An introduction to prior information derived from probabilistic judgements: Elicitation of knowledge, cognitive bias and herding

Michelle C. Baddeley[1]

Andrew Curtis [2,3]

Rachel A. Wood [2,4]

1. *Faculty of Economics and Politics; Gonville & Caius College, Cambridge CB2 1TA, UK*
(e-mail: mb150@econ.cam.ac.uk)

2. Schlumberger Cambridge Research, High Cross, Madingley Road, Cambridge CB3 0EL, UK.

3. Grant Institute of Earth Science, School of GeoSciences, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JW, UK.

4. Department of Earth Sciences, Cambridge University, Downing Street, Cambridge CB2 3EQ, UK.

Abstract: Opinion of geological experts is often formed despite a paucity of data, and is usually based on prior experience. In such situations humans employ heuristics (rules of thumb) to aid analysis and interpretation of data. As a result, future judgements are bootstrapped from, and hence biased by, both the heuristics employed and prior opinion.

This paper reviews the causes of bias and error inherent in prior information derived from the probabilistic judgements of people. Parallels are developed between the evolution of scientific opinion on one hand, and the limits of rational behaviour on the other. We show that the combination of data paucity and commonly employed heuristics can lead to herding behaviour within groups of experts. Elicitation theory mitigates the effects of such behaviour, but a method to estimate reliable uncertainties on expert judgements remains elusive.

We have also identified several key directions in which future research is likely to lead to methods that reduce such emergent group behaviour, thereby increasing the probability that the stock of common knowledge converges in a stable manner towards facts about the Earth as it really is. These include a) measuring the frequency with which different heuristics tend to be employed by experts within the geosciences; b) develop geoscience-specific methods to reduce biases originating from the use of such heuristics, c) create methods to detect scientific herding behaviour; and d) research how best to reconcile opinions from multiple experts to obtain the best probabilistic description of an unknown, objective reality (in cases where one exists).

Introduction

Geological information is often partially based on personal opinion or judgement. The aim of any good theorist must be to form judgements as rationally as possible, the goal being to coincide with some unobservable, but nonetheless objective reality. For this to be possible, well-reasoned, probabilistic judgements must have the potential to guide the evolution of scientific thought. However, subjectivity inevitably biases opinion. This is particularly true in economics where belief has a causal role i.e., changes in belief can change the reality of the phenomena of interest (e.g., stock market share prices, Kahneman &Tversky 1982). In the geosciences, human belief does not usually affect the underlying data-generating system – Earth processes (although there are exceptions, see Wood & Curtis (2004)). However, accepted or prior opinions of existing experts certainly affect the judgement of others, including future experts in the making. It is therefore desirable to understand typical human biases and errors that may be implicit within experts' opinion-forming, cognitive processes so that their effects can be reduced rather than propagated.

This paper has two purposes: first, it explores the cognitive issues surrounding prior information based on probabilistic judgements, and methods to acquire such information reliably. Second, it describes areas requiring future research in order to reduce bias in judgements within the geosciences. These purposes are considered in parallel throughout the paper.

The first section below explains differences and parallels in terminology and concepts used in different fields of research, in order that results from other disciplines (statistics, economics, psychology) can be understood within the geosciences. There is then an explanation of how subjective beliefs can be analysed using Bayesian analysis. This is followed by a discussion of probabilistic judgements and cognitive bias, both from an individual and group perspective, and explains phenomena such as herding. A discussion of methods to elicit expert opinion in such a way as to reduce the effects of such phenomena is then presented, followed by a particular geoscientific interpretation example in which expert opinion was sought. A summary of potentially fruitful and useful further research directions concludes the paper.

Limits on Quantification

A basic distinction that is common to several frameworks of probability and uncertainty found in different academic disciplines (at least in statistics, geoscience and economics) is that between subjective versus objective probability. This distinction is important for the rest of this paper as we generally discuss subjective probabilities, which describe opinions or beliefs. Consequently we begin by reconciling variations in terminology between these different fields, and describing key results from each field.

Statistical literature makes the distinction between *statistical probability* and *inductive probability* (Carnap 1950; Bulmer 1979). A *statistical probability* is the limiting value of the relative frequency of an event over many trials. Statistical probability is therefore an empirical concept about some objective reality, and can be verified via observation and experiment (Bulmer 1979, p. 4). Statistical probabilities or frequencies are usually associated with some *ex post* calculation and/or a complete knowledge of a data-generating process; they may therefore have little to do with fundamental forms of uncertainty emerging from incomplete knowledge. Classical or frequentist statistical approaches have tended to assume implicitly that probabilities are statistical.

In contrast, *inductive probabilities* describe rational expectations of a future event. They act as a guide to life and are formed even when an anticipated event is unprecedented; they therefore have no necessary association with frequency ratios. In contrast to statistical probabilities, inductive probabilities are to do with *ex ante* prediction; they are formed in the face of uncertainty and incomplete knowledge. In most areas of academic investigation, inductive probabilities are of greater practical importance than statistical probabilities because knowledge of an underlying objective reality is either limited or absent. With incomplete knowledge, statistical probabilities based upon past outcomes and an assumption of stationarity, are often inappropriate to the analysis of expert judgement in complex situations, either in natural scientific (such as in geoscience) or social scientific (such as in economics) contexts.

A similar distinction made in the geosciences is that between statistical probabilities and knowledge-based or conceptual uncertainty (see Pshenichny, this volume). Pshenichny defines conceptual probability as a measure of conceptual uncertainty – uncertainty that arises from incompleteness of knowledge – that is clearly associated with the inductive probabilities described above.

In analysing some of the limitations on quantification of economic probabilities, Keynes (1921) distinguishes between *Knightian risk* (the quantifiable risks associated with frequentist concepts) and *Knightian uncertainty* (which is unquantifiable). Under Knightian uncertainty people can say no more than that an event is probable or improbable; they cannot assign a number or ranking in their comparison of probabilities of different events. In the simplest terms the probabilities of Knightian risk and statistical/objective probabilities can be understood to be the same thing: Knightian risk events can easily be calculated using the frequency concepts associated with Classical statistical theory. These events tend to be repeatable and the outcome of a deterministic and immutable data generating mechanism, such as an unloaded die or a lottery machine. In a world of Knightian risk and quantifiable uncertainty it may be easy to assess and monitor expert judgement just by understanding the mathematics of the data generating process.

Keynes (1921) argues, however, that in only a limited number of cases can probabilities be quantified in a cardinal sense; in some cases, ordinal comparisons of probability may be possible, but often, particularly in the context of unique events, probabilities may not be quantifiable at all. In reality there may be little consensus in expert (or amateur) opinions – particularly in economic decision-making. Keynes (1921) therefore argues that events characterised by Knightian uncertainty are more common than those characterised by Knightian risk, at least in the economic and social sphere.

Such issues are of particular importance in economics because much economic behaviour is forward looking, experiments may not be repeatable, and conditions cannot be controlled. People often make subjective probability judgements about events that have not occurred before, for which the data generating mechanism cannot be known. This makes the quantification and assessment of probabilities particularly problematic because it becomes impossible to match subjective probability judgements with an objective probability distribution. Also, endogeneity (i.e. the path a system takes is determined by events within the system) will limit the accuracy of probabilistic judgements of future events when beliefs about the future are affected by beliefs about the present. Shiller (2003) analyses such phenomena in the context of feedback theory, describing the endogeneity in belief formation: beliefs about the system determine the path of that system (e.g. stock prices go up because people believe they will go up). In this sort of world, no matter how much experts know there are no objective probability distributions waiting to be discovered; probabilistic judgements will always concern subjective beliefs rather than an immutable reality.

These problems are more worrying for economists than for most natural scientists and certainly for most geoscientists, as there is a limit to what can be done to change existing geology and geological processes (an exception to this may exist in the area of climate change – see Wood & Curtis this volume). However, even though natural scientists often attempt to find out about an immutable, objective reality, the analysis of subjective probabilities is still of fundamental importance in the evolution of knowledge about natural phenomena under conceptual uncertainty. Lack of knowledge of the immutable reality limits the ability to match objective probability distributions and subjective probability assessments. Without knowledge of the mechanisms generating future outcomes, experts must rely on their subjective assessment of prior information.

Statistical probabilities rely on large data sets and assume an absence of subjectivity (Gould 1970). It is unlikely that frequentist approaches will have much resonance in analysing the elicitation of expert opinion on more complex geoscientific or economic issues: on such issues the experts whose opinion might usefully be sought are often few in number. Other approaches to quantification have focussed on stochastic modelling strategies using genetic algorithms and simulated annealing, and chaos or catastrophe theory (Smith et al. 1992; Gleick 1987; Brock 1998; Sornette 2003). These approaches adopt the assumption of some underlying order that might not be immediately obvious but is nonetheless theoretically quantifiable. If this is true, then the statistical probabilities will coincide with judgements of probabilities as long as the procedures of logical inference adopted are correct (Pshenichny, this volume). If experts are assumed to be consistent, rational and not prone to making systematic mistakes, then the distinction between conceptual probabilities and statistical probabilities disappears as uncertainty is reduced and as experts increase their knowledge of underlying data generating processes. However, experts can never be assumed to possess such qualities, as we show below.

Subjective Probabilities and Bayesian Analysis

Subjective beliefs are important in a world of conceptual uncertainty, and subjective probabilities can be analysed more effectively within a Bayesian approach than within a classical statistical approach. Bayesian analysis focuses on the subjective confidence that people have in a hypothesis about a single event and can be used to analyse the process by which subjective probabilities or judgements of confidence are updated as new information arrives.

Subjectivity can be thought of as a negative quality, particularly in science. However, the formation of subjective judgements is not necessarily problematic if these subjective judgements are derived in a consistent way (Cox 1946). If any given set of information always generates the same set of probability judgements, then judgements can be said to have formed in a systematic way. The recognition of this insight has made the old subjectivist versus frequentist debates somewhat redundant as focus has shifted towards Bayesian methods, and thus the sting of the term 'subjective' has been drawn.

The starting point in Bayesian analysis is the *prior* probability, which represents the odds that a researcher would place on a particular hypothesis before considering new data. This prior probability is combined with new data using Bayes's Theorem, resulting in a *posterior* probability. Bayes shows that the posterior probability of a hypothesis is given by the product of the prior probability derived from all relevant background information *B*, and the relative likelihood of the data having been recorded if the hypothesis were true (Lee 1997).

This theorem is best illustrated by means of an example. The posterior probability of finding oil (*O*) given particular, new geological data (*G*) in addition to the background knowledge (*B*) is calculated as:

$$P(O \mid G, B) = \frac{P(G \mid O, B)\ P(O \mid B)}{P(G \mid B)}$$

P(O|G,B) is the posterior probability of finding oil given data *G,*

P(O/B) is the prior probability of finding oil, conditioned on the model *B,*

$P(G|O,B)/P(G|B)$ is the relative likelihood of seeing data G if there is oil.

The use of Bayes's Theorem to augment geological prior information with new, geophysical data, is illustrated and developed in Wood & Curtis (2004).

The Bayesian approach differs in at least two significant ways from a Classical frequentist approach. First, the output, i.e. the posterior, is a probability density function, not a point estimate. In addition, it is directly related to the beliefs about a population parameter rather than being a sampling distribution of *estimates* of a population parameter (Kennedy 1998).

The virtue of a Bayesian approach in analysing expert judgement is that it captures the process by which subjective beliefs or degrees of confidences can be updated as new information arrives. Reckhow (2002) for example argues that the Bayesian approach provides a systematic procedure for pooling and combining knowledge in order to make decisions. The posterior is combined with a loss function (or a utility function representing gains) and a decision is made such that the expected loss is minimised (maximising expected gain).

There are, however, a number of problems with the Bayesian approach. First, there are practical problems in its application, e.g. in economics, there is often a paucity of data that can be used to quantify subjectively formed probability judgements (Kennedy 1998, p. 205). Also, human intuitive cognitive processes do not deal well with Bayesian concepts. Anderson (1998) argues that this is a consequence of the nature of memes (the cultural analogy of genes – see below). Anderson suggests that Bayesian approaches can be refined using the advantages of a frequentist approach, e.g. using mental /visual imagery. For example, consistent methods should be developed: probabilistic information should be represented in graphical or pictorial form, and more generally frequentist approaches should be adopted in the presentation of information, attention should be paid to devices for cognitive interpretation, and Bayesian analysts should develop conventions for graphic displays. In other words, some frequentist methods can be used effectively within a Bayesian framework such that human cognition will process subjective probabilities more effectively.

Probabilistic Judgements and Cognitive Bias

A Bayesian approach assumes some sort of order in the process of forming subjective beliefs. As outlined below, recent research within cognitive psychology suggests that perhaps expert opinion may not be the outcome of rational, systematic calculation. Some of the most common errors that characterise human assessment and processing of probabilities are now outlined.

In making probabilistic judgements, research has shown that most ordinary people make common mistakes in their judgements of probabilities (e.g., Anderson 1998). Experts are susceptible to similar biases, both on an individual basis and in terms of group biases. This is because of cognitive limitations in the processing ability of the human mind (Gould 1970; Tversky & Kahneman 1974; Anderson 1998).

The problem originates in the input format of data, and in algorithms used: if prompted by clear signals, the human brain is able to deal with probabilities effectively (Anderson 1998). For example, if students are asked to judge the probability of two coincident events within the context of a statistics class, then they will know what to do. However, if outside their classes they are confronted with a problem requiring probability judgements in a situation in which it is not obvious that this is what is required, then they may make a judgement using instincts and intuition rather than statistical reasoning (see Kyburg 1997). The key sources of inconsistency emerge from either *individual bias* or *group bias*.

Individual Bias

At least two main types of individual bias can be distinguished: motivational bias and cognitive bias (Skinner 1999). *Motivational biases* reflect the interests and circumstances of the expert (e.g., does his or her job depend on this assessment? If so, s/he may be overconfident in order to appear knowledgeable). Motivational biases such as these can often be significantly reduced or entirely overcome by explaining that an honest assessment is required, not a promise. Also, it may be possible to construct incentive structures encouraging honest assessments of information. Motivational biases can be manipulated because they are often under rational control.

Cognitive biases are more problematic because they emerge from incorrect processing of the information; in this sense they are not under conscious control. Cognitive biases are typically the result of using heuristics, the common-sense devices or rules of thumb derived from experience, used by people to make relatively quick decisions in uncertain situations. They are used because a full assessment of available information is difficult and/or time consuming or when information is sparse. For example, when thinking about buying/selling shares from their portfolio, a potential investor may have little real knowledge about what is going to happen to share prices in the future; given this limited information, they will adopt the heuristic of following the crowd, i.e. buying when the market is rising and selling when it is falling.

At least four types of heuristics that produce cognitive bias are commonly employed: *availability, anchoring and adjustment, representativeness,* and *control* (Kahneman *et al.* 1973; Tversky & Kahneman 1974). *Availability* is the heuristic of assessing an event's probability by the ease with which occurrences of the event are brought to mind. This often works quite well, but can be biased by the prominence of certain events rather than representing their frequency.

For example, headline news of airplane crashes will be brought to mind more readily than bike crashes, even though the latter are far more frequent. Similarly, the availability heuristic may cause geologists to recall the most interesting, attractive or complex field examples rather than those that are most often encountered, biasing their future interpretations.

Anchoring and adjustment is a single heuristic that involves making an initial estimate of a probability called an anchor, and then revising or adjusting it up or down in the light of new information (Tversky & Kahneman 1974). This typically results in assessments that are biased towards the anchor value. For example, in deciding about an appropriate wage demand to make in the face of an uncertain economic environment, workers will tend to anchor their demands around their existing wage.

The *control* heuristic is the tendency of people to act as though they can influence a situation over which they have no control. People value lottery tickets on which they have chosen the numbers more highly than those with random number selection, even though the probability of a win is identical in both cases.

The *representativeness* heuristic is where people use the similarity between two events to estimate the probability of one from the other (Tversky & Kahneman 1982). Consider the following example:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Please check off the most likely alternative:

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

If this problem were to be expressed in probabilistic / statistical terms: i.e. is 1. more likely than 1 AND 2, most people with a basic knowledge of probability would realised that two events happening together is less probable than each event happening irrespective of whether the other occurred. However, when confronted with the details about Linda, most people find the second option more likely than the first, simply because the description appears to be more representative of a feminist stereotype, and hence more plausible. This is a *conjunction fallacy*: the former option is the most likely since the probability of the conjunction of two events can never be more probable than either event independently. That is, a more detailed scenario is always at best equally (and usually less) probable than a simple scenario. In the same way that the probability of events compounded using logical AND is often overestimated, the probability of events compounded using logical OR is often underestimated (Bar-Hillel 1973).

Such biases can also create an unbounded probability problem: subjects tend to over-estimate each probability in a set of exhaustive and mutually exclusive scenarios, so that the estimated sum of all probabilities is greater than one (Anderson, 1998, p.15). Also, people do not correct their probability estimates when the set of exhaustive but mutually exclusive outcomes is augmented, again leading to an estimate of total probability in excess of one.

Other well-known biases introduced by the representativeness heuristic include the gambler's fallacy and base-rate neglect. The *gambler's fallacy* is the belief that when a series of trials have all had the same outcome then the opposite outcome is more likely to occur in the following trials, since random fluctuations seem more representative of the sample space. *Base-rate neglect* is neglect of the relative frequency with which events occur. Consider the following example: The group was told that Dick came from a population of 30 engineers and 70 lawyers:

> Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

This description provides no information about Dick's profession, but when subjects were asked to estimate the probability of Dick being an engineer, the median probability estimate was 50%, whereas the correct answer is 30%. Subjects ignored the base rate and judged the description as equally representative of an engineer and a lawyer (Tversky & Kahneman 1974, p. 1126).

An interesting and commonly used combination of the gambler's fallacy and base rate neglect is called *probability matching,* a heuristic known to be used by humans and some other primates (e.g., Bliss *et al.* 1995). This is where a reaction from a given range is chosen in proportion to the probabilities of occurrence of various consequences. An example given by Lo (2001) was from World War Two. Bomber pilots were allowed to carry either a flak jacket or a parachute, but not both because of the extra weight. They knew that their probability of getting strafed by enemy guns (requiring a flak jacket for protection from shrapnel) was three times that of being shot down (requiring a parachute). Pilots were observed to take flak jackets three times out of every four and parachutes on the fourth occasions. This is not an optimal assessment of the probabilities. Pilots were more likely to have survived if they had taken a flak jacket 100% of the time because the probability of getting strafed by enemy guns was always more likely than the probability of being shot down – the flak jacket was always more likely to be of use.

Probability matching might occur in a geological context if a human was asked to estimate the type of geology that is most likely at a set of sub-surface locations in a reservoir, knowing that wells would be drilled at those locations and their estimates checked. If all they knew about the geology in the reservoir was that it was either of a type 1 or a type 2, and type 1 was three times as likely to occur as type 2, then it is possible that they would posit geology type 1 as the most likely on average three times out of every 4. Although this would be a non-optimal prediction, this would be a natural tendency for any human, including an expert, who is not intimately familiar with basic probability theory.

Other cognitive biases reflect emotional responses. For example in most cases where elicitation has involved experts, the experts have individually been *overconfident* about their knowledge. Multiple experts undergoing the same elicitation procedure often produce barely- or non-overlapping estimates of elicited parameter values. Even groups of experts are observed to display overconfidence in their consolidated results (see below). Overconfidence is especially a problem for extreme probabilities (close to 0% and 100%) which people tend to find hard to assess.

Other forms of emotional response affecting the heuristics employed include mood: people in a happy mood are more likely to use heuristics associated with top-down processing, i.e. relying on pre-existing knowledge with little attention to precise details. By contrast, people in a sad mood are more likely to use bottom-up processing heuristics, paying more attention to precise details than existing knowledge (Shwarz 2000, p.434). Minsky (1997, p. 519) analyses some of the emotional constraints in the case of expert knowledge, arguing that the 'negative knowledge' associated with some emotional states may inhibit whole strategies of expert thought.

Of all of the biases described above, the most prevalent may be overconfidence and base-rate neglect (Baecher 1988). However, the frequency with which different heuristics are employed within the geosciences has never been assessed and would be an area of fruitful and useful future research.

Group bias

Until now we have assumed that experts are acting as atomistic agents. In reality, experts collect and confer and, in so doing, generate and perpetuate more complex forms of bias associated with group interactions. Difficulties in acquiring probabilistic knowledge by individual experts are compounded because mistakes and misjudgements may be communicated to other experts. This process is made more complex by a source of individual bias that emerges from anchoring effects - if one individual's judgements are 'anchored' to another's (Tversky & Kahneman 1974; Eichenberger 2001).

This implies that expert knowledge will not necessarily evolve along a pre-determined, objective path but instead may exhibit path dependency and persistence. Traditionally, the evolution of expert knowledge has been analysed in philosophical terms in the context of Kuhn's theories of scientific revolution. More recently, the evolution of group knowledge has been explained using approaches from evolutionary biology and economic models of mimetic contagion, e.g. in stock markets. Each of these three approaches is briefly described below.

Kuhn's Theory of Scientific Revolutions

Kuhn (1962) and Pajares (2001) explain how scientific thought emerges in the context of given paradigms, emerging from past scientific achievements. Students learn from experts and mentors and this moderates disagreement over fundamental theorems. This is *paradigm anchoring*: experts' beliefs are anchored to the existing dominant approach. Such paradigm-based research, however, forces thought within certain boundaries and this may lead to herding of expert opinion towards prevailing hypotheses and approaches. Experts will be unprepared to overthrow an old paradigm without a new paradigm with which to replace it, and it may take time for new paradigms to gain acceptability. In normal times 'mopping-up' exercises occur in which the anomalies that do not fit with the current paradigm are discarded (a form of cognitive dissonance may occur - cognitive dissonance being the process of rationalising information that does not fit with preconceived notions of how the world works). Once a theoretical paradigm has become established, alternative approaches are resisted and paradigms will shift only when evidence and anomalies accumulate to such an extent that a scientific crisis develops. When this happens, a scientific revolution is precipitated.

The problem with this deterministic, descriptive account of how new theories emerge is that it does not illuminate the *processes* underlying the formation of expert opinion. By contrast, biological and economic research has focussed more closely on underlying mechanisms, as explained below.

Analogies from Evolutionary Biology

Path dependency in the evolution of scientific beliefs can be described using biological analogies, e.g. those based around the concept of a *meme*, the cultural equivalent of a gene (Dawkins 1976). Imitation is a distinguishing characteristic of human behaviour and a meme is a unit of imitation (Blackmore 1999). The discovery of 'mirror neurons' (neurons in the pre-motor areas of primate brains that are activated without conscious control and generate imitative behaviour in primates) has lent some scientific support to these biological explanations for imitative behaviour (Rizzolatti *et al.* 2002). This biological approach is also consistent with the use of neural networks for information processing: i.e. mathematical approaches that emulate adaptive learning processes observed in human brains.

Biological insights can be applied in the analysis of belief formation in a human context. Anderson (1998) asserts that successful memes survive (that are remembered) and reproduce (are transmitted) effectively when they a) map effectively onto human cognitive structures, b) incorporate a standardised decision structure, and c) have been reinforced by dominant members of the scientific community. Lynch (1996, 2003) applies these insights in his analysis of the evolutionary replication of ideas and argues that 'thought contagion' affects a wide range of human behaviours and beliefs, including the analysis of stock market behaviour.

Economic Analogies: Herding and Mimetic Contagion

An interlocking literature assessing various possibilities for thought contagion has developed in economics, beginning with Keynes's analysis of uncertainty, rationality, subjective probabilities, herd behaviour and conventions (Keynes 1921, 1936, 1937). In Keynes's analysis, herding behaviours are linked back into an analysis of probabilistic judgement in a Bayesian setting. Differences in posterior judgements of probable outcomes may not reflect irrationality but instead may emerge as a result of differences in prior information. Rational economic agents may have an incentive to follow the crowd and herding will result as a response to individuals' perceptions of their own ignorance. This herding will be rational if an individual has reason to believe that other agents' judgements are based upon better information than their own: other people's judgements become a data-set in themselves. In this way, people will incorporate others' opinions into their prior information set and their posterior judgements may exhibit herding tendencies. Shiller (2000, 2003) analyses these ideas in the context of feedback theories of endogenous opinion formation in which beliefs about the system determine the path of that system, e.g. as is seen in stock markets. These ideas are also developed in Topol (1991), Schleifer (2000), Brunnermeier (2001) and Sornette (2003), amongst others.

Ideas about herding can be applied to the literature on the acquisition of expertise in an academic context in recognising that divergent expert opinions reflect uncertainty rather than irrationality or misguided thought. The incorporation of the judgement of other experts into experts' prior information sets explains herding tendencies. However, whilst expert-herding behaviour can be explained as a rational phenomenon, the existence of herding may still contribute to instability if the herd is led down the wrong path. Stable outcomes will only be achieved if the herd can be led along a path of increasing the stock of common (real) knowledge. In such cases, increases in the stock of reliable prior information will contribute to convergence in posterior probabilities.

If, however, the herd path fosters increasing noise within the system then the process of opinion formation will become unstable. Further research is needed to assess the extent to which expert herds move in either stable or unstable directions. This can be done by assessing the extent to which herd leaders (experts) are selected on objective versus subjective grounds, and by assessing the extent to which herd leaders turn out to be right in the end. This would be another direction for fruitful and useful future research.

Expert elicitation techniques (methods of interrogating experts for information – see below) address these issues. It should be noted, however, that the implications for the social sciences versus the physical sciences may be different because the existence of an objective and immutable reality in many situations in the physical world contrasts with the endogenous, mutable nature of reality in the social, economic and cultural sphere. Feedback-loops between belief and reality are therefore less likely to occur in the geosciences.

Elicitation Theory

The preceding sections have outlined in general terms how and why people use new information to form probabilistic judgements. This section describes how these ideas are addressed when eliciting and processing expert knowledge and opinion. Below, the key elements of elicitation methods are outlined, highlighting various sources of individual and group bias and how these may be reduced.

How experts think

The first step in expert elicitation involves identifying experts. Wood & Ford (1993) outline four ways in which an expert's approach to problem solving differs from a novice's approach to problem-solving: expert knowledge is grounded in specific cases; experts represent problems in terms of formal principles; experts solve problems using known strategies; experts rely less on declarative knowledge (the what) and more on procedural knowledge (the how).

Eliciting and documenting expert judgement

Biases in knowledge and judgement defined earlier will emerge for experts just as they emerge for ordinary people making everyday decisions. Suitable elicitation methods can sometimes correct the biases in expert opinion and the problem tackled in the field of Elicitation Theory is to design the best way to interrogate experts or lay-people in order to obtain accurate information about a subject in question.

There are no universally accepted protocols for probability elicitation and there is relatively little formal empirical evaluation of alternative approaches (though some examples and practical guidelines are outlined in Meyer *et al.* 1991, 2003). There are, however, three common assessment protocols: the Stanford/SRI protocol, Morgan and Henrion's protocol, and the Wallsten/EPA protocol (Morgan & Henrion 1990). Both the Stanford/SRI and Morgan and Henrion's protocols include 5 principal phases. These include motivating the experts with the aims of the elicitation process, structuring the uncertain quantities in an unambiguous way, conditioning the expert's judgement to avoid cognitive biases, encoding the probability distributions and verifying the consistency of the elicited distributions. The Wallsten/EPA protocol includes the preparation of a document that describes the objectives of the elicitation process, descriptions of cognitive heuristics and biases, and other relevant issues. The expert reads the document before the elicitation process, which is similar to a step in Morgan & Henrion's (1990) protocol.

Within each protocol the elicitor must decide exactly what problems to tackle or what questions to ask in order to maximise information about the topic of interest. Coupe & van der Gaag (1997) showed how a sensitivity analysis might sometimes be carried out in order to see which elicited probabilities would have most influence on the output of a Bayesian belief network. This issue is developed in Curtis & Wood (this volume) who optimise the design of the elicitation process in real time. They take into account all information available, as it is elicited, and design the elicitation procedure using the experimental design method of Curtis *et al.* (2004). This must be an optimal strategy, although the details of any particular method must be tailored to particular tasks.

Editing Expert Judgement: the Problem of Calibration

If expert judgement is affected by the biases described earlier, then the results of elicitation from each individual expert will need to be calibrated against each other. This in turn requires some statistical model of the elicitation process. The main model proposed in the literature is that of Lindley *et al.* (1979) requires that there be an objective assessor who will consolidate the results derived from subjective experts. It is not clear, however, why an assessor should be any more objective than the experts.

Other work that attempts to calibrate experts' judgement includes that of Lau & Leong (1999) who created a user-friendly JAVA interface for elicitation that includes graphical illustrations of possible biases and any inconsistencies in elicited probability estimates. The interface then enters into dialogue with the expert until consistency is achieved. This allows the experts themselves to try to compensate for their natural biases and inconsistencies, without the need for a statistical model applied by an external elicitor. Other methods to deliver the questions asked of the experts in graphical form were reviewed by Renooij (2001). While it is clear, however, that a careful choice of graphical representations of the questions and answers can reduce bias, it is likely that this can only provide partial compensation.

It should be noted that some of the heuristics described earlier perform extraordinarily well in some situations (Gigerenzer & Goldstein 1996; Juslin & Persson 2002). Gigerenzer & Goldstein (1996) examined a particular controlled task using a 'take-the-best' algorithm. This algorithm selects a best guess at the answer to a question from a set of possibilities by using the minimum number of heuristics from a ranked list such that a guess can be made. The 'take-the-best 'algorithm worked as well as an algorithm that used full probabilistic information to make the guess, but at a fraction of the cost. The problem is that in practical situations it is not clear from the results alone whether the heuristics work well or not since there is no objective answer with which to compare them. The role of the elicitor is to try to reduce the use of heuristics unless there is no alternative but to use them. In the latter situation, at least the heuristic used should be explicitly understood by all involved so that the results can be treated as conditional on this heuristic being effective.

Group elicitation

As discussed previously, natural biases in human cognitive processes and biases caused by paradigm anchoring and scientific revolutions, mimetic contagion and herding effects, result in individuals' cognitive biases being compounded in groups. If expert opinion evolves along a particular path just because others have started on that path then the link between subjective probabilities and underlying objective probability distributions may be completely broken. Expert opinion may be led further and further way from an objective grounding and the evolution of knowledge becomes endogenous, determined by events within the system itself. Group biases may particularly affect the evolution of expert opinion in geoscientific contexts where relatively weak information is available compared to that which would be necessary to constrain belief to be close to reality.

Say we wish to assess probabilities based on the estimates of several experts, each of whom have different background knowledge. Few studies have addressed the issue of how best to combine such knowledge into a single probability distribution function (PDF). Individual assessments will almost certainly differ, sometimes by orders of magnitude (see below). In order to reconcile these differences we can either combine the individual assessments into a single one, or we can ask the experts to reach a consensus. The former approach assumes that nothing extra is gained by sharing knowledge and ideas among the experts. The second approach can be jeopardised by group interaction problems (the dominance of one expert over others, or the pressure for conformity). It remains unclear which method produces more accurate final probability estimates.

Philips (1999) studied a case where two groups of experts of varying relevant backgrounds assessed, both individually and in groups, the PDF of very long term corrosion rates of carbon steel drums containing nuclear waste, after the containers have been sealed in a concrete, underground bunker. Philips's first important result was that the two approaches described above lead to different probability assessments: the average of individual assessments of the PDF did not match the group consensus PDF.

Second, individuals' median estimates in each group initially spanned three orders of magnitude. Individuals' PDF 's were assessed at three stages during the ensuing discussions, and it was found that this median spread decreased consistently during the formation of a joint consensus distribution (i.e., inter-expert discussion resulted consistently in some convergence between the individuals' views) in only one of the two groups. The observed convergence was, however, accompanied by an increase in the variance or spread of each individual's PDF estimates. This is a typical feature of anchoring and adjustment – individuals simply increase the range of their initial PDF's in order to encompass the range of the consensus PDF.

Third, even after reaching consensus within each of the groups of experts, the resulting two groups' PDF's differed in their median estimates by three orders of magnitude. This is the same magnitude of difference as was observed between individuals in each group. On further analysis, the reason turned out to be that each group agreed different basic assumptions in the initial discussions (there was a three-fold difference between the two groups' estimates of alkalinity on the outside of the steel drums, and one group considered this sufficient to accelerate the rate of corrosion). We can conclude that particular attention should be paid to surfacing initial assumptions during the elicitation process (but note that deciding on appropriate assumptions may in turn require another elicitation process!)

Geoscientific Example

In relatively simple situations there have been apparently successful attempts to elicit uncertain information from experts within the geosciences. One such case is described by Rankey & Mitchell (2003). Six seismic interpreters with different levels and types of experience were asked to define the depth of the top of a roughly circular buried Devonian pinnacle reef (a hydrocarbon reservoir) in the Western Canadian Basin. They were provided with 3D seismic data spanning the reef, and initially with log data from two wells to which the seismic had been matched in depth. After a first-pass interpretation they were provided with logs and depth matches from a further four wells, and also with literature on the geology of the area, after which they were allowed to revise their interpretation. Four of the six interpreters chose to do so, although all changes made were minor. To maintain consistency the base of the reef was picked prior to the experiment, considered "known" for this experiment, and given to all interpreters. The purpose of the exercise was to elicit information both on the location of the top of the reef, and on its uncertainty.

Particular features to note about this elicitation process are:

- interpreters were kept separate from one another during the exercise so that there were no group interaction effects,
- there would appear to be few motivational biases – none of the interpreters (to our knowledge) had a vested interest in any particular results,
- there may have been some anchoring after the first pass interpretation: one of the interpreters who did not change his interpretation noted, "I did… not want to change any of my picks based on the additional well data – looks like I had it nailed." There may have been an element of pride for some interpreters in appearing to get the first pass interpretation correct.
- The basic assumptions intrinsic to the experiment, and made by all interpreters (e.g., the depth of the base of the reef, the assumed accuracy of the depth-matches to well logs) themselves required an elicitation (interpretation) process. This was not detailed in the report by Rankey & Mitchell (2003), but it is clear that errors in such assumptions could impact both the interpretations made in this experiment, and potentially their uncertainty.

This exercise was agreed by all interpreters to be relatively simple: the seismic was of good quality, the stratigraphy was simple, and the well ties appeared to be good. One interpreter commented that there were few critical decision points, the main one being whether to pick a high or a low arrival on the seismic data through the reef's southern flank. No interpreters found any decisions difficult to make. The southern-central part of the reef is extensively dolomitised, the rest of the reef is limestone.

Most interpreters picked very similar top surfaces all over the reef, other than around the pinnacle margins. The largest variation was in the southern-central section. In fact, interpreters were picking a "zero-crossing" on the 3D seismic data volume (a point at which the seismic data changes from positive to negative values) because this coincided with the reef crest on one of the wells. This zero-crossing became less distinct and appeared to bifurcate into an upper and lower zero-crossing over the dolomitised area.

Subsequent modelling showed that the zero-crossing was probably due to interference in the seismic waves ('tuning') between reflections from the limestone and from layers just above the reef (the reef crest was not the largest seismic impedance contrast in its vicinity). The difficulty in the southern-central section was probably due to different tuning effects caused by the switch from limestone to dolomite, causing the zero crossing to deviate from the reef top. Interpreters did not know the results of this modelling during this exercise.

Each interpreter was fairly confident of their own interpretation (their individual estimated uncertainty was low), yet the spread of interpretations around the southern-central margin was significant over an area of 200 m laterally. In this area, differences in gross pore volumes predicted using seismic attributes around the interpreters' top reef estimates differed by up to 24%; predictions of gross rock volume varied by up to 13%. Interpreters had correctly identified this area as containing a significant decision point, but had assumed that they had made the correct decision. It therefore seems reasonable to conclude that the over-confidence bias occurred in most interpreters.

Ultimately there is no proof that the final uncertainty estimates obtained in this experiment were reasonable, as it is not reported that a well was drilled in the more complex southern-central area to check whether the true top-reef depth lay within the bounds defined by the range of interpretations. However, this is one of the few geological interpretation cases within the hydrocarbon industry where a genuine attempt was made to hold a set of well-controlled trials in order to assess this uncertainty. In most cases a single individual interprets the seismic data, the interpretation being 'checked' by one other. In the above experiment this would have lead to an over-confident and biased interpretation, hence the results of this experiment at least demonstrated a successful improvement on more usual methods.

Conclusions

This paper explains how the judgements of experts can be biased by their use of heuristics to guide the formation of their opinions. We present research from cognitive psychology showing that such heuristics are used both by experts and by lay people alike, and often cause biases in individuals' perceived knowledge – commonly overconfidence and base-rate neglect. We show that models from economic and biological literature explain how in conditions of uncertainty or asymmetric prior information, such biases can cause herding behaviour, potentially leading to instability in the stock of common, accepted knowledge. Elicitation theory is described that attempts to elicit robust information from experts by mediating the effects of such biases.

During the course of this discussion paper we identified several key directions in which future research is likely to be both useful and fruitful:

- Measure the frequency with which different heuristics tend to be employed by experts within the geosciences. This will help to define the range of biases that may occur, and their likely prevalence.
- Develop geoscience-specific methods to reduce biases originating from the use of such heuristics.
- Create methods to detect herding behaviour amongst scientific experts.
- Assess whether herd leaders are selected on objective or subjective grounds, or whether they are better described as emergent or self-created. Also assess (in retrospect) the frequency with which they turned out to be correct in their contested and uncontested opinions.
- Research how best to reconcile opinions from multiple experts to obtain the best probabilistic description of an unknown, objective reality (in cases where one exists).

Geological prior information is often partially opinion or judgement based. When assembling prior information in order to augment it with new data, at the very least it is necessary to be aware of the heuristics and biases discussed herein so that their propagation to future, accepted knowledge can be limited. In itself this will help to reduce emergent group behaviour such as herding, thereby increasing the probability that the stock of common knowledge becomes stable, and converges towards facts about the Earth as it really is.

References

Anderson, J. L. 1998. Embracing Uncertainty: the Interface of Bayesian Statistics and Cognitive Psychology. Conservation Ecology [online] 2(1), 2. World Wide Web Address: http:www.consecol.org/vol2/iss1/art2.

Baecher, G. B. 1988. Judgemental probability in geotechnical risk assessment. Technical report, prepared for The Office of the Chief, US Army Corps of

Engineers. World Wide Web Address:http://www.ence.umd.edu/~gbaecher/papers.d/judge_prob.d/judge_prob.html

Bar-Hillel, M. 1973. On the subjective probability of compound events. *Organizational behaviour and human performance,* **9,** 396-406.

Blackmore, S. 1999. *The Meme Machine.* Oxford University Press, Oxford.

Bliss, J. P., Gilson, R. D., & Deaton, J. E. 1995. Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics,* **38**, 2300-2312.

Brock, W. A. 1998. Chaos and Complexity in Economic and Financial Systems.*In:* Von Furstenburg, G. (ed.) *Acting Under Uncertainty: Multidisciplinary Approaches* Boston, MA: Kluwer Academic Publishers.

Brunnermeier, M. K. 2001. *Asset Pricing Under Asymmetric Information.* Oxford University Press, Oxford.

Bulmer, M.G. 1979. *Principles of Statistics.* Dover, New York.

Carnap, R. 1950. *Logical Foundations of Probability.* Routledge & Kegan Paul, London.

Coupe, V. M. H., & Van Der Gaag, L. C. 1997. Supporting Probability Elicitation by Sensitivity Analysis. Lecture notes in computer science. *Knowledge acquisition,Modeling and Management,* **1319**, 335-340.

Cox, R.T. 1946. Probability, Frequency and Reasonable Expectation, *American Journal Physics* **14,** 1-13.

Curtis, A., Michelini, A., Leslie, D. & Lomax, A. 2004. A deterministic algorithm applied to experimental design applied to tomographic and microseismic monitoring surveys. *Geophysical Journal International*, In Press.

_________ & wood, R. 2004. Optimal elicitation of probabilistic information from experts. *This volume.*

Dawkins, R. 1976. *The Selfish Gene.* Oxford University Press, Oxford.

Eichenberger, R. 2001. Economic incentives transform psychological anomalies. *In:* Frey, B.S. (ed) *Inspiring Economics Human Motivation in Political Economy.* Cheltenham, Edward Elgar, xxx-xxx.

Gigerenzer, G. & Goldstein, D. G. 1996. Reasoning and fast and frugal way: Models of bounded rationality. *Psychological Review,* **103,** 650-669.

Gleick, J. 1997. *Chaos – Making a New Science.* Harmondsworth, Penguin.

Gould, P. 1970. Is *statistix inferens* the geographical name for a wild goose? *Economic Geography*, **46**, 438-448.

Juslin, J. & Persson, M. 2002. Probabilities from Examplars (PROBEX): a “lazy” algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, **26**, 563-607.

Kahneman, D. & Tversky, A. 1973. On the psychology of prediction. *Psychological Review*, **80**, 237-51. [Reprinted In: Kahneman, D., Slovic, P., & Tversky, A. (eds) 1982. *Judgement under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, 48-68.

_____________ & ____________1982 Subjective probability: A judgement of representativeness. *In:* Kahneman, D., Slovic, P. & Tversky, A. (eds) *Judgement under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, 32-47.

Kennedy, P. 1998. *A Guide to Econometrics* (4th edition). Blackwell, Oxford.

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

____________ 1936. *The General Theory of Employment, Interest and Money.* Macmillan, London.

____________ 1937. The General Theory Of Employment. *Quarterly Journal Of Economics*, **51**, 209-223.

Kuhn, T. 1962. *The Structure of Scientific Revolutions.* University of Chicago Press, Chicago.

Kyburg, H. E. 1997. Expertise And Context In Uncertain Inference. *In:* Feltovich, P.J., Ford, K.M. & Hoffman, R.R. (Eds) *Expertise In Context: Human And Machine* Cambridge Massachusetts. MIT Press And American Association for Artificial Intelligence, 499-514.

Lau, A.-H. & Leong, T.-Y. 1999. Probes: a framework for probability elicitation from experts. *Amia Annual Symposium, American Medical Informatics Association,* 301-305.

Lee, P. M. 1997. *Bayesian Statistics: An Introduction* (2nd Edition). Arnold, London.

Lindley, D. V., Tversky, A. & Brown R. V. 1979. On the reconciliation of probability assessments (with discussion). *Journal of the Royal Statistical Society,* **142,** 146-180.

Lo, A. W. 2001. Bubble, rubble, finance in trouble? Edited Luncheon Address Delivered June 1st At The Third Annual Institute of Psychology and Markets Conference.

Lynch, A. 1996. *Thought contagion: how belief spreads through society.* Basic Books, New York.

________2003. An introduction to the evolutionary epidemiology of ideas. *The Biological Physicist,* **3**, 7-14.

Meyer, M. A. & Booker, J. M. 1991. Eliciting and analyzing expert judgement: a practical guide. *Asa-Siam Series on Statistics and Applied Probability, Philadelphia,* 7.

_____________ Butterfield, K. B., Murray, W. S., Smith, R. E. & Booker, J. M. 2003. Guidelines for eliciting expert judgement as probabilities or fuzzy logic. *In:* Ross, T., Booker, J. & Parkinson, W. (Eds) *Fuzzy Logic and Probability Applications. Bridging The Gap.* ASA-SIAM Series On Statistics And Applied Probability, Philadelphia, 7.

Minsky, M. 1997. Negative Expertise. *In:* Feltovich, P.J., Ford, K.M., & Hoffman, R.R. (Eds) *Expertise In Context: Human and Machine.* Mit Press and American Association for Artificial Intelligence, Cambridge Massachusetts, 515-521.

Morgan, M. G. & Henrion, M. 1990. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press, Cambridge.

Pajares, F. 2001. The structure of scientific revolutions. Philosophers Web Magazine. World Wide Web Address: Http://Www.Philosophers.Co.Uk.

Philips, L. D. 1999. *Group elicitation of probability distributions: are many heads better than one? Bayesian research conference; decision science and technology, reflections on the contributions of ward edwards,* Klumer Academic, Boston And London, 313-330.

Pshenichny, C. A. 2003. Obtaining probability assessments from concepts. *This Volume.*

Rankey, E. C. & Mitchell, J. C. 2003. That's why it's called interpretation: impact of horizon uncertainty on seismic attribute analysis. *The Leading Edge,* **22,** 820-828.

Reckhow, K. H. 2002. Bayesian approaches in ecological analysis and modeling. *In:* Canham, C. D., Cole, J. J. & Lauenroth, W. K. (eds) *The Role of Models in Ecosystem Science,* Princeton University Press.

Renooij, S. 2001. Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review,* **16**, 255-269.

Rizzolatti, G., Craighero, L. & Fadiga, L. 2002.The Mirror System In Humans. *In:* Stamenov, M. I. & Gallese, V. (eds) *Mirror Neurons And The Evolution Of Brain And Language.* Advances In Consciousness Research, **42,** John Benjamins, Amsterdam, 37–59.

Shiller, R. J. 2000. *Irrational Exuberance.* Princeton University Press, Princeton.

_________2003. From efficient markets theory to behaviour finance. *Journal of Economic Perspectives*, **17**, 83-104.

Shleifer, A. 2000. *Inefficient markets: an introduction to behavioural finance.* Oxford University Press, Oxford.

Shwarz, N. 2000. Emotion, cognition, and decision making. *Cognition and Emotion,* **14,** 433-440.

Skinner, D. C. 1999. *Introduction to decision analysis.* Probabilistic Publishing.

Smith, M., Scales, J.A. &Fischer, T. 1992. Global search and genetic algorithms. *The Leading Edge,* **11**, 22-6.

Sornette, D. 2003. *Why stock markets crash: critical events in complex financial systems.* Princeton University Press, Princeton.

Topol, R. 1991. Bubbles and volatility of stock prices: effect of mimetic contagion. *Economic Journal,* **101,** 786-800.

Tversky, A. & Kahneman, D. 1974. Judgement under uncertainty: heuristics and biases. *Science,* **185**, 1124-1131.

________& ___________1982. Judgements of and by representativeness. *In:* Kahneman, D., Slovic, P. & Tversky, A. (eds) *Judgement under Uncertainty: heuristics and biases.* Cambridge University Press, Cambridge, 84-98.

Wood, R. & Curtis, A. 2004. Geological prior information, and its application to geoscientific problems. *This Volume*

Wood, L. E. & Ford, J. M. 1993. Structuring interviews with experts during knowledge elicitation. *International Journal of Intelligent Systems***,** **8**, 71-90.