The Probabilistic Retreat From Biases: Implications for Man-Machine Communication?

Mike Oaksford [1] and Nick Chater [2]

[1] Department of Psychology, University of Wales at Bangor, Bangor, Gwynedd, LL57 2DG, UK. (E-mail: PSSØ27@bangor.ac.uk OR mike@cogsci.ed.ac.uk)
[2] Department of Psychology, University of Edinburgh, 7, George Square, Edinburgh, EH8 9JZ, UK. (E-mail: nicholas@cogsci.ed.ac.uk)

Abstract A recent trend within cognitive psychology has been the move away from the depiction of human reasoning as errorful and prone to bias and towards more appropriate probabilistic models of the behavioural data. This paper reviews this trend in three areas - probabilistic reasoning, causal reasoning and hypothesis testing - and discusses some possible implications for user modelling in man-machine communication.

Key Words: Reasoning, psychology, probabilistic reasoning, causal reasoning, hypothesis testing, heuristics, biases, man-machine communication

1 Introduction

Until recently the psychological study of reasoning has been interpreted to reveal that untutored human reasoners fall far short of the standards of rationality provided by normative theories. In laboratory tasks normal adults appear to reveal a variety of apparently irrational and systematic biases from the prescriptions of logic and mathematics (see, Evans, 1989, or Evans, Newstead & Byrne, 1993, for overviews). Theoretical responses to these findings have included appeal to short cut processing strategies, i.e., heuristics, and the use of special representational formats, i.e., mental models. Over the last few years, however, it has become clear that many of the tasks taken to reveal biases are susceptible to more rational interpretations. It seems that the wrong logico-mathematical frameworks may have been taken to define rational behaviour in these tasks and that subjects may use probabilistic strategies that are wholly consistent with mathematical probability theory.

An awareness of these developments may be important to user modelling in man-machine communication. According to Norman and Draper (1986), in designing profitable human-computer interactions it is important that a task be "presented so that it matches human skills." The skills to be matched invariably involve some element of reasoning. Hence in constructing methodologies for interacting with computers a minimal requirement may be an understanding of how people reason. Moreover, on briefly scanning the contents of some recent edited collections in this area we were struck by how many made appeal to the theoretical constructs - such as heuristics and mental models - that have become popular as a result of work on biases in human reasoning. To the extent that the invocation of such constructs is being re-thought in the cognitive psychology of reasoning, a similar re-evaluation may be necessary in the area of man-machine communication.

In this chapter we discuss three areas where a probabilistic re-interpretation has been successful: probabilistic reasoning, causal reasoning, and hypothesis testing. We first introduce each area, outlining the principle tasks used. We then discuss the new probabilistic accounts of these data due, in the case of probabilistic reasoning, to Birnbaum (1983) and Gigerenzer and Murray (1987), in the case of causal reasoning, to Cheng and Novick (1990), and in the case of hypothesis testing, to Oaksford and Chater (Oaksford & Chater, 1993).

2 Biases in Human Reasoning

2.1 Probabilistic Reasoning and Base Rate Neglect

Perhaps one of the most well known examples of bias in the reasoning literature concerns Kahneman and Tversky's (1972) demonstration that subjects tend to ignore base rate information. We illustrate this using Tversky and Kahneman's (1980) cabs problem. Subjects were provided with the following information:

> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are give the following data:
> (i) 85% of the cabs in the city are Green and 15% are Blue
> (ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.
> *Question*: What is the probability that the cab involved in the accident was Blue rather than Green.

In terms of Bayes' theorem the probability of the cab being blue given the witnesses testimony ($P(Blue|"Blue")$) is,

$$P(Blue|"Blue") = \frac{P("Blue"|Blue)P(Blue)}{P("Blue")}$$

That is the likelihood that the witness correctly identifies the cab

($P("Blue"|Blue)$) times the ratio of the base rates of Blue cabs in the city and the witness giving the evidence "Blue." The probability $P("Blue")$ can be calculated by conditioning on the probabilities of the cab being Green or Blue (since these are exhaustive and mutually exclusive events), so,

$$P("Blue") = P("Blue"|Blue)P(Blue) + P("Blue"|Green)P(Green)$$

The Cabs problem therefore provides all the information needed to deploy Bayes' theorem to calculate the probability that it was indeed a Blue cab that was involved in the hit-and-run accident:

$$P(Blue|"Blue") = \frac{0.80 \times 0.15}{0.80 \times 0.15 + 0.20 \times 0.85} = 0.41$$

However, most undergraduate student subjects report the probability of the cab being Blue as around 0.80 (this was the median response and the most frequent response). This concurs with the likelihood that the witness correctly identifies the cab, i.e., 0.80. Tversky and Kahneman therefore concluded that their subjects neglect base rate information in calculating posterior probabilities.

2.2 Bias in Causal Attributions

Another area where bias has been commonly observed is causal attribution in social psychology. We illustrate these biases by reference to Kelley's (1967) ANOVA model which represents the attribution process as relying on the principle of co-variation embodied in the analysis of variance. The analysis of variance is of course typically regarded as a normative inductive procedure.

The co-variation principle states that for an event C to be viewed as the cause of event E, E must occur when C does and when C does not occur neither should E. Kelley proposed three dimensions in the attribution of the cause of an event. The cause could be due to persons (P), stimuli (S) or times (T). Three informational variables contribute to "who, what or when" gets the blame, these are:

> **Consensus** between P's response to S and other's response to S at T.
> **Distinctiveness** of P's response to S from P's response to other stimuli at T.
> **Consistency** of P's response to S at T with P's responses to S at other Ts.

Only distinctiveness is directly proportional to covariation. So *high* distinctiveness corresponds to a *high* covariation between S and the response, i.e., high distinctiveness means the response rarely occurs in the absence of S. This contrasts with, for example, consensus information, where *high* consensus corresponds to a *low* covariation between P and the response, i.e., lots of other people also make this response. It is then a trivial corollary of the model that the

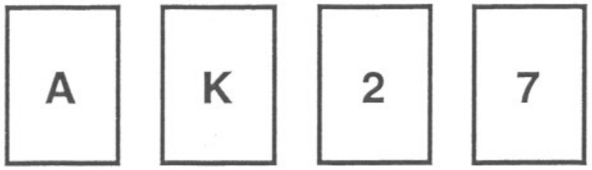

Figure 1: The four cards in the abstract version of Wason's selection task.

following patterns of high or low consensus, distinctiveness or consistency information (in that order) should lead to the indicated attributions: **LLH** = Person, **HHH** = Stimulus, **HLL** = Time. However, while there has been some good general agreement between model and data, some systematic biases have also been observed.

I concentrate on two of these biases. First, it has been found that consensus information is underused particularly in determining person (P) attributions. Recall that low consensus means that there is a high covariation between that person and the response, i.e., the response is rarely observed without this person being involved. Consensus information should therefore be important in determining person attributions. However, McArthur (1972) found that consensus information accounts for only 6% of the variance in person attributions and only 3% of the total variance in causal attributions. Second, there would appear to be a strong bias towards person attributions as opposed to stimulus or time attributions. In McArthur's (1972) results (noted by Jaspars *et al*., 1983), 82% of subjects made person attributions when P covaried with the response, as opposed to 62% who made stimulus attributions when S covaried with the response, and 33% who made time attributions when T covaried with the response.

2.3 Hypothesis Testing

Perhaps the most notorious area where biases have emerged in the reasoning literature concerns human hypothesis testing and in particular work carried out using Wason's (1966, 1968) selection task. Wason's task, which is probably the most replicated task in cognitive psychology, requires subjects to assess whether some evidence is relevant to the truth or falsity of a conditional rule of the form *if p then q*, where by convention "*p*" stands for the antecedent clause of the conditional and "*q*" for the consequent clause. Subjects are presented with four cards each having a number on one side and a letter on the other (see Figure 1). The subjects are also given a rule, e.g., *if there is a vowel on one side* (*p*), *then there is an even number on the other side* (*q*). The four cards show an "A"(*p* card), a "K"(*not-p* card), a "2"(*q* card) and a "7"(*not-q* card). By convention these are labelled the *p* card, the *not-p* card, the *q* card, and the *not-q* card respectively. Subjects have to select those cards they must turn over to determine

whether the rule is true or false. Typical results were: *p* and *q* cards (46%); *p* card only (33%), *p*, *q* and *not-q* cards (7%), *p* and *not-q* cards (4%) (Johnson-Laird & Wason, 1970).

The normative theory of the selection task was derived from standard logic and Popper's (1959) account of falsification. Popper argued that a scientific law can not be shown to be true because it is always possible that the next instance of the law observed will be falsifying. However, one can be logically certain that a law is false by uncovering a single counter-example. Furthermore, this means that scientific reasoning is fundamentally deductive in character - what must be established is a logical contradiction between putative laws and observation. Hence looking for false (*p* and *not-q*) instances should be the goal of scientific inquiry. However, in the selection task subjects typically select cards that could *confirm* the rule, i.e., the *p* and *q* cards.

Confirmation bias, however, does not adequately account for the data on the selection task. Evans and Lynch (1973) showed that when negations are varied between antecedent and consequent of a rule subjects reveal a bias towards selecting those cards that are named in the rules, ignoring the negations. For example, take the rule "if there is not an A on one side, then there is not a 2 on the other side", and the four cards showing an "A", a "K", a "2" and a "7". The matching response would be to select the A and the 2 cards. In contrast according to confirmation the K and the 7 card should be selected and for falsification the K and the 2 card should be selected. Evans and Lynch (1973) refer to this phenomenon as *matching bias*. Matching could also account for the data from the standard affirmative selection task (see above) where no negations are employed in the task rules. This is because the matching and confirmation strategies coincide on the selections they predict for the affirmative rule, i.e., the A and the 2 cards. Thus subjects behaviour on the selection task may not reveal any hypothesis testing strategy at all!

3 Interlude

Having summarised three areas where biases have typically been observed in human reasoning we now turn to how these "biases" are being reconciled with normative theory by the adoption of more appropriate normative models based largely on probability theory. It is interesting to first note that there were always three possible reactions to the observation of biases (Thagard, 1988, p. 123):

> 1. *People are dumb*. They simply fail to follow the normatively appropriate inferential rules.
> 2. *Psychologists are dumb*. They have failed to take into account all the variables affecting human inferences, and once all the factors are taken in to account it should be possible to show that people are in fact following appropriate rules.
> 3. *Logicians [Mathematicians,* my insertion*] are dumb*. They are assessing the inferential behaviour of human thinkers against the wrong set of normative standards.

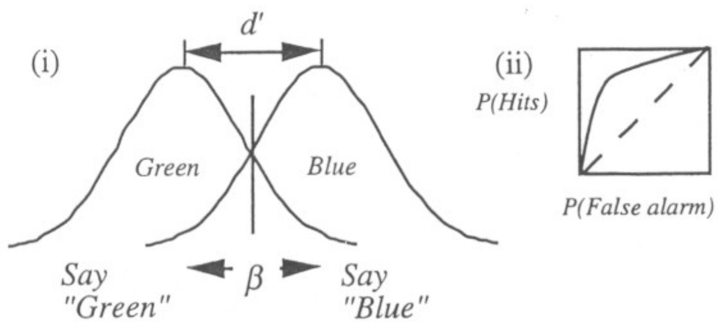

Figure 2: Signal detection theory. (i) Area in "Blue" distribution to the right of β = *P(hit)*; area in the "Green" distribution to the right of β = *P(false alarm)*. (ii) The Receiver Operating Curve (ROC)

Psychologists seem to have largely opted for 1 and explained why in terms of cognitive limitations. Our preference is go after 2 and 3 However, our target in this paper is mainly 3 and we will only briefly touch on work pursuing 2 when we discuss Cheng and Novick's account of biases in causal attribution. (For work pursuing 2, see, e.g., Beyth-Marom, 1982; Gigerenzer, Hell & Blank, 1988; Oaksford & Stenning, 1992.)[1]

4 The Probabilistic Retreat from Biases

4.1 Probabilistic Reasoning and Base Rate Neglect

Birnbaum (1983) has shown that the cabs problem of Tversky and Kahneman (1980) can be recast in the language of Signal Detection Theory (SDT) which is based on Neyman and Pearson's statistics. He shows that when this is done the resulting model concurs with Tversky and Kahneman's (1980) data.

SDT represents the discrimination problem in terms of two partially overlapping normal distributions along which a decision boundary or criterion (β) may be specified (see Figure 2(i)). The discriminability of the two distributions is determined by the separation between their mean (peak) values, this is d' which is measured in z-score units. In the case of the cabs problem one of the distributions corresponds to Green and the other to Blue. The eye witnesses testimony defines a hit rate (80%) and a false alarm r[1] ate (20%) from which d' can be calculated. In this case it equals 1.68. Given one point, a whole ROC curve is defined which corresponds to a plot of pairs of false alarm rates and hit rates for a single d' as the criterion β is varied (see Figure 2(ii)). The

[1] It is worth noting that logicians/mathematicians do not always see their work as directly applicable to these data in the ways that psychologists would wish. Hence often it is

standard psychophysical procedure for plotting an ROC curve is to vary the base rates of the distributions (Green and Blue) thereby moving the criterion to obtain another false alarm rate/hit rate pair. The critical observation in accounting for the cabs problem is that the base rates in the court's test (50%-50%) were different from the night when the witness saw the accident (85%-15%). Hence the witnesses criterion will have been set differently on the night of the accident in comparison to when the test was conducted. However, the lighting conditions were carefully mimicked in the test so it can be assumed that d' remains the same.

What is the effect of altering the base rates? The problem is that SDT does not specify where the criterion is to be placed when the base rates change. However, Birnbaum assumes that witnesses set their criterion to minimise incorrect testimony in the knowledge that there are 15% Blue cabs. This leads to a hit rate of 0.302 and a false alarm rate of 0.698. If these values are then translated back in to likelihoods and used in Bayes' theorem, a value of 0.82 is returned for the posterior probability that the cab was indeed blue given the witnesses testimony. This is remarkably close to the 0.80 actually observed in Tversky and Kahneman (1980).

As Gigerenzer and Murray (1987) point out this analysis depends on *attending* to bases rates not *neglecting* them. They also observe that just by taking Tversky and Kahneman's own proposals seriously with respect to which base rates should be used can lead to the observed data. Subjects may regard the base rates *in the city* as irrelevant seeking instead base rates of cabs *in accidents*. In the absence of relevant information to the contrary, they may assume that the Green and Blue cabs are equally likely to be involved in an accident, i.e., the relevant base rates are 50%-50%. Assuming these base rates, the posterior probability that the cab was indeed blue given the witnesses testimony is 0.80! In sum, it seems that the conclusion of a bias towards base rate neglect may have been premature.

This account of the cabs problem has probably not been influential because (i) of its isolation in a sea of data apparently supporting the existence of biases and (ii) the corresponding lack of promising normative analyses of these data. However, such analyses are now increasing in number as the next two more recent examples demonstrate.

4.2 Bias in Causal Attributions

Cheng and Novick (1990) argue that the problem of causal attribution can be framed in terms of their normative probabilistic contrast model. They show that their model predicts the detailed pattern of results on causal attribution when subjects are provided with complete information.

Cheng and Novick (1990) note that only incomplete information is

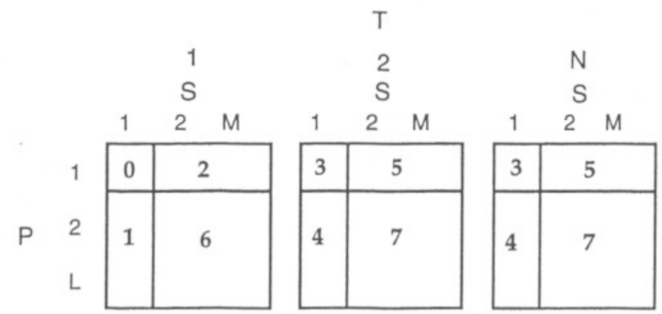

Figure 3. Kelley's ANOVA cube for persons (P), stimuli (S) and times (T), showing the eight regions of information.

these informational sources can be treated as specifying information about specific regions of the cube (see Figure 3).

Figure 3 shows the eight regions of the cube. Consensus information provides information about the agreement between person 1's response to stimulus 1 at time 1, i.e., it provides information about region 1 of the cube (distinctiveness provides information about region 2 and consistency about region 3). Consensus does not provide information about the agreement between person 1's response and others' responses to the other stimuli S_M - S_1 at other times T_N - T_1, i.e., it fails to provide information about regions 4 and 5 of the cube. Cheng and Novick (1990) note that the ANOVA uses information in all regions to determine main effects and interactions. Hence biases in causal attribution may be due to incomplete information. They therefore used materials that provide the missing information. For example, the following materials were used for an **LLH** problem (see above).

> *Jane had fun washing the dishes on this occasion.*
> 1. In fact, Jane always has fun doing all chores.
> 2. But nobody else ever has fun doing any chores.

The first sentence provides information about region 0, i.e., there is a positive relationship between this person (Jane) on this occasion for this stimulus (washing the dishes) and having fun (the response). Clause 2 provides information about regions 2, 3 and 5 (+ relationship) and clause 2 provides information about regions 1, 4, 6 and 7 (- relationship).

Cheng and Novick (1990) derive predictions for the detailed patterns of attributions (including interactions) using their probabilistic contrast model (for details of this approach, see, Cheng & Novick, 1990). They show that their model predicts the detailed pattern of results on causal attribution when complete information is specified (like the ANOVA complete information is required for their model). Moreover, their data reveal that with complete information all signs of bias evaporate. Taking the bias against using consensus information, Cheng

and Novick (1990) found that consensus accounted for 31% of the variance in their data, distinctiveness 33%, and consistency 24%, i.e., no indication of bias was observed. Similarly, for the bias towards person attributions they found that for the problems where only persons (**LLH**), stimuli (**HHH**), or times (**HLL**) covaried with the response the proportion of subjects making the appropriate attribution were 75%, 80% and 85%, i.e., no indication of bias was observed.

Again conclusions of bias have proved to be premature, when complete information is provided subjects show no bias against using consensus information or towards making person attributions. Moreover, their behaviour is predicted by a normative probabilistic model.[2]

4.3 Hypothesis Testing

Oaksford and Chater (1993) show that Wason's selection task can be recast as a problem of optimal data selection that is susceptible to a Bayesian analysis. When this is done they show that all the main results on the selection task can be derived from their normative model.

Any problem of deciding what experiment to perform next, or which observation is worth making, is a problem of optimal data selection. Consider assessing the truth of a mundane regularity, such as that eating tripe makes people feel sick: to collect evidence should we ask people who complain that they feel sick, or perhaps those who do not feel sick, whether they have eaten tripe; should we investigate people who are known tripe eaters or tripe avoiders. This case is analogous to the selection task: the rule *if tripe then feeling sick* is of the form *if p then q*, and the various populations which may be investigated correspond to the various visible card options, *p*, *not-p*, *q* and *not-q*. The subject must judge which of these people should be questioned concerning either how they feel or what they have eaten.

Oaksford and Chater (1993) suggest that those data points expected to provide the greatest information gain should be selected. The information gain given some data is the difference between the uncertainty *before* receiving the data and the uncertainty *after* receiving the data. Uncertainty is measured using the Shannon-Wiener information measure and hence the information gain of a data point D may be defined as follows:

$$I_g(D) = -\sum_{i=1}^{n} P(H_i)\log_2 P(H_i) + \sum_{i=1}^{n} P(H_i|D)\log_2 P(H_i|D) \qquad (3)$$

[2] It could be argued that Cheng and Novick's probabilistic contrast model is redundant if it makes the same predictions as the ANOVA. However, (i) their model also accounts for other areas of causal reasoning where bias had been observed, (ii) the computations involved in their model are far simpler and therefore more psychologically plausible, (iii) a direct indication of facilitatory or inhibitory causation is provided.

That is the difference between the information contained in the *prior* probability of a hypothesis (H_i) and the information contained in the *posterior* probability of that hypothesis given some data D. In calculating the posterior probability Oaksford and Chater assume that the hypothesis under test, which indicates a dependence between p and q, is compared to an alternative hypothesis which indicates that p and q are independent (the principle of indifference between these hypotheses is assumed, so they are both taken to be equally probable, i.e., the *priors* for both are 0.5).

In the selection task, however, when choosing which data point to examine further (that is, which card to turn), the subject does not know what that data point will be (that is, what will be the value of the hidden face). However, what can be calculated is the *expected* information gain. If we assume that there are two possible data points that might be observed (in the case of the *not-q* card, for example, corresponding to the hidden face being, say, p or *not-p*), then the expected information gain of the *not-q* card is:

$$E(I_g(\neg q)) = \sum_{jk} P(H_j)P(p_k|\neg q, H_j)I_g(p_k, \neg q) \qquad (4)$$

Oaksford and Chater then show that arriving at appropriate probability models for the dependence and independence hypotheses only involves specifying the frequency in the population of the properties described by p and q. They label these parameters a and b respectively. With a and b specified all the values needed to compute (4) for each card are available.

In accounting for the selection task Oaksford and Chater (1993) carried out an extensive meta-analysis of the data which revealed that over some 34 studies involving almost 900 subjects the frequency of card selections was ordered such that $p > q >$ *not-q* $>$ *not-p*. To model this data they calculated the expected information gain of each card with $a = b = 0.1$. That is, they assumed that the properties described in p and q are rare in the population. A similar assumption was made by Klayman and Ha (1987) in accounting for related data on Wason's (1960) 2-4-6 task. On this assumption the order in expected information gain is:

$$E(I_g(p)) > E(I_g(q)) > E(I_g(not-q)) > E(I_g(not-p))$$

That is, if subjects base their card selections on the expected information gain of the cards then the observed order in the frequency of card selections is exactly as Oaksford and Chater's (1993) model of optimal data selection predicts. Oaksford and Chater (1993) also show how their model generalises to all the main patterns of results in the selection task, including the manipulation described above of introducing negations into the task rules.

In sum, if the selection task is regarded as a problem of probabilistic optimal data selection then the putative biases observed in this task disappear.

5 Conclusions and Some Possible Implications for Man-Machine Communication

5.1 Summary

We began this chapter by claiming that much of the bias literature needs to be re-interpreted in the light of the growing number of normative probabilistic re-analyses of the data upon which claims of bias have been based. These analyses, as the first example of Birnbaum (1983) reveals, were often contemporaneous with the heyday of the "heuristic and biases" tradition. However, these analyses were largely set aside in the rush to explain these biases in process terms. Of course these accounts simply ignored the fact that they were now assigning apparently irrational functions to the processes they invoked. Recently, however, there has been a re-emergence of concern with rationality and rational analysis (see, e.g., Anderson, 1991; and the collection edited by Manktelow & Over, 1993) and with it we suspect that we are witnessing a degenerative problem shift (Lakatos, 1970) for work in the area of heuristics and biases. Lopes (1991) has observed that the rhetorical emphasis of the heuristics and biases tradition was on Thagard's 1 above, i.e., on human dumbness. She points out that this de-emphasised the importance of rationality in cognitive explanation. In witnessing a problem shift back towards rational function we should be wary that the psychological insights of the heuristics and biases account are not similarly abandoned. As Gigerenzer and Murray (1987, see also, Gigerenzer, Hoffrage & Kleinbölting, 1991) observe there has always been an essential subjective, psychological aspect to probability theory. As we have seen assumptions about which *priors* to employ or their effects on the decision criterion, for example, are not decideable within probability theory. In arriving at psychological theories of reasoning in any domain there will be a need for essentially psychological accounts of how these values are arrived at and how people tractably implement their probabilistic competence.[3] Nevertheless, following Marr (1982), first identifying the rational functions that the human cognitive system needs to compute is again being seen as a feasible and necessary first step in specifying an adequate cognitive theory.

[3] For further discsussion of the computational tractability of reasoning theories, see Chater & Oaksford, 1990, 1993; Oaksford, 1993; Oaksford & Chater, 1991, 1992, 1993; Oaksford, Chater & Stenning, 1990, 1992.

5.2 Man Machine Communication

We now outline an area where these considerations may be relevant to man-machine communication: the apparent success of graphics interfaces like the highly usable Macintosh interface. We make the assumption that all interfaces are deterministic systems. That is, the actual rules that determine what happens given user input are deterministic, i.e., as long as the system is functioning normally, a given input, key press sequence, mouse click in a menu location etc., yields a single specific action with probability 1. Now at some metaphysical level of description the world may be a deterministic system. Certainly classical mechanics which for the bulk of our everyday macroscopic needs is all we need to predict the physical world, regards the world deterministically. However, this does not mean that in acquiring knowledge of the macroscopic world human beings model it deterministically. Irremediable ignorance of the disposition of a complex system whose next state we wish to predict often precludes deterministic prediction. Instead scientists must settle for probabilistic theories in order to make adequate predictions. Similarly, when confronted with a complex user interface novice users may model the system not deterministically but probabilistically.

This provides a possible explanation for the success of interfaces like the Macintosh. Deterministic command line interfaces are all or nothing affairs, either you know the appropriate command or you don't. The organisation of access is also very flat in that if you don't know the specific command there are no higher level exploratory actions you can perform in order to find out. This kind of organisation is only suited to rote learning of the appropriate command language and rules. This contrasts with the Macintosh interface, where there are often many ways of performing the same action. Moreover, there are a set of high level actions, like pulling down menus, clicking on *labelled* buttons, scrolling etc. that enable the user to interact with the computer in an exploratory manner. This is analogous to our normal inductive interactions with the world. In acquiring practical skills the ability to manipulate and explore bits of our world permits us to learn much faster. It seems possible, therefore, that these interfaces succeed because they exploit the same abilities that allow us to learn so effectively about the world in order to inductively explore the interface. In sum, novices may model their interactions probabilistically in accordance with their natural competences; the very competences that are now being revealed in the probabilistic retreat from bias.

References

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, **14**, 471-517.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory & Cognition*, **10**, 511-519.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, **96**, 85-94.

Chater, N. & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. *Cognition*, **34**, 93-107.

Chater, N. & Oaksford, M. (1993). Logicism, mental models and everyday reasoning: Reply to Garnham. *Mind & Language*. **8**, 72-89.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, **58**, 545-567.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391-397.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human Reasoning*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 513-525.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilisitic mental models: A Brunswickian theory of confidence. *Psychological Review*, **98**, 506-528.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, N.J.: Lawrence Eerlbaum Associates.

Jaspars, J. M. F., Hewstone, M. R. C., & Fincham, F. D. (1983). Attribution theory and research: The state of the art. In J. M. F. Jaspars, F. D. Fincham, & M. R. C. Hewstone (Eds.), *Attribution theory: Essays and experiments*, (pp. 3-36). Orlando, FL: Academic Press.

Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, **1**, 134-148.

Kahneman, D., & Tversky, A. (1972). On prediction and judgement. *ORI Research Monograph*, **12**.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*, Cambridge: Cambridge University Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In Levine D. (Ed.), *Nebraska Symposium on Motivation* Vol. 15, (pp. 192-238). Lincoln: University of Nebraska Press.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, **94**, 211-228.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos, & A. Musgrave (eds.), *Criticism and the growth of knowledge*, (pp. 91-196). Cambridge: Cambridge University Press.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology*, **1**, 65-82.

Manktelow, K. I., & Over D. E. (Eds.) (1993). *Rationality*, London: Routledge.

Marr, D. (1982). *Vision*. San Fransisco, CA: W. H. Freeman & Co.

McArthur, L. Z. (1972). The how and what of why: Some determinants and consequecnces of caual attribution. *Journal of Personality and Social Psychology*, **22**, 171-193.

Norman, D. A., & Draper, S. W. (1986). *User-centred System Design*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Oaksford, M. (1993). Mental models and the tractability of everyday reasoning. *Behavioural & Brain Sciences*. **16**, 360-361.

Oaksford, M. & Chater, N. (1991). Against logicist cognitive science. *Mind & Language*. **6**, 1-38.

Oaksford, M. & Chater, N. (1992). Bounded rationality in taking risks and drawing inferences. *Theory & Psychology*. **2**, 225-230.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. I. Manktelow & D. E. Over (Eds.) *Rationality*. pp. 31-60, London: Routledge.

Oaksford M, & Chater N. (1993). A rational analysis of the selection task as optimal data selection. Technical Report No. CCN-UWB-93-05: Centre for Cognitive Neuroscience, University of Wales, Bangor. (In submission).

Oaksford, M., Chater, N. & Stenning, K. (1990). Connectionism, classical cognitive science and experimental psychology. *AI & Society*, **4**, 73-90.

Oaksford, M., Chater, N., & Stenning, K. (1992). Connectionism, classical cognitive science and experimental psychology. In Clark, A. & Lutz, R. (Eds.) *Connectionism in Context*. pp. 57-74, Berlin: Springer-Verlag.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **18**, 835-854.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgements under uncertainty. In Fishbein M (Ed.), *Progress in social psychology*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

Wason, P. C. (1960). On the failure to elminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, **12**, 129-140.

Wason, P. C. (1966). Reasoning. In B. Foss (ed.), *New horizons in psychology*, Harmondsworth, Middlesex: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273-281.