JUDGEMENT, POLICY AND THE HARMONY MACHINE

John Fox

Imperial Cancer Research Fund Laboratories, London

"Men cannot be reconciled by appealing to their rational minds, only by appealing to their hearts" Robert Owen, quoted on a plaque at New Harmony, Illinois.

Science and technology confer their benefits through our understanding of the world, and the consequent ability to choose actions which limit its unpleasant aspects or enhance its pleasant ones. The natural sciences and their technological offspring have produced methods and devices which, on balance, have made the world more predictable, more controllable and more comfortable.

Broadly speaking these features of predictability and controllability are greatest in our dealings with the physical world. In biological systems they are, as yet, less evident. In social systems and personal affairs uncertainty is dominant.

The successes of the natural sciences and the growing successes of the life sciences have encouraged many people to ask whether we can achieve comparable levels of control over social events, and even our private lives, through the use of rational, scientific techniques. Liberal, societies traditionally leave these matters to human judgement. It is my judgement whether or not I should seek a medical treatment, or private education for my children. It is at the discretion of company executives to pursue new markets or develop new products. It is the judgements of governments and their officers whether policies should be established which turn on certain economic indices, or which introduce industrial changes with social or environmental implications, or which determine rules of citizenship.

Most people view judgement as something akin to art and not science. A contemporary view is that it is amenable to articulate, rational analysis which could lead to better decisions and, consequently, greater personal contentment and less conflict between individuals and groups.

THE ART OF JUDGEMENT

When faced with a bald challenge we may defend individual and professional judgement, notably our own, but it is not reliable. Judgement is no better than the information it is based upon, or the mind that formulates it. We cannot always rely upon our own judgement to make the right decisions in our lives, or upon the collective judgement of others to make the right decisions for our society. We make mistakes and it is only the tolerance of others and the flexibility of social institutions that prevents many of our mistakes being costly.

It is natural therefore to use tools, such as computers, to collate and refine information prior to making judgements. Some argue that we should go further and rely upon rational calculation in preference to judgement. The last thirty years or so have seen the development of quantitative techniques in psychology, sociology, ecology and economics whose forms are reminiscent of those employed in the natural sciences. Behavioural scientists argue that human decision making and personal judgements are all too frequently 'irrational' by comparison with the prescriptions of scientific decision theory, and that we would be wise to evaluate our otions mathematically. Proponents of nuclear power attack opponents with bundles of calculations of 'risk'. Operations research techniques are used to calculate the manpower needs of companies and to plan complex industrial or military projects. The influence of economic models on national policies is well known.

Clearly we can use information more effectively to make personal and social judgements, and we should explore the use of information technology to do that. However the dominance of quantitative methods may be mistaken. Some of my complaints are technical objections but I shall not go into these much here. Rather I want to suggest that qualitative methods, such as those pioneered in AI, could radically change our understanding of, and competence in, the processes of social judgement.

THE LIMITS OF MATHEMATICS

One way or another most 'rational' techniques for dealing with personal, ecological, economic and political questions depend upon mathematical probability theory and statistics. The technical attraction of probabilistic methods, unlike the traditional deterministic techniques of physics, is that they do not try to squeeze uncertainty out of complexity, only to get the measure of it. Using statistics we can make predictions in the face of uncertainty, and provide a rational basis for individual decisions and social policies. We can calculate our 'expected utility' of a surgical operation, the 'risk' of ecological disaster, the 'inflationary pressure' of a public spending programme. Strong advocates of such methods believe that rational judgement must be founded on rational mathematics, and that its widespread use must lead to greater individual satisfaction, industrial efficiency and social harmony. Critics consider them naive.

There are many tacit critics. Although mathematical tools are available for dealing with individual and social questions most of us don't use them at all. The techniques can be effective, so why don't we use them? I suggest four related reasons. I shall leave aside the fact that most of us don't know how to use them. If that were all there were to it then presumably most of us would be happy to employ specialist consultants if it seemed advantageous. The reasons for the doubts about mathematics that I prefer have to do with what people understand, and what they believe and do as a consequence.

The first observation is that mathematics is abstract. It ignores, quite deliberately, the meaning of the events it deals with and concentrates on the formal structure of problems. It ignores the idiosyncrasies of the situation and its participants, and it ignores details of organisational or social contexts which can exert powerful influences on the very interpretation of a problem, as well as on the form of an acceptable solution. The risk of a nuclear disaster may be calculably remote, but that is irrelevant if my concern is with the competence of engineering contractors, the standard of routine maintenance, or the possibility that plutonium may fall into the wrong hands.

Second, mathematics is unintelligible to most of us. This is not just because we lack some education, but more important it is a consequence of its abstractness. The statistics on smoking are conclusive, but the effect on smokers has been limited. The Treasury's model of the economy might be as good a predictive device as humanity can develop but citizens can't see how, or if, the model reflects the rise or decline of economic activities that affect them. We just don't understand models that depend upon the interaction of fifty parameters. In effect we have no basis for assessing the personal implications of such calculations, and they are inevitably disregarded in individuals' personal, political and economic behaviour.

Thirdly, and consequent upon this lack of intelligibility, many of us cannot trust those who define 'rational social policies' even if we should. If we accept the mathematics, without understanding, we accept the power of technical elites to decide matters that affect us without being accountable to us. Disastrous medical advice can easily be dismissed as "the luck of the draw". The effects of a bad regulation or economic policy "could not have been foreseen". Who can check? Claims that professional ethics or public accountability provide proper controls are just seen as special pleading and status quoism. Human judgement may be rough and ready but it has the huge political advantage that it can be examined, and challenged, in the public arena or a court of law because we share a common language for discussing it. In a political sense the most rational attitude is to regard an obscure technical argument as simply a hostage to fortune.

Finally there is the ancient problem of personal values and interests. I believe that the development of much of mathematics and the natural sciences would have arrived more or less as it is if the beliefs, politics and values of the discoverers had been different. We might have had a different selection of discoveries, but not a contradictory selection. This is certainly not true of personal judgement or social policy. Efforts to achieve harmony must recognise differences of value. It is here that mathematical techniques have been at their weakest. To be sure there are •objective[1] quantitative notions of personal utility and economic rationality but they are, in the view of many, unreliable and unconvincing.

I conclude that certain aspects of human judgement are flawed, but the mathematical tools which may claim to correct or assist judgement are too abstract to substitute for it entirely, too unintelligible for us to know what they do and do not address, and too neutral to be confident that they protect our interests rather than those of others.

THE CONTRIBUTION OF AI

So what does all this have to do with Artificial Intelligence? AI claims to offer a radically new framework for understanding the manifestations of intelligence. Judgement is one such manifestation. Could AI offer a different sort of technology for making judgements that is more compatible with human understanding? Could it let us build a new generation of harmony machines without the vices of the old?

The most prominent practical development in Artificial Intelligence so far has been the introduction of expert systems. Expert systems, it is said, use 'knowledge' to give assistance in specialised problem solving and decision making. Many people see knowledge based systems as providing new capabilities for making decisions, interpreting information, planning and designing, and even making scientific and commercial innovations. Some see a role for knowledge based systems in the formulation of law and social legislation.

The technical capabilities of expert systems are probably only a little ahead, and in some ways behind, the capabilities of classical mathematical systems. The importance of knowledge based systems, however, is not their current achievements but in the way that they solve problems and some side effects of the techniques they use rather than their current capabilities. These features might address the the problems that I have outlined.

Expert systems emphasise knowledge, not numbers. AI workers have an idea of what knowledge is, or at least a partial one. The information that an expert system uses is primarily qualitative, including •facts' such as:

```
hopelessness is a cause of
                    social alienation
```

and •rules* like:

```
if Client is unemployed
   and period of unemployment
                of Client is long
   and opportunity
                of employment is low
then risk of social alienation
             of Client is high
```

(The syntax of these fragments is that of the PROPS package developed at the Imperial Cancer Research Fund. Capitalised terms are variables.)

There are several consequences of representing knowledge in this way.

First, the emphasis on qualitative facts and rules of thumb expresses fairly directly what we know, or at least what we think we know. Qualitative statements are imprecise, but they reflect ordinary thinking. Precision is often an illusion or irrelevant anyway. Brian Gaines has a nice comment that there is little point in saying that something "will be delivered at gate no 5 at 10.00 on Saturday morning" when all you mean is it "will be dropped off round the back over the weekend". The apparent precision of calculation may merely give an air of rationality without its substance.

Extensions of such ideas let us represent the meaning of the concepts referred to in the rules and facts. As more and more rules and facts about "unemployment", "opportunity", "social

alienation" are added the computer becomes more and more able to use the concepts in ways which are isomorphic with the ways in which we use them. If the concepts are complex and varied, then the computer's representation of them is complex and varied. The ability to represent the details and idiosyncrasies of the problem is greater than if we limit ourselves to formal idealised models.

Dreyfus, Weizenbaum and Searle, deny that this is "true meaning" in a human sense, but even if this is correct, which many question, it is an observation which may have little force. If the computer behaves in such a way that people can understand and even predict then the practical consequences are that its actions are intelligible.

This intelligibility is a pivotal point for the present argument, as well as expert systems generally. We may consider these fragments of knowledge, and assert that they are simplistic, inconsequential or just plain wrong. Quite possibly, but little or no training in computer science is required to understand, and therefore challenge, the judgements they embody. Although they are in effect fragments of a computer program, they are intelligible fragments that can be examined and debated.

A side effect of representing knowledge in this way is that the computer system becomes accountable to those it affects. It is well known that one of the features of expert systems is that they can give explanations. If I want to know how a conclusion or recommendation is arrived at I can ask. The computer must report the facts that it assumed and the line of reasoning it followed. Admittedly current techniques of explanation are primitive but they will improve - and my experience is that an expert system's clumsy attempts at explanation are more understandable than many legal documents that I encounter.

Knowledge based systems will be more credibly competent, or openly laughable, than their predecessors. The potential for argument, challenge and the exercise of individual discretion are thereby increased, and the commissioners, designers and operators of the machinery of policy become more accountable. Interestingly, the habit of explanation could be catching. Refusal of insurance cover; imposition of zoning regulations; taxation demands; denial of promotion or citizenship; public statements of changes to fiscal policy, or announcements of public works, would increasingly be expected to be accompanied by intelligible documents of explanation. A future Freedom of Information Act might insist upon it.

Finally, how do we analyse individual values? We can't, at least not entirely. AI does not solve all the deep problems of philosophy. However even if our understanding of such matters is sketchy, a knowledge based system could still allow for (if not fully comprehend) individual values or attitudes. To give just one example we might imagine a home computer asking "which is more important to you, having your baby at home near your family and friends, or within reach of trained staff in case of problems?" and reflecting the answer in its advice.

I think this all boils down to the possibility that, contrary to many expectations, AI could be a liberating force. The influence of technology and the mechanical handling of information are growing at a rapid rate. Many of these influences are hidden by virtue of their incomprehensibility. AI, properly managed, could lead to needed checks and balances in a technology based society, and more participation by its members in the formation of policies.

CONCLUSIONS AND CAVEATS

As we are increasingly dominated by complexity and change we need help to predict and control their consequences. One response is to introduce rational mathematical tools. But by themselves these tools are too alien to gain much ground, and where they are used they are liable to improper or unaccountable use. There are areas where technique can enhance judgement and policy making, but technique will only be acceptable if it reflects human understanding and is accountable to human authority. It is worth exploring what AI techniques have to offer.

However the huge growth of interest in AI has not been driven by liberal aspirations, but by commercial ones. Many of us feel that this has unbalanced its development. Most of the technical community is far more interested in the new capabilities, the new efficiencies and the new markets that AI seems to offer than the social benefits. Although the administrators of research programmes and organisations established

to foster the development of AI appear to be sympathetic to the idea that it should be exploited for social as well as economic benefit the response from technical, social science and political groups has been disappointing. I hope that the IJCAI panel will contribute to altering this.

I do not suggest that the necessary AI developments are around the corner or will be easily achieved. It must be said that AI is subject to political direction and management, and its application to socially valuable aims will have to be consciously encouraged. The liberalisation that I think AI could deliver would also be a painful discipline for our masters; they may prefer obscure mathematics, or nothing at all.

ACKNOWLEDGEMENT

I wish to thank Maggie Boden for inviting me to prepare this paper, and to take the opportunity of drawing attention to her own paper in the Journal of Mathematical Sociology, 1984, 9, pp 341- 356, which deals with rather similar themes.