The Knowledge Engineering Review, Vol. 10:1, 1995, 3–19

Cognitive expertise research and knowledge engineering

FERGUS BOLGER
University College London, Gower Street, London WC1E 6BT, UK

Abstract

This paper is a review of research into cognitive expertise. The review is organized in terms of a simple model of the knowledge and cognitive processes that might be expected to be enhanced in experts relative to non-experts. This focus on cognitive *competence* underlying expert performance permits the identification of skills and knowledge that we might wish to capture and model in expert systems. The competence perspective also indicates areas of weakness in human experts. In these areas, we might wish to support or replace the expert with, for example, a normative system rather than attempting to model his or her knowledge.

The review of studies of experts in a number of domains reveals that, indeed, experts show enhanced competence relative to non-experts in many instances. This conclusion is arrived at by taking into consideration numerous methodological problems with the study of expertise. These problems have often resulted in poorer performance by experts than they were capable of in more ideal circumstances. However, circumstances in the real world are not always more favourable towards experts than those in the psychological laboratory. Hence in some cases, one finds no difference between expert and non-expert performance and in some extreme cases experts may perform at a lower level than non-experts due to the particular contingencies of their working environment. On the whole, it is possible through *task analysis* to identify those circumstances where expert competence is likely to be enhanced relative to non-experts, and those where it is not. Further, the research indicates that there are some sorts of tasks which people, expert or not, find particularly difficult. These tasks, such as ones that require complex probabilistic reasoning or weighting and combination of large numbers of cues, demand special attention from knowledge engineers.

1 Defining expertise: The competence model

In this paper, I will first propose a definition of expertise in terms of *competence*. This kind of definition, I will argue, is the most useful for the purposes of knowledge engineers. After outlining some of the difficulties of inferring competence from observed performance, and of conducting research into cognitive expertise *per se*, I will briefly review research findings from the fields of judgment and decision-making, and from cognitive science, which speak to the competence of experts.

Note that I say "briefly review". There are by now a very great number of studies concerned with cognitive expertise. I will only select a few which I consider to be representative and illustrative. There are plently of good reviews of the expertise literature (for example, Schwartz & Griffin, 1986; Mumpower et al., 1987; Chi et al., 1988; Dowie & Elstein, 1988; Ericsson & Smith, 1991; Wright & Bolger, 1992), so I direct the reader to these if greater detail is required. My aim in this paper is not to produce a comprehensive review of the expertise literature but rather to extract from this literature—by means of the competence model as an organizing principle—the main lessons (and questions) about expertise relevant to knowledge engineering. These lessons will form the focus of the final section of this paper.

An obvious definition of expertise is that of some improved performance over people or systems regarded as "inexpert". Of course, this is rather simplistic in that it avoids some important questions (for example, how much of an improvement in performance do we require? How do we go about measuring improvement? How do we identify the baseline non-experts?) However, the emphasis on performance is an improvement over commonly used social definitions of expertise. Here professional qualifications, salary, position within an organization, number of publications, media profile and so forth are taken as an indication of expertise. It should become clear that socially defined expertise may not be very well correlated with actual ability.

Although defining experts in terms of improved performance is essential to the identification of experts it is not a definition which is at an appropriate level for the purposes of discovering what, if anything, is special about experts which we might wish to capture, model or support. For these purposes I propose that it is necessary to consider performance as a reflection of an underlying competence, in other words, enhanced knowledge and skills relative to the non-expert. It is these superior skills and knowledge which must be identified and described for knowledge engineering purposes.

I have enumerated below some candidate forms of knowledge and skills which might be enhanced in an expert. These I have organized in terms of a simple model of the cognitive system whereby information is collected and filtered by peripheral input systems, passed to central processing systems for deeper analysis, then finally passed to peripheral output systems for translation into an action, for example, a decision, a plan or a forecast (which may lead to new information being received by input systems, etc.).

Input processes

- attentional and perceptual skills (for example, problem recognition and interpretation, information gathering);

Central processes and knowledge

- cognitive skills (for example, memorial abilities, problem-solving strategies, learning)
- personality and affective traits (for example, confidence in judgment, motivation)
- "declarative" domain knowledge (for example, facts, rules)
- "procedural" domain knowledge (for example, algorithms, heuristics)
- "meta-knowledge" (for example, knowledge about the limitations of what one knows, and/or of one's abilities and strategies);

Output processes

- social skills (for example, communication, self-presentation)
- production and application skills (for example, motor skills, finding necessary resources).

This list is by no means meant to be exhaustive. Also the precise nature of cognitive processes and the representation of knowledge is unknown and debated (for example, to what extent is knowledge "declarative" or "procedural" or either?). Further, not all the above skills and knowledge need to be enhanced for someone to be "expert" although, at minimum, one would expect to find at least domain knowledge improved. Finally, faster, more automatic responding seems also to be a characteristic of experts relative to non-experts in certain domains.

Although the above list of skills and knowledge has its limitations, it does have its use for organizing a review of expertise research and for focusing our discussion of the implications of this research for knowledge engineering.

2 Limitations on acquiring and determining expertise

2.1 Acquisition of expertise

In some task domains it may not be possible to attain much enhanced performance due to, for example:

- low predictability (performance can never be good, for instance, in long-range weather forcasting)

- uniqueness of problems (this creates difficulties for learning generalizable rules and occurs, for example, in law and clinical psychology)
- lack of objective standards. (How does one assess whether performance is improving? This is a problem in, for instance, psychiatry)
- complexity/amount of information (it may be too difficult to process information even with assistance because the information is being received too fast and/or it is changing dynamically as in, for example, on-line share dealing)
- unavailability of relevant/useable outcome feedback. (If you cannot see whether what you did in the past was correct you cannot learn. This is the case, for instance, in life insurance underwriting where claims are usually only made many years after the underwriting decision).

In certain circumstances, the practices of experts (for example, professional practices) might even lead to experts demonstrating worse (for example, more biased) performance than non-experts. For example, "defensive medicine" (practising medicine so as to avoid litigation) is likely to lead to over-diagnosis and consequently to over-treatment.

Where expert performance is attained in a domain it may not be generalizable to another even seemingly similar domain (as will become apparent later, expertise seems to be domain specific).

2.2 Determining expertise

Identification of experts (for example, to study or make use of for system development) may not be easy due to the problems of definition outlined above. After identification, experts may be unwilling to cooperate (for example, because of time constraints, threat to their status) and at any rate, by their very nature, will tend to be rare. For these reasons many studies have used very small samples (and therefore may have lacked the statistical power necessary to detect enhancements in performance) whilst others may not have studied "true" experts (for example, social experts may not necessarily be competent due to the limitations listed above.

Assessment of expertise is also problematic. Inferences about competence from measurement of performance can be incorrect for a number of reasons:

- the experimental task may not be within the expert's domain of competence (for example, because of domain specificity of expertise);
- the expert may not be able to express competence in a metric or language which is meaningful to the researcher (for example, because of automaticity of expertise);
- there may not exist relevant standards against which to assess performance (for example, because of the domain problems listed above; using non-experts as the standard is problematic if you cannot specify what "good" performance is).

In many cases, inappropriate subjects and/or tasks have been used in expertise research, hence "ecological validity" (the degree to which the laboratory tasks correspond to what the experts really do, and therefore experimental findings can be generalized to the real world) has often been poor. For a more detailed discussion of the influence on conclusions regarding the quality of expert judgment of the interaction between factors affecting the ability to acquire good performance, and the ecological validity of tasks used, see Bolger and Wright (1994).

3 Review of research

As already indicated, I will conduct this review in terms of what the research findings tell us about expert competence relative to that of non-experts. I will also refer to the problems of definition and assessment of expertise mentioned in the previous section wherever appropriate.

3.1 Input and output processes

Although this review is chiefly concerned with cognitive processes, these do not operate in a vacuum—information must be extracted from the world and passed to cognitive processes and the results of cognitive processing must be translated into some motor action, even if this is merely writing or speaking. Either perceptual or motor skills or both may be more highly developed in experts than non-experts and this may account for much of their superior performance. This is likely to be particularly true in certain domains such as music, sport and transcription typing. However, in adults, it is unlikely that there is much development of the physiological aspects of the perceptual system, hence any improvements associated with expertise must be at a cognitive level. With respect to motor skills, it may be more difficult to disentangle the effects of physical training regimes from the acquisition of more efficient information-processing routines, so care must be taken when assessing cognitive expertise in domains which require a large component of physical training.

There is evidence of enhanced perceptual and motor abilities in experts from a number of sources. For example, de Groot (1978) found that chess masters could recognize familiar patterns of pieces, and consequently recall these positions or use them to generate moves, after only very brief exposure times. This observation is supported by other experiments in the domain of chess suggesting enhanced perceptual abilities in stronger compared to weaker players for making same-difference comparisons of quarter boards (Ellis, 1973), predicting capture (Milojovic, 1982), and piece-enumeration (Saariluoma, 1990). An example of enhanced motor performance in experts comes from transcription typing (Salthouse, 1984, 1985). In this case not only are experienced typists faster than less-experienced typists, but they can also type at speeds which are far in excess of speeds that should be attainable on the basis of reaction-time experiments requiring a key to be pressed in response to presented stimuli. Salthouse's experiments show that such speeds are possible in typing because the typist is able to process the information to be typed ahead of actually typing it (this is not possible in reaction-time experiments where stimuli are presented sequentially). Typists are therefore able to prepare for a subsequent keystroke (for example, by moving a finger whilst performing the current keystroke). In this way, natural processing limitations can be circumvented. Similarly, in the chess examples, it has been suggested that recognition can be accelerated by grouping individual pieces into commonly occurring configurations (Chase & Simon, 1973). As we shall see, this "chunking" of items of information also has consequences for memory performance.

In not all cases has enhanced perceptual or motor skills been found in experts. For example, in the domain of chess again, Waghorn (1988) found no differences between rated and non-rated players in their speed of check detection. An explanation for this null result is that the check-detection problem used is a relatively easy one (it involved few pieces) so that even the non-rated players tested already had the low-level pattern recognition skills needed to perform well. This illustrates one of the problems faced in attempting to assess expert competence: the difficulty of identifying tasks which are appropriate to the actual enhancements in competence possessed by experts. Use of inappropriate tasks may be one of the reasons why experts have not been found to perform better than non-experts in many expertise studies. Another methodological point arising from this example is that of appropriate comparison groups. There may well be a hierarchy of skill in most domains such that some skills are shared by people who are at different levels of competence with respect to other skills.

Another aspect of input processes is the ability to select, attend to, or in some way filter relevant information from irrelevant. On the whole, research into this aspect of expert competence has not been very favourable to experts. A number of studies have demonstrated that professionals do not always select the most appropriate information as the basis of their decisions. Ebbesen and Konecni (1975) found that in making sentencing decisions, court judges used only a very restricted subset of available dimensions. Further, Gaeth and Shanteau (1984) found that expert soil judges referred to materials, when categorizing samples, which are irrelevant to the discriminations they were trying to make. Similar findings with respect to limited information use and/or use of irrelevant cues comes from studies of State Registered Nurses (Shanteau et al., 1981), audit managers (Bamber, 1983), personnel selectors (Nagy, 1981), pathologists (Einhorn, 1974), stockbrokers (Slovic, 1960), and clinical psychologists (Goldberg, 1970). These findings seem somewhat at odds to what we might expect on the basis of both intuition and also the majority of the work on chess experts, some of which is reported above. What reasons might there be for such results other than merely that experts have no special competence in this respect?

One possible explanation for findings that very few cues are used by many experts is that in fact very few cues need to be used for good performance. Whilst much of the literature on the use of linear models has focused on the fact that simple linear models can usually outperform the expert judge (Meehl, 1954, 1986), it is also true that in many of these instances there is not much difference between the models and the experts (Camerer, 1981). This suggests that there is a low performance ceiling in the tasks studied (typically clinical diagnosis), thus experts could not do better even if they used more cues. The reason that the models usually outperform the experts is that they are more consistent with respect to the cues used and their relative weightings. This is particularly true of "bootstrap" models, where the cues and weights are derived from the judgment of the experts themselves. Further evidence for a low performance ceiling in clinical judgment comes from a review by Garb (1989) of over fifty comparisons of judgments by clinical psychologists and novices. The conclusion was that training has a significant effect on accuracy but experience virtually none, hence graduate students were as good as those who had been practising clinical psychologists for a number of years. Thus it seems that there is not much to learn in this domain.

This "low-performance ceiling" may be imposed for some of the reasons described above. In particular, there may be inadequate feedback about which information is relevant and which is not and/or the definition of a piece of information to which one might attend is fuzzy or constantly changing. Certainly in most of the domains listed above it would seem plausible that such a ceiling on performance exists although a thorough task analysis would be necessary in each case to establish this.

The poor performance of soil judges, however, does not seem to be easily explained in such terms. In this case there exist objective measures of soil analysis against which subjective judgment can be assessed. These objective analyses are not very hard to obtain so experts should have plenty of opportunity to calibrate their judgments. A possible alternative reason for the poor performance of the soil judges is that the irrelevant dimensions which they considered are not, in fact, irrelevant for many other frequently occurring analysis tasks. Thus the "irrelevant" dimensions may have been considered in order to rule-out these possibilities or perhaps just out of habit. It is perhaps worth mentioning here so-called, "broken-leg cues" (Meehl, 1954). These are pieces of information which are rare but if encountered instantly rule-out a number of routine interpretations of the information in front of you. Meehl's original example, from which the term comes, is that if you are trying to predict the movements of a person the knowledge that s/he has broken his or her leg affects seriously the prediction you would have otherwise made on the basis of the person's habits. It has been suggested that people are good at detecting broken-leg cues and that this gives them an advantage over statistical models induced from past examples. Studies of loan officers (Casey & Selling, 1986; Chalos, 1985) offer some support for this view (see also Blattberg & Hoch, 1989, and Johnson, 1988 for additional discussion of the role of broken-leg cues in human judgment compared to statistical models). Perhaps the "irrelevant" cues used by soil judges fulfil a similar function to broken-leg cues in that they permit certain radical alternatives to be selected between.

Camerer and Johnson (1991) argue that experts use configural rules which are not captured by linear models. Configural rules are non-linear because they encode interactions between different cues rather than treating them independently. Configural rules are easy to use and have plausible causal explanations. However, they are often faulty due to perceived, but illusory, correlations between cues. Also they are often based upon consideration of only very few data points and then over-generalized. Camerer and Johnson give an example of configural rule use from baseball. A normally average player named Bucky Dent played exeptionally well for the New York Yankees in one World Series and was made manager of the team on the basis that he could "come through when it mattered". In other words, "Dent" was not predictive of success but in interaction with "World Series", it was. Unfortunately, Dent's managerial spell was not a success because during this time the Yankees had their worst ever season. However, configural rules may have an advantage over linear rules in tasks not well described by linear models such as judgments of creditworthiness (for example, Whitred & Zimmer, 1985) although linear models are rather robust even in the face of quite major violations of linearity.

Finally, with respect to input and output processes, there is some evidence that experts may have enhanced capabilities in problem recognition (or "framing") and information search. For example, Selnes and Troye (1989) found that for purchasing decisions those with expert knowledge of the products to be considered sought out different types of information and devoted more effort to identifying the problem than non-experts. Selnes and Troy concluded that their experts attempted to frame the problem so as to efficiently serve their informational needs whereas the non-experts simply reacted to the stimuli in the order in which they were presented to them.

3.2 Central processes and knowledge

It is generally accepted that experts know more about their domain than non-experts. However, greater knowledge does not automatically result in improved performance—the knowledge could be incorrect or inappropriately applied. There are several studies which have demonstrated that the heuristics and biases found in students in the psychology laboratory can also be observed in the expert in the field. For example, Smith and Kida (1991) reviewed 25 studies of auditing and accounting professionals of varying levels of experience. These studies found evidence of use of the anchoring-and-adjustment heuristic (insufficient adjustment is made away from an initial, and possibly arbitrary, baseline "anchor"); miscalibration of probabilities (probability judgments do not match the true long-run frequences of the target events—usually the probability judgments are too high or "overconfident"); incoherence of probabilities (probabilities are assessed or combined in a manner which is inconsistent with the probability laws for example, the "conjunction fallacy" where the probability of the co-occurrence of two events is seen as being greater than the probability of either of the two events occurring by themselves); base-rate neglect (for example, the likelihood of someone with certain symptoms of having a particular disease is not adequately adjusted in the light of the incidence of the disease in the population); insensitivity to sample size (i.e. that information derived from a large sample is more reliable than that derived from a small); and confirmation bias (information is sought out in order to confirm rather than to reject hypotheses).

Although these biases have all been identified in professional auditors (and in some other experts, for example, estate agents: Northcraft & Neale, 1987; physicians: Christensen-Szalanski et al., 1983; and restaurateurs: Dube-Rioux & Russo, 1988) Smith and Kida concluded that many of them were mitigated by increased experience of the domain and/or by the use by the experimenters of familar, everyday (in contrast to artificial) tasks. This analysis is consistent with the idea mentioned above that many of the findings of poor expert judgment are due to lack of ecological validity of the tasks and methods used. However, there is also some reason to suspect that, in some domains, the more experienced someone is the more biased s/he might become (i.e. the exact reverse of what Smith & Kida propose). For example, Evans (1989) comments that the paediatricians who were involved in the Cleveland child abuse case (where an unprecedented number of children in the UK were diagnosed as having been sexually abused and subsequently were taken into care, thereby causing a public outcry) were caught in a vicious circle of overdiagnosis (for example, due to hypervigilance and the desire to avoid errors, see below) leading to an increase in estimates of prevalence of sexual abuse in the community resulting in even greater overdiagnosis, and so on.

In many jobs carried out by "experts" the results of their decisions can have serious consequences on the lives of others. This responsibility may often have a biasing effect on the judgments and decisions of such experts. Specifically, it is likely that the experts may err on the side of caution. Hence doctors may overdiagnose conditions on the basis that unnecessarily treating patients is better (for them, but not the hospital budget) than failing to treat a few patients who actually needed treatment. This argument could also be used to explain why doctors are unwilling to use statistical techniques such as linear models. These models may be recognized as being more accurate but—because they will be held responsible—doctors are unwilling to accept the "misses" that such systems are inevitably going to make. Katz (1984) went even further than this to suggest that by "donning a mask of infallibility" doctors are able to get patients to conform and also to maintain their professional control. However, a side-effect of appearing to be infallible may be that uncertainty is not dealt with in a rational manner. Certainly there is plenty of evidence of poor probabilistic reasoning by medics (for example, Eddy, 1982, for diagnosis of breast cancer; Christensen-Szalanski & Bushyhead, 1981, for diagnosis of pneumonia; and Dolan et al., 1986 for diagnosis of heart disease). I will return to a discussion of failures in probabilistic reasoning in experts shortly but it suffices to say here that such failures may be particularly prevalent in the context of medical diagnosis for the reasons given above and also as a consequence of medical training. For example, a number of medical textbooks explicitly warn the reader not to apply statistical analyses but to focus on each case individually (see, for example, Eddy, 1982).

A large number of studies of experts have been concerned with the quality of their probabilistic judgment. The reason for this focus is that from a normative perspective most decision-making can be cast as a trade-off between options in terms of their expected value (i.e. the product of the probability of that option occurring and its value or utility if it does occur). The application of expectancy-value principles is "normative" because it should result in optimal decision making (at least within the domains where such principles apply). A number of technologies are based on normative, expectancy-value models, mostly notably Decision Analysis. Probabilities are required as input to such technologies and mostly these probabilities are not available from historical data and therefore the subjective judgments of experts must be elicited. The quality of expert probability judgments is thus of crucial practical importance. It is also of interest to cognitive psychologists who wish to determine the extent to which normative models describe the way people actually make decisions.

The quality of probabilistic judgment is usually assessed in one of two ways. First, probability judgments are tested for consistency with the laws of probability theory (for example, as described by Bayes' Theorem, this requirement is often referred to as "coherence"). Second, judged probabilities are compared with the true relative frequencies of the target outcomes (for example, rain should fall on 70% of those days where a 0.7 probability of rain was assessed, this requirement is usually referred to as "calibration"). As with most of the expertise research so far described the picture regarding the quality of expert probability judgment is fairly mixed.

A number of studies have demonstrated incoherence in non-experts' probability judgments but as far as I am aware there are virtually no studies of coherence in experts. However, Eddy (1982) found that physicians misunderstood the relationship between marginal probabilities (for example, the probability of cancer, the probability of a positive test) and conditional probabilities (for example, the probability of a positive test given cancer). Schaefer et al. (1977) found that in two out of three tests, self-rated "experts" in soccer and statistics were slightly more coherent than non-experts. In a similar study of coherence in self-rated snooker "experts" we found the experts to be significantly more incoherent than non-experts on conditional probability questions, but significantly more coherent on transitivity questions (Wright et al., 1994). On closer examination of our task we observed that conditionalities imputed to independent events (for example, the outcome of two quarter-final matches) might have some validity. For example, knowledge about the outcome of one game might well affect the way players played the other game. In a study of restaurant managers, Dube-Rioux and Russo (1988) found failure to conform to the additivity axiom (mutually exclusive probabilities for example, rain and no rain should add to one) with respect to the disjunction and conjunction of restaurant failures.

Studies of expert calibration are more plentiful than those of expert coherence. Experts have been found to be just as overconfident as naive subjects in a number of instances, for example, clinical psychology (Oskamp, 1965), weather forcasting (Stael von Holstein, 1971), oddsmaking for basketball games (Yates & Curley, 1985). Further, expert calibration has in some cases been found to be worse than a uniform judge (i.e. one who gives the same probability response on every occasion and therefore can be regarded as having no knowledge): prediction of stock prices (Stael von Holstein, 1972), diagnosis of pneumonia (Christensen-Szalanski & Bushyhead, 1981) and diagnosis of heart disease (Dolan et al., 1986). On the other hand, in several instances, experts have been found to be very well-calibrated (for example, weather forecasting: Murphy & Brown, 1985; diagnosis of heart disease: Levi, 1986; forecasting the technical success of R&D projects: Balthasar et al., 1978; and predicting the number of tricks to be won in bridge: Keren, 1987). These examples provide what Wallsten and Budescu (1983) term an "existence demonstration". In other words, experts can make valid probability judgments under certain circumstances. What are these circumstances?

As I have suggested above and elsewhere (Bolger & Wright, 1994), the conditions for the existence of good probability judgment include: repeated events of the same type; the unambiguous occurrence of events; the availability of relevant and useable feedback about performance; and experience at using probabilities. These conditions do not apply in that many instances. In weather forecasting and prediction of the number of tricks to be won in bridge all these four conditions are satisfied (and good performance has been observed). However, more usually each event or case is somewhat different from another (for example, in medical diagnosis) and/or determining the occurrence of an event may be difficult (for example, the exact nature of some ailment may not be discovered until autopsy). Further, feedback is often not available. For example, in insurance underwriting we discovered that no claims made over five years after the policy was taken out were reviewed by underwriters (see Bolger et al., 1989). Even if there is feedback it may be influenced by the actions one takes. For instance, in governmental forecasting, policy decisions taken on the basis of forecasts may influence what actually happens in the world. In fact, such forecasts are often referred to as "projections" because they are targets rather than objective forecasts. Further, in most cases experts are not used to using numeric probabilities, and therefore may have difficulties expressing their feelings of uncertainty in the manner demanded by experimenters. Weather forecasters in the USA have been routinely using numeric probabilities for a number of years now and their performance has been observed to improve over this period (Murphy & Winkler, 1977, Murphy & Brown, 1985).

Knowledge of the rules of probability would also seem to be an important criterion for good probabilistic reasoning, especially with regard to the coherence requirement. However, it is unlikely that experts in most domains are any more sophisticated regarding the probability axioms than non-experts, given that they will have received no specific training. In this case we should anticipate that experts will make similar errors to non-experts in any task requiring complex probabilistic reasoning. However, as already indicated, the empirical evidence for this hypothesis is rather limited at present. There is some suggestion, though, that even expert statisticians may manifest biases on probabilistic tasks which they do not immediately recognize as being probabilistic. For example, Wagenaar and Keren (1986) found that professional statisticians were no better than lay people when predicting the probability of particular card combinations occurring in blackjack. This implies that training in the probability laws might not in itself be sufficient to ensure unbiased probabilistic reasoning in experts. In addition, they must also know how to apply these rules within the context of their daily work.

Experts may (or may not) know more but do they also have enhanced memory abilities relative to non-experts? Superior memory has been demonstrated for experts in a number of domains (for example, chess: Chase & Simon, 1973; de Groot, 1978; music notation: Sloboda, 1976; electronic circuit diagrams: Egan & Schwartz, 1979 and computer programming: McKeithen et al., 1981). The commonly-held view appears to be that this superiority is not due to actual increased capacity and/or speed of processing (which are assumed to be fixed as a consequence of our physiology) but instead is due to the ability to overcome processing and capacity limitations through improved strategies (for example, mnemonics) and organization (for example, chunking). This is the same argument as for the enhanced perceptual abilities which I described earlier. It should be noted that the strategies and methods of organization may be explicitly represented and applied by experts, as in the case of the long-distance runner who used race times to memorize long sequences of digits (Chase & Ericsson, 1981). However, in other cases, such as the chunking of chess pieces, experts may not always have conscious access to the knowledge underlying their enhanced abilities and therefore be unable to verbalize how they do what they do. In the latter case, the strategies and organizing principles will have to be inferred from controlled experiments rather than through some kind of verbal protocol.

There seems to be little evidence for inherited differences in cognitive abilities in general, or memory in particular, being implicated in expert performance, except perhaps personality traits such as motivation (for example, Galton, 1869). IQ appears not to be very well correlated with levels of performance in the arts and sciences (Tyler, 1965), and there is little evidence of stable individual differences in memory ability (Kelley, 1964). Moreover, specific memory ability and IQ are poorly related—the high performance of idiot savant "calendrical calculators" being a spectacular example of this. Further, good performance on particular memory tasks can be attained after relatively short periods of practice (for example, Ericsson, 1985). However, for complex tasks such as chess many, many hours of practice are required to achieve the highest levels of performance (Simon & Chase, 1973, estimated that approximately 30,000 hours of practice are required to become a chess master).

Do experts learn any differently than do non-experts? Anderson (1982) has proposed that the learning of experts may be different from non-experts. He suggests a progression from cognitively effortful attempts to understand tasks, through rehearsal and restructuring so as to make cognitive processes more efficient, to automatic, stimulus-driven performance requiring relatively little cognitive effort. Although this account seems plausible from the research, both individual and task differences affecting learning make it difficult to provide strong support for this proposal. For example, some people appear not to learn as well as others from the same experience. This was the case in a study of waiters' and waitresses' strategies for remembering orders where one waiter stood out by far from his colleagues (Ericsson & Polson, 1988). In the case of domain differences, I have already mentioned the problems of lack and quality of feedback for learning, and also of differences between domains in terms of their complexity and dynamism. In some domains it might also not be possible to incrementally acquire and refine skills. For example, there is evidence from sports that nonoptimal movements cannot be shaped into optimal. Instead bad learning must be undone and the optimal movement acquired if expert levels are to be achieved. A further problem with studying learning in experts is that they usually try to avoid unfamiliar tasks (where learning might be manifest). Thus learning could be regarded as an indication that the experimental task has been chosen badly.

Is expertise transferable to other domains? As mentioned earlier, expertise seems to be very domain-specific, for example Anderson (1990) concluded that "Chess experts do not appear to be better thinkers for all their genius in chess". However, there are also findings which suggest that experts may be able to generalize their skill to similar but simpler tasks such as the "schematic" tasks (where certain features of the real task are missing) often used in laboratory experiments. For example, Bennet (1983) replicated the finding that expert cocktail waiters and waitresses have a superior ability to remember drinks orders than non-experts using a simulated situation where dolls were substituted for customers.

With respect to expert novice differences in problem solving and reasoning, it has been found that experts tend to retrieve a solution as part of their comprehension of the task whereas non-experts have to construct a representation of the task and then generate a step-by-step solution (for physics problems: for example, Chi et al., 1982; and algebra word problems: Hinsley et al., 1977). Further Chi et al. found that experts categorized physics problems according to principles of physics whereas novices categorized in terms of specific features mentioned in the statement of the problem. This is evidence of concept versus data-driven strategies being used by experts and novices respectively. In contrast, Patel and Groen (1991) found that medical experts use "forward reasoning" (diagnoses are based on an examination of the symptoms) whereas medical students tend to use "backward reasoning" (symptoms are sought to confirm or disprove some hypothesis). Greater use of forward reasoning was strongly associated with number of correct diagnoses. Patel and Groen argue that experts can use forward reasoning because their extensive specialist knowledge allows them easily to match symptoms to diseases. In certain tasks, such as the diagnosis tasks used by Patel and Groen, this is a highly efficient and accurate method. In other situations, such as ones where there are lots of irrelevant cues, backward reasoning is more effective. Also backward reasoning must be used when little or no specialist knowledge can be brought to bear, as is the case with non-experts or with experts faced with unfamiliar tasks.

Murphy and Wright (1984), however, found that clinical psychologists' categorization of affective types was less distinctive than novices' (i.e. experts would allocate a type to several categories whereas novices would allocate a type to only a single category). Murphy and Wright attribute this rather surprising finding to the increased realization by experts of the similarities between different concepts.

Experts have in some cases been found to plan further ahead than non-experts (for example, bridge: Charness, 1989; medicine: Patel & Groen, 1991; physics: Anzai, 1991). However, such planning advantages have not always been found for chess (de Groot, 1978; Holding, 1985), but this may be because the tasks are not really designed to test depth of search (i.e. time constraints emphasize efficient rather than deep search). Chess masters can search very deeply if given unlimited time (Ericsson & Staszeweski, 1989).

Very little research has been conducted into expert-novice differences in meta-knowledge. However, Chi et al. (1987) used verbal-protocol techniques to investigate whether good performance in solving physics problems was associated with more elaborate explanations of solution processes. It was found that the better students tended to generate the most self-monitoring statements. Further, the explanations of the highest performers contained many more justifications in terms of conditions, actions and goals than did the explanations of the lowest performers and also contained more naming of quantitative expressions. In certain cases, however, we might expect less meta-knowledge in experts than novices. As I have already mentioned, in some domains reasoning appears to become more automatic with experience and thereby less accessible to consciousness.

Shanteau (1987) argues that experts often display certain personality characteristics including: good communication skills, the ability to adapt to new situations, and a sense of responsibility. However, the evidence for this assertion is somewhat anecdotal. Shanteau also argues that experts manifest strong self-confidence. As already indicated when discussing probabilistic reasoning, there is evidence to suggest that confidence is indeed a trait of experts but may not always be justified by equivalent performance. I suggested that overconfidence might occur as a result of a desire to avoid mistakes (or being seen to make mistakes) coupled with an over-narrow focus on a particular problem. It has also been proposed that medics might appear confident in order to ensure compliance from their patients. Further, in the calibration literature it is often observed that the easier a task is, the more people tend to display overconfidence (see Lichtenstein et al., 1982). This implies that experts, who presumably find tasks within their domains particularly easy, may be more overconfident than non-experts.

Finally, Catell (1963) investigated whether the personality profiles of leading science researchers are any different from those of teachers or administrators in the same fields. Catell concluded that the expert group was more reflective and emotionally unstable than the non-expert group but also more self-sufficient and dominant. Interestingly he also concluded that the experts were more introverted than non-experts which conflicts somewhat with what I have just written regarding confidence in experts. One explanation is that researchers may not have to project the same self-confidence as those such as medics who have to deal constantly with the public. It is also necessary to contrast outwardly directed confidence (as in self-presentation) with inwardly directed confidence (for example, certainty in ones judgments). It can be seen then that some aspects of personality are strongly implicated in output processes, particularly with respect to communication and self-presentation.

4 Implications for knowledge engineering

I will now discuss in turn the implications for knowledge engineering of cognitive expertise research findings in each area of competence outlined in Section 1.

4.1 Input and output processes

Capturing and modelling enhanced perceptual and/or motor skills may not be a worthwhile activity because they appear, on the whole, to be mechanisms for overcoming limitations of the human body and nervous system. On the other hand, there may be lessons to be learnt from the study of these mechanisms. For example, we may learn how to organize input and output so as to increase efficiency of processing in a particular domain. This may be of use for optimizing systems for certain applications, for example, those requiring responses in real-time.

The work on experts' ability to select, attend to, or "filter" relevant information from irrelevant also suggests that modelling expert competence per se may not be the most profitable course of action. At least this may be true in tasks where linear models work well, such as in clinical diagnosis, or any task where predictions need to be made on the basis of a number of variables observable over time. These variables do not even need to be reliably measured over time to justify the use of linear models. The work on bootstrapping shows that elicitation of cues, and weights for these cues, from an expert and then incorporation of this information into a linear model produces performance at least on par with the expert, and often a little bit better. Of course, bootstrapping is a form of knowledge engineering, although expert knowledge is being shoe-horned into a processing scheme which is almost certainly not the one being used by the experts themselves. A different knowledge-engineering scheme might be favoured, however, if task analyses reveal a significant role for broken-leg cues and/or serious violations of linearity such as multiplicative interactions between variables. In these cases, modelling of the expert's specialist knowledge and configural rules should lead to improved performance relative to a linear model (perhaps linear models and expert systems could profitably be used in conjunction for some tasks). It must be stressed again, though, that linear models are fairly robust and hence can be quite widely applied.

4.2 Central processes and knowledge

The emerging conclusion of research suggests that heuristics and biases may not be as widespread amongst professionals as initially thought on the basis of laboratory work with students. In many cases where biases have been demonstrated in experts, it is possible to identify problems of ecological validity in the experimental procedure. Nonetheless, knowledge-engineers should be mindful that experts may use heuristics and these may result in biases in some situations (note that heuristics can also lead to improved performance—I will return to this point shortly). Biases are most likely to be manifest in situations where there is a low-ceiling on expert performance due to the factors listed in Section 2.1 or because of professional practices such as those of defensive medicine. If heuristic use is identified then attempts should be made to establish how likely the heuristic is to result in good performance and how likely bad. When we (Bolger et al., 1989) were developing an expert system to assist in the training of life insurance underwriters we discovered that our expert underwriters were using a form of the representativeness heuristic, in this case stereotyping, to make some of their judgments. One memorable example was that if a male applicant stated his occupation as being a hairdresser then the underwriter would suspect the applicant of being homosexual and therefore at high risk from HIV infection. A rather intrusive "lifestyle" questionnaire would then be sent to this person. Although there may be some correlation between male hairdressers and homosexuality it is not clear that this relationship is strong enough to justify potentially losing the customer by sending them the questionnaire. Further, the evidence at the time was that homosexuals had most heeded the "safe sex" message and consequently may be at less risk than heterosexuals who had not changed their behaviour (in the US, figures were emerging to suggest that young heterosexual women were at highest risk). We decided against building this particular heuristic into our system even though our informal analysis of its validity may have been incorrect. However, this example illustrates that care must be taken not to build expert knowledge into a system uncritically as there is a danger that biases (and indeed prejudices) might then become reified within the system.

The studies of probabilistic reasoning suggest that usually experts have no special competence in this area. This indicates that caution should be exercised in eliciting numeric probabilities from experts for input into knowledge-based or normative systems. There are several reasons why experts may have difficulties with numeric probabilities but lack of knowledge about probability theory, limited practice at applying this theory, and simple lack of experience at expressing uncertainty as numbers, are all important factors. Also of importance is to be an expert in a domain where probabilities are meaningful (for example, there are repeated events of the same type) and where there is appropriate feedback for calibrating these probabilities. Where these conditions are fulfilled then experts' probabilities should be reliable (but in these instances it is likely that probabilities could equally be derived from actuarial sources). Where conditions are less favourable but probabilities are still required as input to some system then there are strategies which can be used to ensure that these probabilities are as reliable as possible:

- use some metric which is more familiar to the expert than numeric, point probabilities for example, odds or verbal labels (see, for example, Kahneman et al., 1982; Teigen, 1994).
- if complex probabilities are required then elicit simpler probabilities and derive the more complex target probabilities by combining using the probability laws (see, for example, Edwards et al., 1968; Bolger & Wright, 1992)
- use techniques whereby elicited probabilities are represented to the expert so that they can see the practical implications of the numbers they have given (for example, by using them in a simulation); computer technologies such as Influence Diagrams can help with this (see, for example, Bunn, 1992)
- use more than one expert to produce aggregate probabilities (see, for example, Rowe, 1992) or to agree on some values through, for example, a decision-conferencing process (see, for example, Reagan-Cirincione & Rohrbaugh, 1992)

Note that these procedures are not just restricted to the elicitation of probabilities but can be usefully applied elsewhere, particularly for capturing quantitative information.

Implicit aspects of improved memory and representation learnt through lots of practice with exemplars in a domain may not be easily accessed verbally by experts. An alternative approach for capturing this sort of knowledge may be to use a neural network rather than traditional knowledge-elicitation techniques. However, neural networks are not necessarily the panacea they might seem to be as their success is highly dependent on being provided with (lots of) the right sort of training. Neural networks are usually quite good at forgetting things also. Explicit memory strategies, such as mnemonics, should be more easily identified and could reasonably be expressed as rules within a knowledge-based system. However, much depends upon whether one wants to model experts or to develop some system as good as or better than the expert using whatever strategies possible (i.e. not necessarily the ones used by the experts). If one has the latter goal then, since both implicit and explicit memory strategies appear to be directed at overcoming processing limitations to which computers are not necessarily subject, then capturing the expert memory and representation competence may not be a worthwhile enterprise. On the other hand, brute-force strategies (presumed not to be used by experts) have not yet allowed the creation of a computer chess program which can consistently beat the best human players using heuristics. Thus heuristic-use can be a very powerful strategy in the right context.

Given that a huge number of hours practice seem to be required to acquire the highest levels of perceptual and memorial competence a potentially useful knowledge-engineering exercise might be to provide decision-support systems in the form of external memory aids and devices to assist in the organization of information. One problem with this approach is that in certain domains there might not be much else to the expert's task than being able to remember a lot and/or to recognize things quickly (for example, is expert chess performance "simply" a case of remembering good moves associated with particular configurations of pieces?). If this is the case then perhaps it is better to replace the expert entirely (there would not be much point doing this in chess, of course!).

Obviously, any attempt to replace all, or a good deal, of a person's expertise with a computer system is likely to meet with considerable opposition from the expert. The provision of explicit knowledge (by which I mean verbalizable) in an advice-giving system may be more acceptable to experts than some black box containing implicit knowledge (by which I mean non-verbalizable and not necessarily available to consciousness) which simply provides the "correct answer". One function that experts could retain (which is difficult to make computers do well) is as an intelligent front-end to some knowledge-based or normative system. In this way the experts can utilise their enhanced communication and social skills and not have to worry about engaging in years of expensive training or practice intended at honing-up their cognitive skills (which, as we have seen, may not ultimately be all that honed-up anyway). For example, it is quite easy to envisage the General Practitioner of the future who is selected and trained for counselling skills, and gathering of data about symptoms, but with no medical training *per se*. All medical knowledge would instead be available from a knowledge-based system.

The implications for knowledge-engineering from studies of learning and transfer seem to be twofold. First, expert competence may often be automatic and therefore not easily expressed, for example, through verbal protocols. Instead inferences may have to be made from observations of expert behaviour on typical and untypical tasks. Another strategy may be to look at sub-experts who may have expert competence but still have access to how they are performing tasks. However, from the Chi et al. (1987) study of physicists it seems that not all experts have less insight than non-experts into their strategies and in some cases they may have more. Second, expert knowledge is very specific to a particular task domain. This implies that any expert task to be modelled should be analysed carefully so as to determine its key features. Any knowledge elicited from experts should be generalizeable to, and only to, any other task containing only the same key features.

Regarding problem-solving and reasoning, the expertise research suggests that there are expert-novice differences in competence that might be worth modelling from a knowledge-engineering point of view. However, the finding that in some cases experts prefer top-down strategies (as in categorization tasks) and at other times appear to prefer bottom-up strategies (as in diagnosis tasks) demonstrates the importance of the context within which problem-solving occurs. Perhaps a distinguishing feature of expertise is the possession of a large repertoire of problem-solving strategies which one can apply flexibly and appropriately to suit the context. This context sensitivity has so far proved rather difficult to emulate in knowledge-based systems but case-based reasoning systems show some potential here, perhaps because they model some of the more genuine aspects of experts' enhanced competence.

5 Summary and conclusions

To summarize, experts have demonstrated superior performance relative to non-experts thereby implying enhanced competence in all the areas specified in Section 1. However, the research by no means implies that all experts are always more competent in all respects. I have suggested that in some domains it is not possible to learn very much if anything, hence any bright person with the appropriate presentation skills but no domain knowledge should be able to perform at a level with someone with several years experience in the domain. I have also suggested that in some domains experts may even perform worse than non-experts in some respects. This may occur, for example, as a consequence of practical contingencies of the experts' jobs, such as avoiding litigation. There also seem to be some domains which demand competence(s) that may be difficult for people to acquire even if the conditions for learning are available. For example, those domains where multiple cues must be weighted or which require complex probabilistic reasoning. Unfortunately, the picture regarding the precise relationship between expert competence and the nature of task domains is difficult to obtain due to various methodological problems such as defining and identifying experts and non-experts, designing tasks which are controlled yet ecologically valid, and obtaining analyzable responses from the subjects. One thing at least seems clear. That is that expertise is acquired rather than inherited (with the possible exception of certain personality traits needed to be successful and to persevere in acquiring expertise) and that it usually takes many years of practice to attain the highest levels of performance and underlying competence.

Where does all this leave us from a knowledge-engineering perspective? In general terms it tells us that, under certain circumstances, experts have a special competence that is worth modelling or capturing. It also tells us that, under other circumstances, it may not be such a good idea to follow this approach and that a better approach is to replace or support expert judgment, perhaps with normative/statistical systems. In yet other circumstances it tells us that it might be better not to engage in knowledge-engineering at all since the conditions are such that neither knowledge-based nor normative systems are likely to be successful. It further suggests which are the circumstances to which these three approaches are likely to be suited although further empirical research and, just as important, task analyses, are required to clarify the picture we already have. Finally, it indicates that, since expertise is largely acquired, knowledge and its organization are key to expert performance. Hence knowledge-engineering is a worthwhile activity. If it had been the case that inherited differences accounted for most of the variance in expert performance then we might have suspected that acquired knowledge does not play much of a role in expertise relative to, for example, speed and efficiency of neurological functioning. In such a scenario the case for modelling expert knowledge for system design would be harder to defend (although there would still be clear arguments for knowledge-engineering for cognitive ergonomics and cognitive psychology). Also, because humans take a long time to acquire this knowledge it is a useful thing to design systems which may speed-up this process. Expertise should consequently be rendered a less rare commodity thereby reducing the dangers of abuse of power and the incompetence and/or biases of particular individuals.

References

Anderson, JR, 1982. "Acquisition of cognitive skill" *Psychological Review* **89** 369–406.

Anderson, JR, 1990. *Cognitive Psychology and its Implications (3rd ed.),* Freeman.

Anzai, Y, 1991. "Learning and use of representations for physics expertise" In: KA Ericsson and J Smith (eds.), *Toward a General Theory of Expertise,* pp 64–92, Cambridge University Press.

Balthasar, HU, Boschi, RAA and Menke, MM, 1978. "Calling the shots in R and D" *Harvard Business Review* May–June, 151–160.

Bamber, EM, 1983. "Expert judgment in the audit team: A source reliability approach" *Journal of Accounting Research* **21** 396–412.

Bennet, HL, 1983. "Remembering drink orders: The memory skill of cocktail waitresses" *Human Learning* **2** 157–169.

Blattberg, RC and Hoch, SJ, 1989. "Database models and managerial intuition: 50% model and 50% manager". Report from the Center for Decision Research, Graduate School of Business, University of Chicago.

Bolger, F and Wright G, 1992. "Reliability and validity in expert judgment" In: G Wright and F Bolger (eds.), *Expertise and Decision Support*, pp 47–76, Plenum.

Bolger, F and Wright, G, 1994. "Assessing the quality of expert judgment: Issues and analysis" *Decision Support Systems* **11** 1–24.

Bolger, F, Wright, G, Rowe, G, Gammack, J and Wood, R, 1989. "LUST for life: Developing expert systems for life assurance underwriting" In: N Shadbolt (ed.), *Research and Development in Expert Systems VI*, pp 128–139, Cambridge University Press.

Bunn, D, 1992. "Synthesis of expert judgment and statistical forecasting models for decision support" In: G Wright and F Bolger (eds.), *Expertise and Decision Support*, pp 251–268, Plenum.

Camerer, CF, 1981. "The validity and utility of expert judgment." Unpublished doctoral dissertation, University of Chicago.

Camerer, CF and Johnson, EJ, 1991. "The process-performance paradox in expert judgment: How can experts know so much and predict so badly?" In: KA Ericsson and J Smith (eds.), *Toward a General Theory of Expertise,* pp 195–217, Cambridge University Press.

Casey, C and Selling, TI, 1986. "The effect of task predictability and prior probability disclosure on judgment quality and confidence" *The Accounting Review* **61** 302–317.

Catell, RB, 1963. "The personality and motivation of the researcher from measurements of contemporaries and from bibliography" In: CW Taylor and F Barron (eds.), *Scientific Creativity: Its Recognition and Development,* pp 119–131, Wiley.

Chalos, P, 1985. "The superior performance of loan review committee" *Journal of Commercial Bank Lending* **68** 60–66.

Charness, N, 1989. "Expertise in chess and bridge" In: D Klahr and K Kotovsky (eds.), *Complex Information Processing: The Impact of Herbert A. Simon,* pp 183–208, Erlbaum.

Chase, WG and Ericsson, KA, 1981. "Skilled memory" In: JR Anderson (ed.), *Cognitive Skills and their Acquisition,* pp 141–189.

Chase, WG and Simon, HA, 1973. "The mind's eye in chess" In: WG Chase (ed.), *Visual Information Processing,* pp 215–281, Academic Press.

Chi, MTH, Bassok, M, Lewis, MW, Reimann, P and Glaser, R, 1987. "Self-explanations: How students study and use examples in learning to solve problems" *Cognitive Science* **13** 145–182.

Chi, MTH, Glaser, R and Rees, E, 1982. "Expertise in problem solving" In: RS Sternberg (ed.), *Advances in the Psychology of Human Intelligence (Vol. 1),* pp 1–75, Erlbaum.

Chi, MTH, Glaser, R and Farr, MJ (eds), 1988. *The Nature of Expertise,* Erlbaum.

Christensen-Szalanski, JJJ and Bushyhead, JB, 1981. "Physicians' use of probabilistic information in a real clinical setting" *Journal of Experimental Psychology: Human Perception and Performance* **7** 928–935.

Christensen-Szalanski, JJJ, Beck, DE, Christensen-Szalanski, CM and Koepsell, TD, 1983. "Effects of expertise and experience on risk judgments" *Journal of Applied Psychology* **68** 278–284.

de Groot, A, 1978. *Thought and Choice in Chess,* Mouton. (Originally published in 1946.)

Dolan, JG, Bordley, DR and Mushlin, AI, 1986. "An evaluation of clinicians' subjective prior probability estimates" *Medical Decision Making* **6** 216–223.

Dowie, J and Elstein, A, 1988. *Professional Judgement: A reader in clinical decision making,* Cambridge University Press.

Dube-Rioux, L and Russo, JE, 1988. "An availability bias in professional judgment" *Journal of Behavioural Decision Making* **1** 223–237.

Ebbesen, E and Konecni, V, 1975. "Decision making and information integration in the courts: The setting of bail" *Journal of Personality and Social Psychology* **32** 805–821.

Eddy, DM, 1982. "Probabilistic reasoning in clinical medicine: problens and opportunities" In: D Kahneman, P Slovic and A Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases,* Cambridge University Press.

Edwards, W, Phillips, LD, Hays, WL and Goodman, BC, 1968. "Probabilistic information processing systems" *IEEE Transactions on Systems Science and Cybernetics* **4** 248–265.

Egan, DE and Schwartz, BJ, 1979. "Chunking in recall of symbolic drawings" *Memory and Cognition* **7** 149–158.

Einhorn, HJ, 1974. "Expert judgment: Some necessary conditions and an example" *Journal of Applied Psychology* **59** 562–571.

Ellis, SH, 1973. "Structure and experience in the matching and reproduction of chess patterns." Unpublished doctoral dissertation, Carnegie-Mellon University.

Ericsson, KA, 1985. "Memory skill" *Canadian Journal of Psychology* **39** 188–231.

Ericsson, KA and Polson PG, 1988. "Memory for restaurant orders" In: M Chi, R Glaser and M Farr (eds.), *The Nature of Expertise,* pp 23–70, Erlbaum.

Ericsson, KA and Smith, J, 1991. "Prospects and limits of the empirical study of expertise: An introduction" In: KA Ericsson and J Smith (eds.), *Toward a General Theory of Expertise,* pp 1–38, Cambridge University Press.

Ericsson, KA and Staszeweski, J, 1989. "Skilled memory and expertise: Mechanisms of exceptional performance" In: D Klahr and K Kotovsky (eds.), *Complex Information Processing: The Impact of Herbert A. Simon,* pp 183–208, Erlbaum.

Evans, J, St, BT, 1989. "Some causes of bias in expert opinion" *The Psychologist* **2** 112–113.

Gaeth, GJ and Chanteau, J, 1984. "Reducing the influence of irrelevant information on experienced decision makers" *Organizational Behavior and Human Performance* **33** 263–282.

Galton, F, 1869. *Hereditary Genius,* Macmillan.

Goldberg, LR, 1970. "Man versus model of man: A rationale, plus some evidence, for a method of improving clinical inferences" *Psychological Bulletin* **73** 422–432.

Hinsley, DA, Hayes, JR and Simon, HA, 1977. "From words to equations: Meaning and representation in algebra word problem" In: MA Just and PA Carpenter (eds.), *Cognitive Processes in Comprehension,* pp 89–108, Erlbaum.

Holding, DH, 1985. *The Psychology of Chess Skill,* Erlbaum.

Johnson, EJ, 1988. "Expertise and decision under uncertainty: Performance and process" In: M Chi, R Glaser and M Farr (eds.), *The Nature of Expertise,* Erlbaum.

Kahneman, D, Slovic, P and Tversky, A (eds.), 1982. *Judgment under Uncertainty: Heuristics and Biases,* Cambridge University Press.

Katz, J, 1984. "Why doctors don't disclose uncertainty" *Hastings Centre Report* **14** 35–44.

Kelley, HP, 1964. "Memory abilities: A factor analysis" *Psychometric Society Monographs* **11** 1–53.

Keren, G, 1987. "Facing uncertainty in the game of bridge: A calibration study" *Organizational Behaviour and Human Decision Processes* **39** 98–114.

Levi, KR, 1986. "Numerical likelihood estimates from physicians and linear models." Unpublished doctoral dissertation, University of Michigan.

Lichtenstein, S, Fischhoff, B and Phillips, LD, 1982. "Calibration of probabilities: The state of the art to 1980" In: D Kahneman, P Slovic and A Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases,* Cambridge University Press.

McKeithen, KB, Reitman, JS, Rueter, HH and Hirtle, SC, 1981. "Knowledge organization and skill differences in computer programmers" *Cognitive Psychology* **13** 307–325.

Meehl, PE, 1954. *Clinical versus Statistical Prediction. A Theoretical Analysis and Review of the Evidence,* University of Minnesota Press.

Meehl, PE, 1986. "Causes and effects of my disturbing little book" *Journal of Personality Assessment* **50** 370–375.

Milojovic, JD, 1982. "Chess imagery in novice and master" *Journal of Mental Imagery* **6** 125–144.

Mumpower, JL, Phillips, LD, Renn, O and Uppuluri, VRR (eds.), 1987. *Expert Judgment and Expert Systems,* Springer-Verlag.

Murphy, AH and Brown, BG, 1985. "A comparative evaluation of objective and subjective weather forecasts in the United States" In: G Wright (ed.), *Behavioral Decision Making,* Plenum.

Murphy, AH and Winkler, RL, 1977. "Reliability of subjective probability forecasts of precipitation and temperature" *Applied Statistics* **26** 41–47.

Murphy, GL and Wright, JC, 1984. "Changes in conceptual structure with expertise: Differences between real-world experts and novices" *Journal of Experimental Psychology: Learning, Memory and Cognition* **10** 144–155.

Nagy, GF, 1981. "How are personnel selection decisions made? An analysis of decision strategies in a simulated personnel selection task." Unpublished doctoral dissertation, Kansas State University.

Northcraft, MA and Neale, GB, 1987. "Experts, amateurs and real-estate: An anchoring and adjust perspective in property pricing decisions" *Organizational Behavior and Human Decision Processes* **39** 84–97.

Oskamp, S, 1965. "Overconfidence in case-study judgments" *Journal of Consulting Psychology* **29** 261–265.

Patel, VL and Groen, GL, 1991. "The general and specific nature of medical expertise: A critical look" In: KA Ericsson and J Smith (eds.), *Toward a General Theory of Expertise,* pp 93–125, Cambridge University Press.

Reagan-Cirincione, P and Rohrbaugh, J, 1992. "Decision conferencing: A unique approach to the behavioral aggregation of expert judgment" In: G Wright and F Bolger (eds.), *Expertise and Decision Support,* pp 181–202, Plenum.

Rowe, G, 1992. "Perspectives on expertise in the aggregation of judgment" In: G Wright and F Bolger (eds.), *Expertise and Decision Support,* pp 155–180, Plenum.

Saariluoma, P, 1990. "Chess players' search for task relevant cues: Are chunks relevant?" In: D Brogan (ed.), *Visual Search,* pp 115–121, Taylor & Francis.

Salthouse, TA, 1984. "Effects of age and skill in typing" *Journal of Experimental Psychology: General* **113** 345–371.

Salthouse, TA, 1985. "Anticipatory processing in transcription typing" *Journal of Applied Psychology* **70** 264–271.

Schaefer, RE, Borcherding, K and Laemmerhold, C, 1977. "Consistency of future event assessments" In: H Jungermann and G de Zeeuw (eds.), *Decision Making and Change in Human Affairs,* pp 331–345, Reidel.

Schwartz, S and Griffin, T, 1986. *Medical Thinking: The Psychology of Medical Judgment and Decision Making,* Springer-Verlag.

Selnes, F and Troye, SV, 1989. "Buying expertise, information search and problem solving" *Journal of Economic Psychology* **10** 411–428.

Shanteau, J, Grier, M, Johnson, J and Berner, E, 1981. "Improving decision making skills of nurses" In: *ORSA-TIMS Proceedings,* Houston, Texas: ORSA-TIMS.

Shanteau, J, 1987. "Psychological characteristics of expert decision makers" In: JL Mumpower, O Renn, LD Phillips and VRR Uppuluri (eds.), *Expert Judgment and Expert Systems,* Springer-Verlag.

Simon, HA and Chase, WG, 1973. "Skill in chess" *American Scientist* **61** 394–403.

Sloboda, J, 1976. "Visual perception of musical notation: Registering pitch symbols in memory" *Quarterly Journal of Experimental Psychology* **28** 1–16.

Slovic, P, 1969. "Analyzing the expert judge: A descriptive study of a stockbroker's decision processes" *Journal of Applied Psychology* **53** 255–263.

Smith, JF and Kida, T, 1991. "Heuristics and biases: Expertise and task realism in auditing" *Psychological Bulletin* **109** 472–489.

Stael von Holstein, CS, 1971. "An experiment in probabilistic weather forecasting" *Journal of Applied Meteorology* **10** 635–645.

Stael von Holstein, CS, 1972. "Probabilistic forecasting: An experiment related to the stock market" *Organizational Behaviour and Human Performance* **8** 139–158.

Teigen, KH, 1994. "Variants of subjective probabilities: concepts, norms and biases" In: G Wright and P Ayton (eds.), *Subjective Probability,* pp 211–238, Wiley.

Tyler, LE, 1965. *The Psychology of Human Differences,* Appleton-Century-Crofts.

Wagenaar, WA and Keren, GB, 1986. "Does the expert know? The reliability of predictions and confidence ratings in experts" In: E Hollnagel, G Mancini and DD Woods (eds.), *Intelligent Decision Support in Process Environments,* Springer-Verlag.

Waghorn, K, 1988. "Chess players' use of task-specific processes in a perceptual classification task." Unpublished honors thesis, University of Waterloo.

Wallsten, TS and Budescu, DV, 1983. "Encoding subjective probabilities: a psychological and psychometric review" *Management Science* **29** 151–173.

Whitred, G and Zimmer, I, 1985. "The implications of distress prediction models for corporate lending" *Accounting and Finance* **25** 1–13.

Wright, G and Bolger, F (eds.), 1992. *Expertise and Decision Support,* Plenum.

Wright, G, Rowe, G, Bolger, F and Gammack, J, 1994. "Coherence, calibration and expertise in judgmental probability forecasting" *Organizational Behavior and Human Decision Processes* **57** 1–25.

Yates, JF and Curley, SP, 1985. "Conditional distribution analysis of probabilistic forecasts" *Journal of Forecasting* **4** 61–73.