to Karl Popper, is that genuinely scientific theories can be tested. Instead of starting with data, Popper starts with a theory. Popper has little to say on where scientific theories come from.

There must be a motivation for collecting data. There must be a sense that some data are worth collecting and some are not. Researchers who are unclear why they are collecting data, and are not selective about what data they collect, typically end up with data that are of little use. Effective researchers are highly selective about the data they collect. They seek data that will address the questions that are of interest to them.

Any legitimate scientific theory will make predictions. For example, Newton’s gravitational theory predicts that the earth and other planets will move around the sun in elliptical orbits. This prediction seems to be born out by the observed facts. So Newton’s theory survives that particular test⁴.

A scientific theory will not be rejected just because it cannot explain particular observations or results from a particular experiment. Kuhn (1970) argues that for a new theory to replace an old theory two conditions must be satisfied

There must be serious cracks in the old theory, i.e. important facts that the old theory does not explain.

A new theory must be available.

Why replace a theory, even one that has evident flaws, unless something better is available with which to replace it?

There are further issues:

When observations or an experiment give results that are contrary to a well-established theory, is it the theory or the experiment that is mistaken? There may have been a flaw in the experimental procedure.

Flaws in experimental procedure are especially common when one is working at the limits of experimental technology. It may be at these limits that theory has its most extreme test.

Often, a small modification to the theory may be enough to accommodate a newly discovered anomaly.

Scientists may be so deeply wedded to the existing theory that they refuse to accept the new theory. This is particularly likely if the new theory is itself incomplete, i.e. many of the theoretical details have not been worked out. There are many examples of this.

8 Strategies for Managing Complexity

Complex systems defy ready understanding. Easily the most successful scientific strategy has been to restrict attention to limited aspects of a system where simple models may work. Once

⁴ It almost survives it. Later work found small anomalies in the orbit of the planet Mercury. Einstein’s theory of relativity is required to give a completely accurate description of the orbit of Mercury.

the subsystems are well enough understood, the hope is that it will be possible to bring the separate pieces of information together to give a useful account of the total system.

This reductionist approach has been spectacularly successful in physical science, biology and medicine. As Wilson (1998, p.58) says, "Reductionism is the search strategy used to find points of entry into otherwise impenetrably complex systems." In the end however, the aim is to describe and explain the rich complexity of the systems under investigations. There is no virtue in naïve simplicity unless it leads, finally, to insights that enable us to get a handle on the complexity.

In practical applications of science, this complexity may extend far beyond the specific issues that motivated the scientific study. As an example of this complexity, consider the salinity that has affected or is threatening huge areas of Australian farmland. There are a large number of scientific issues that bear on this problem, some of which I list below. However none of the studies that one might conduct under these individual headings will, on their own, give the information needed to address the problem. Somehow the information from these various sources must be brought together.

An Example – The Desertification of Australian Land

Over large areas of Australia the destruction of forests has removed the trees that formerly soaked up water in the soil, leading to a rise in the water table. Salts are naturally present in the soil, in some places in substantial quantities. Irrigation brings in further dissolved minerals. These remain after the water has evaporated and build up slowly, adding to what is already in the soil. As long as the water table is well below the surface, rain will wash any salts down into the ground water, where they are not a problem. Once the water table rises to close to ground level, it brings the salts with it. Trees that have been left standing, and other vegetation, die off. In the end, the land becomes unusable. Coram (1998) quotes an estimate of 120,000 hectares of land in New South Wales that was affected by dryland salinity in 1996, with a further 5 million acres considered to be at risk.

There are many individual components to any study of this salinity problem.

1. Extent of the problem: What is the present and expected future size of the land areas that are affected?
2. Vegetation Effects: What is the extent of continuing damage from new clearing of vegetation? What is the potential remediation role of new tree plantings? Is it possible to find tree species that will grow and survive in saline soil?
3. Irrigation practices: How much of the problem is the result of past and current irrigation practices? How might changes in irrigation practices assist remediation? How effective (and cost-effective) would it be to use bores to replace the use of water from irrigation channels?
4. Groundwater draining and pumping: Is draining and/or pumping of groundwater a viable potential remediation strategy in some areas? Which areas?
5. Engineering of irrigation channels: What effects (e.g. damage to adjacent roads from the build-up of salt in the soil and/or from waterlogging) arise from loss of water from irrigation channels? What engineering solutions (e.g. better lining of channels) are available?
6. Land use strategies: What changes in patterns of land use might assist remediation. The replacement of agriculture by forestry can be highly effective. Those crops are preferable that do not require heavy irrigation.
7. Flow-on effects: How much of the problem in one or another area is the result of practices in other areas, perhaps more elevated or perhaps upstream?
8. Ecology: What are the effects on fauna and flora? How would alternative remediation strategies affect fauna and flora?

9. Social issues: What steps will ensure that remediation measures do not unduly disadvantage individual communities?

Also open to scientific study are political and economic consequences, flowing both from the present degradation of land and from proposed remedies.

There must be strategies for gathering whatever information is needed under each of these headings, and for creating from them an integrated plan of understanding and action.

Questions worth considering are:

1. Are there changes that would be easy and cheap, and that would make substantial inroads on the problem?
2. What changes, ignoring for the moment their costs, would make the largest inroads?

Questions: Why is it hard to get action on the degradation of Australian land that is a result of salinity? Are there no good strategies? Or is the problem an inability to implement the strategies that are available? Is the needed co-operative action too difficult for our society's political and economic structures?

9 Cause and Effect

It is one thing to establish a correlation between two variables. It is another to establish a causal link. The direction of causation is sometimes obvious. It is rain that causes the wheat to grow, not growth of wheat that causes the rain. Heavy drinking causes the subsequent hangover. But what is the relationship between hard work and business success. Does success come first, leading people to work hard to maintain and improve their position? Or does hard work come first. Often, both variables are driven by a third variable. Weight and height are strongly correlated, but it makes no sense to claim that one causes the other. These issues have generated fierce continuing debate in the social science literature. References in Freedman (1999, p.248) represent a range of perspectives. See also pp.78-80 of Greenhalgh (1997).

Cause and effect issues have appeared at several points earlier in these notes. Does salt in the diet cause high blood pressure? Does an increase in the minimum wage cause reduced employment? What long-term effects flow from sudden and unexpected traumatic loss?

Claims of causation are convincing when there is a cogent theory that establishes the causal chains of connection. Where the theory is complex, built from many individual components, those components must be open to testing. Complex theories must often rely on computer modelling to link the separate components. One example is the extensive body of theory that is designed to predict the global climatic impacts of human activity. Some might argue that it is a complex of theories rather than a single theory. This is a matter of definition.

10 Computer Modelling

Many of the new biological challenges are of the "how do we put the pieces back together" type. Those problems are horrendously difficult for our current approaches.
[Wilson, 1998, pp.91-92.]

Human impacts on climate change are a serious issue for our time. For science it is a huge problem of the "how do we put the pieces back together type". Many different sources of information and evidence must come together. Computer modelling seems the only viable approach.

Increased atmospheric levels of carbon dioxide and other implicated "greenhouse" gases⁵ increase the effectiveness of the earth's atmosphere as a heat shield. Much of the focus has been on increases in carbon dioxide levels that have resulted from increased use of fossil

⁵ Other gases that are implicated are methane, nitrous oxide and hydrofluorocarbons.

fuels. A 0.5°C average global increase in the temperature of the earth over the past century seems in part due to this and other human activities. Schneider (1996) reports an assessment of tree-ring and other evidence for temperature change in the past ten thousand years that suggests that such a large 100-year change has been unusual over this time, occurring no more than once in a thousand years. See also Crowley (2000).

Projections drawn up by the Intergovernmental Panel on Climate Change predict an average global warming of between 1.0°C and 3.5°C over the next century, a greater rate of climate change than at any earlier time in the past 10,000 years. Predictions are that sea levels will rise, some low-lying areas will be covered by sea, there will be loss of vegetation, farmers may need to change to new crops that are viable in the new climatic conditions, weather patterns will be less stable, and tropical diseases will affect many sub-tropical regions.

How were these figures obtained? It is not sensible to try to project current temperature trends into the future. The world's climate has changed continuously over time, making short-term trends a poor guide to what may happen in the future. Rather the evidence comes from computer modelling, including modelling of the effect of projected ongoing emissions of greenhouse gases in the atmosphere. The predictions from this modelling are unequivocal – present rates of release of CO₂ into the earth's atmosphere will lead to a temperature increase. If these rates continue to increase at about 1.5% per annum as in the recent past, the temperature increase over the next 100 years will be correspondingly larger.

Atmospheric and ocean currents are the moving parts of a huge engine that is driven by the sun's heat. The blanketing effect of the atmosphere, itself affected by life processes on land and in the sea and by human activities that include the use of fossil fuel, are a part of the engine's control mechanisms. Understanding of the functioning of the individual components seems adequate for the building of computer models that make gross predictions, always assuming that ocean (and air) currents continue to follow pretty much their current patterns of movement. A worrying aspect of potential large temperature changes is that they may cause the engine to reconfigure itself. Changes in the flow of major ocean currents, such as have happened in past geological times, would bring changes in climate patterns that would be even more traumatic.

Computer models must accommodate, as best they can, all these different effects. Statistical methodology has a clear role in checking the predictions of individual components against experimental and observational data. Checks that model predictions over several years for different regions of the earth's surface agree with observation are encouraging, but not clinching evidence. By the time that clinching evidence of the accuracy of model predictions is available, the damage will be irreversible. Hence the importance of close critical scrutiny of the separate components of the models, of the way that those components are linked and of sensitivity analyses that check how predictions would change if there were changes to those model assumptions that are open to challenge.

Scientists from many different disciplinary backgrounds have critically scrutinised the computer models. There has been extensive refinement of the details. Qualitative model predictions have withstood these criticisms remarkably well. The most persistent criticism has come from those with a political axe to grind, usually in defence of inaction! Such critics have the option, and the challenge, to build and offer for scientific scrutiny models that give predictions that are more to their taste.

11 Science as a Human Activity

I know that most men, including those most at ease with problems of the greatest complexity, can seldom accept even the simplest and most obvious truth if it be such as would oblige them to admit the falsity of conclusions which they have delighted in explaining to colleagues, which they have proudly taught to others, and which they have woven, thread by thread, into the fabric of their lives.

[Tolstoy, quoted in Gleick, 1988.]

[Scientific theories] . . . are constructed specifically to be blown apart if proved wrong, and if so destined, the sooner the better. “Make your mistakes quickly” is a rule in the practice of science. I grant that scientists often fall in love with their own constructions. I know, I have. They may spend a lifetime vainly trying to shore them up. A few squander their prestige and academic political capital in the effort. In that case – as the economist Paul Samuelson once quipped – funeral by funeral, theory advances.
[Wilson, E.O., 1998, p.56]

Humans are not inherently rational creatures. Much of what passes for reasoned argument is rationalisation – the use of reason to defend positions that we hold for other reasons. An attitude of mind that judiciously balances openness to new ideas with rigorous critical scrutiny does not come easily to our human nature. Prejudice readily takes precedence over the demands of rationality. Scientists are not inherently different from other humans who are prey to idiosyncratic belief systems and spurious claims. Gilovich (1991) is one of many books devoted to the discussion of our irrational foibles.

Fallible Scientists

Scientists are not immune from the tendency to rationalise. Thus craniology – the measurement of the brain capacity, often with the aim of relating brain capacity to racial differences – became a popular subject of study in the nineteenth century. Not surprisingly, much of this work collected and used data in ways that reflected the racial and sexual prejudices of the scientists who undertook it. Gould (1996), in a highly readable book, discusses this and other similar examples. Fortunately the processes of scientific criticism and re-evaluation do in the course of time tend to expose and correct such abuse. (Gould’s book has itself attracted accusations of bias from academic critics.)

Still today, rationalisation and prejudice compromise science. New prejudices and new rationalisations have arisen to replace those that we hoped to have conquered. Such rationalisations find it especially easy to establish and retain a foothold in those areas where there is a dearth of external checks on the exercise of imaginative reconstruction. Dogma easily masquerades as science.

Researchers may become more concerned about maintaining their funding or their position within the profession than about truth. Science easily degenerates, in some times and some corners, into pseudo-science. There is self-deception, there is an often exaggerated deference to authority, there is deliberate manipulation, and there is a yielding to self-interest. There is a challenge to devise ways of funding and directing scientific research that reduce opportunity for manipulation, for deviousness, and for prejudice and dogma that masquerade as science.

Different scientists have different qualities. Some may be receptive to new ideas, but not good at criticism. Others may be good at criticism, but not receptive to new ideas. They may apply high standards of criticism in their own area, but make idiosyncratic judgments when the scientific demands change. They may be hypercritical, not understanding the different nature of the evidence that the new and unfamiliar area demands. Or, failing to note the different opportunities for self-deception that this new area offers, they may be unduly credulous. There are few who can examine claims in medicine or social science or physics with more or less equal critical incisiveness.

Dominant authorities

As in all communities, there are some whose pronouncements carry especial weight, or whose positions give them special authority. They may be editors of major journals, or have a large influence in the decisions of funding agencies. There are practical reasons for listening to the voices of such dominant figures. Their judgments can be effective in weeding out ideas that are not worth pursuing. At the same time they may weed too ruthlessly, their own speculative notions may acquire the force of dogma, and they may resist anything that they find too

novel. This may be a particular danger if there are just one or two dominant figures — individuals who occupy the sort of position that Harold Jeffreys occupied in geophysics in the 1950s. It is healthier if the dominant figures do not altogether agree among themselves.

Jealousy and backbiting also flourish. Other scientists may be seen, not as partners in a common endeavour, but as threats to one's own enterprise who must be cut down by any means available. Political concerns may influence scientific judgements. Even if such attitudes are not overt, they may lurk below the surface. Perhaps we should be surprised that the demands for scientific rationality do so often prevail over these human influences. Only an overarching insistence on rigorous criticism can keep science from becoming prey to irrationality. There will never be total success. There is however plenty of scope for improvement on the way that science is now conducted.

The Logic of Science and the Sociology of Scientific Communities

Above I noted conditions that, according to Kuhn, must be satisfied before a new theory can replace an existing theory. There must be serious cracks in the existing theory, and a new theory must be available.

However Kuhn goes further. He argues that science is driven by powerful social forces, akin to those that drive other human activities. An objective examination of the history of science shows much that confirms Kuhn's claim. A weakness in Kuhn's account is that he does not maintain a clear distinction between the logic of scientific discovery and the sociology of scientific communities⁶. Science has an inherent logic that does often, in the course of time, prevail against the sociological forces that drive one or another scientific community. At least in the physical and biological sciences, it is unusual for reactionary attitudes to hold back progress for more than a decade or two. Individuals who show unusual insight may be denied their PhDs. Their ideas, if they withstand critical scrutiny, do however finally prevail. This is a remarkable feature of scientific discovery. A science that was wholly the product of social forces would be ineffective.

The sociology of scientific communities often works against really good science. I will criticise unhelpful practices, in data collection, in data analysis and in the reporting of results, that are undesirable outgrowths of the sociology of particular scientific communities. My complaint is that they are contrary to the inherent logic of science. Some common failings are:

- uncritical reliance on expert opinion
- exaggerated expectations of what can be learned from observational data
- failure to marry subject area insights with results from statistical analysis
- deficiencies in data-based overview
- unwillingness to bring in other skills when these are clearly required
- deference to pressures from commercial interests.

Reductionist Scientists?

Scientists who wish to publish extensively and advance in their chosen research area will do well to limit their attention to a narrow range of problems that seem likely to yield easily to their skills. This narrowness of focus, which can be beneficial in making initial progress in a closely defined area of research, does not give the breadth of view needed to tackle “big issue” questions. Determining the structure of an organic chemical compound found in the river water, or using radio-isotopes to trace its progress through the river system, does not of

⁶ For a recent wide-ranging critique of Kuhn's views, see Fuller (2000).

itself give the breadth of view needed to tackle such “big picture” problems as dry land salinity.

Wilson (1998, p.40) has apt comments:

The vast majority of scientists have never been more than journeymen prospectors. That is even more the case today. . . . They acquire the training needed to travel to the frontier and make discoveries of their own, and as fast as possible, because life at the growing edge is expensive and chancy. The most productive scientists, installed in million-dollar laboratories, have no time to think about the big picture and see little profit in it.

The skills of a “journeymen prospector” may serve well those who expect to join multi-million dollar research laboratories. A narrow training focus seems clearly inappropriate for anyone whose work is likely to demand skills different from those of their Ph.D. or other research degree, or who is likely at some time to work on “big picture” issues.

Commercial Pressures

Money speaks volumes. Commercial pressures may be a potent influence. Wilkinson (1998) offers a series of case studies that highlight some of the issues. Edmeades (2000) is an interesting study of the aftermath to a celebrated defamation claim that occupied the New Zealand High Court for 135 days. What were the rights and duties of fertiliser scientists who wished to make the results of their research available to the farming community that they had a responsibility to serve?

The Uses of Controversy

Controversy is not of itself bad, it may help fire enthusiasm in the scientific community and in the public at large. The many-talented biologist Thomas Huxley (1825-1895) used his combative nature and his penchant for controversy to great effect. He was a great populariser of science as well as a great scientist (Desmond 1994.) The down side was that his penchant for controversy too often got out of hand, making enemies unnecessarily.

Controversy can be helpful in drawing attention to areas of weakness in the science. It offers an interesting window both into the sociology of scientists and into the logic of scientific discovery. It is an advantage when the different parties to the controversy come from different disciplines, and accordingly offer different perspectives. Novice researchers sometimes find themselves caught, uncomfortably, between the different sides of a controversy. From time to time the views of a PhD examiner will, in spite of care in the choice of supervisors and examiners, be seriously at odds with the ideas and insights that shaped a smaller or larger part of the thesis. It is with these points in mind that I now comment on controversies that have surrounded the study of human abilities and human nature.

12 The Study of Human Nature and Abilities

Know then thyself, presume not God to scan,
The proper study of mankind is man.
[Alexander Pope (1688-1744): An Essay on Man.]

The scientific study of human nature and abilities is a sensitive area, for all sorts of reasons. Are humans able to pursue such studies objectively, with the detachment that science demands? Supposed scientific objectivity readily becomes a vehicle for particular prejudices.

The Heritability of IQ

Studies of the genetic basis of IQ have had a long and tangled history. A key and greatly overplayed concept has been the *heritability* coefficient, the proportion of variation (measured

using the statistical variance) that is due to genetic variation. The heritability coefficient has been widely used in animal and plant breeding studies, where the outcome variable of interest has been weight or milk production. A high heritability suggests a potential to get further improvements from breeding. Comparison between heritability coefficients from different trials makes sense only if environmental variation is comparable. This may be reasonable if, as in many animal and plant breeding studies, conditions are similar across different trials.

Studies of twins, both identical and non-identical and including separated pairs, have been the main source of evidence for the heritability of IQ in human populations. As one might expect, the two members of a separated pair are often reared in very similar circumstances, more similar than for two randomly chosen members of the population. Thus the studies tell us nothing about heritability in a section of the population where the range of social disadvantage is large. Lewontin (1979) has argued, rightly in my view, that

. . . there is no way in human populations to break the correlation between genetic similarity and environmental similarity, except by randomised adoptions.

One would need to randomly assign adoptees to the whole range of social circumstances to which it was intended to generalise results. Such an experiment is surely out of the question.

There is a further issue. Twins share a common maternal environment. Daniels et al. (1997), in a meta-analysis of more than 200 studies, estimate that the shared maternal environment of twins accounts for 20% of the total variance. The ignoring of this component in earlier analyses of data from twin-adoption IQ studies led to a substantial over-estimate of heritability. Assigning to the wrong source a component that turns out to be 20% of the total is perhaps excusable in the initial tentative investigations. Long before one has the 212 sets of results that Daniels et al. analysed, this surely has acquired the status of a fundamental biological error! This analysis still leaves large questions unanswered. What is the relevance of these studies, if any, to a wider population where the range of environmental effects may be far larger than those typically experienced by the separated twins?

IQ tests capture a small part of the rich texture of human abilities. Mental and other abilities continue to change and develop through into old age. *Mind Sculpture* (Robertson 1999) is the title of a book that discusses evidence on how our brains develop and change as a result of demands placed on them. The emphasis should perhaps move from the study of mental testing to the study of *mind sculpture*.

Sociobiology

In his 1975 book *Sociobiology: The New Synthesis*, Wilson defined sociobiology⁷ as “the systematic study of the biological basis of all social behaviour”. Wilson hoped to find a genetic basis for behaviour. Sustained controversy followed its publication. While most of the book was devoted to the study of animal and especially insect societies, the final chapter speculated on genetic influences on human behaviour. Why all the fuss? The discussion that now follows draws at several points on the account in Segerstråle (2000).

Any initial foray into an area that is as complex as genetic effects on animal behaviour must over-simplify. But what if the simplifications that seem required are precisely those that readily feed into racial, sexual, national and other such forms of prejudice? Wilson was aware of the risks of the area into which he had ventured, and took care to protect his words from such misuse. His critics were not satisfied, either with his science or with the care that he had exercised. Criticisms were of several different types:

- Wilson was charged with specific scientific errors.

⁷ Note also the more recent term *evolutionary psychology*, used to describe an area of study that has a large overlap with sociobiology.

- Notwithstanding the generally liberal tenor of Wilson's views, it was argued that they lent support to those opposed to steps that would ameliorate the position of socially and economically disadvantaged groups.
- Criticism of Wilson's book became a convenient starting point for promoting wider scientific and political agendas. In some instances statements were taken out of context, charging Wilson with views that were at variance with specific statements in the surrounding text.

There is a succinct statement of the criticisms in Rose et al. (1984). Segerstråle attempts to disentangle the various strands of this controversy. It is worth noting that a wide spectrum of political views is found both among those who emphasise genetic influences on human behaviour and abilities, and among those who emphasise environmental effects.

The first tentative steps in a new area of study may use overly simplistic models, which will be refined as understanding advances. Problems arise when there are perceived implications for the way that we regard or treat fellow humans. There is a long history of misusing claimed scientific results that is the theme of Gould's *The Mismeasure of Man*⁸. Where such implications are perceived, it behoves scientists to tread with extreme care, to acknowledge obvious limitations in their models, and to acknowledge the tentative character of their results. This may conflict with the motivation that researchers feel to persuade themselves and others of the importance and significance of their work.

A useful outcome of the sociobiology controversies has been a closer scrutiny of the scientific methodology than has been common in other areas of biology that rely extensively on observational data. This scrutiny needs to go further. Such statistical methodologies as regression are too often used uncritically, without regard to traps such as were discussed in section 5.2. Even if the models are correct, estimates of key parameters may be wrong.

References and Further Reading

- Box, Joan Fisher 1978. Fisher — the Life of a Scientist. Wiley, New York.
- Clark, W.R. 1995. At War Within. The Double-Edged Sword of Immunity. Oxford University Press, Oxford, UK.
- Clarke, D. 1968. Analytical Archaeology. Cambridge.
- Coram, Jane (ed.) 1998. National classification of catchments for land and river salinity control. Rural Industries Research and Development Corporation (Australia), no. 98/78.
- Crowley, T.J. 2000. Causes of climate change over the past 1000 years. *Science* 289: 270-277.
- Daniels, M., Devlin, B., and Roeder, K. 1997. Of genes and IQ. Chapter 3 of Devlin, B., Fienberg, S.E. and Roeder, K., eds., *Intelligence, Genes and Success*. Springer, New York.
- Desmond, A., paperback edition 1999. Huxley. From Devil's Disciple to Evolution's High Priest. Perseus Books, Reading MA.
- Diamond, J. M. 1997. *Guns, Germs, and Steel: the Fates of Human Societies*. Random House, London.
- Edmeades, D.C. 2000. Science Friction. The Maxicrop Case and the Aftermath. Fertiliser Information Services Ltd., P.O. Box 9147, Hamilton, N.Z.
- Fuller, S. 2000. Thomas Kuhn: A Philosophical History for Our Times. University of Chicago Press.
- Gilovich, T. 1991. How we know what isn't so. The Free Press, New York.

⁸ Gould's account has itself attracted strong criticism from some academic reviewers.

- Gleick, J. 1987. *Chaos: making a new science*. Viking, New York.
- Gould, S. J., revised and expanded edition, 1996. *The Mismeasure of Man*. Penguin Books.
- Greenhalgh, T. 1997. *How to Read a Paper: the basics of evidence-based medicine*. BMJ Publishing Group, London.
- Hallam, A., 2nd. edn 1989. *Great Geological Controversies*. Oxford University Press.
- Harré, R. 1967. The principles of scientific thinking. In Harré, R., ed.: *The Sciences. Their Origin and Methods*, pp. 142-174. Blackie and Son Ltd., Glasgow.
- Jeffreys, H. 1926. *The Earth, its Origin, History and Physical Constitution*. Cambridge University Press.
- Kuhn, T., 2nd edn, 1970. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.
- Lewontin, R.C. 1979. Sociobiology as an adaptionist program. *Behavioural Science* 24: 5-14.
- Little, M.A. and Haas, J.D., eds. 1989. *Human Population Biology. A Transdisciplinary Science*. Oxford University Press.
- Maindonald, J. H. 1986. Apple transport in wooden bins. *New Zealand Journal of Technology* 2: 171-176.
- Robertson, I. H. 1999. *Mind Sculpture*. Bantam, London.
- Sagan, C. 1997. *The Demon-Haunted World. Science as a Candle in the Dark*. Headline Book Publishing, London.
- Schneider, S.H. 1996. *Laboratory Earth. The Planetary Gamble We Can't Afford to Lose*. Weidenfeld and Nicholson, London.
- Segerstråle, U. 2000. *Defenders of the Truth. The Battle for Science in the Sociology Debate and Beyond*. Oxford University Press, Oxford.
- Silverman, W. A. 1985. *Human Experimentation. A Guided Step into the Unknown*. Oxford University Press, Oxford.
- Taubes, G. 1998. The (political) science of salt. *Science* 281: 898-907 (14 August).
- Wilkinson, T. 1998. *Science Under Siege: The Politicians' War on Nature and Truth*. Johnson Press, Boulder CO.
- Wilson, E.O. 1975. *Sociobiology: The New Synthesis*. Harvard University Press, Cambridge MA.
- Wilson, E.O. 1998. *Consilience. The Unity of Knowledge*. Abacus, London.