Acta Psychologica 43 (1979) 111–121

PSEUDODIAGNOSTICITY

Michael E. DOHERTY, Clifford R. MYNATT, Ryan D TWENEY and Michael D. SCHIAVO
Dept. of Psychology, Bowling Green State University, Bowling Green, Ohio

Revised version received April 1978

Subjects selected data in order to decide from which of two 'islands' an 'archeological find' had come. The results replicated two established phenomena in cognitive psychology: (1) the tendency to ignore base rate data given individuating information, and (2) the tendency to seek confirmatory evidence.

The major outcome of the study was, however, to reveal a new phenomenon in information search. Subjects displayed a surprising and strong tendency to seek diagnostically worthless information. They then altered their conclusion based on that information. For example, subjects who had already obtained $P(D_1/H_1)$ selected $P(D_2/H_1)$ when $P(D_1/H_2)$ was equally easily available, and when they had no relevant experience to bring to bear on the estimation of $P(D_1/H_2)$. This phenomenon, which appears to be a wholly dysfunctional cognitive tendency, was labeled *pseudodiagnosticity*.

Cognition involves a complex interplay between mechanisms of information input, information processing and information search. While this has long been clear, too little is known about the interaction of these mechanisms. In particular, behavioral scientists have discovered much about the limits on cognitive processes imposed by, for example, the limited capacity to store information in short term memory. But little is known about mechanisms that people use to overcome these limitations. The present study was designed to explore limitations of another sort, limitations which operate when subjects seek information needed to make a judgment under uncertainty. The task was such that three basic cognitive limitations were manifested: (1) failure to use statistical base-rate information (Tversky and Kahneman 1974); (2) confirmation bias (Wason 1960; Mynatt et al. 1977); and (3) failure to identify and select diagnostically relevant information. The latter phenomenon, which we have called pseudodiagnosticity, has not previously been described.

While the present paper is not a test of the adequacy of Bayes' theorem as a descriptive model of how subjects revise their opinions, it will be

convenient to define the three limitations of interest in terms of Bayes' theorem. Consider the equation

$$P(H_1/D_i) = \frac{P(H_1)\,P(D_i/H_1)}{P(H_1)\,P(D_i/H_1) + P(H_2)\,P(D_i/H_2)} \qquad (1)$$

in which H and D stand for hypothesis and data, respectively, the subscripts 1 and 2 label two mutually exclusive and exhaustive hypotheses, and the subscript i indexes a set of data. Assume further that the probability (P) values are defined as relative frequencies. The base rate for H_1 is simply the prior probability that H_1 is true, and is denoted by $P(H_1)$. Note that very high or low values of $P(H_1)$ can overwhelm the diagnostic effects of data, even when those data are highly diagnostic. Whether behavior is influenced appropriately (i.e., in accordance with the normative model) by $P(H_1)$ is one way of determining how well people use base rate data. Confirmation bias, the second limitation, may reflect itself in several ways, one being when a subject favoring H_1 does not search for data which would be likely to favor H_2, or does not seek any information about H_2 at all. Finally, Bayes' theorem points to a strong qualitative test of whether subjects possess a working knowledge of the concept of diagnosticity. Suppose a subject in a two hypothesis task is given the opportunity to choose information revealing $P(D_1/H_1)$ and $P(D_1/H_2)$ or $P(D_2/H_1)$ and $P(D_2/H_2)$. If the subject instead chooses information revealing $P(D_1/H_1)$ and $P(D_2/H_1)$ in order to make an inference, then such a subject cannot be said to have a working understanding of diagnosticity. That is, the posterior probability of either hypothesis, say $P(H_1/D_i)$, can be calculated only if *both* the probability of a piece of data given the hypothesis under test, $P(D_1/H_1)$, *and* the probability of the same piece of data given the alternative hypothesis, $P(D_1/H_2)$, are known.

A compelling demonstration of people's failure to use base rate data appropriately is provided by Tversky and Kahnemann (1974). Subjects asked to classify university students ignored radical base rate differentials as soon as they were given very weak individuating data. That is, though values of $P(D_1/H_1)$ and $P(D_1/H_2)$ were much too close to each other to overturn the base rates, subjects were quick to abandon the hypothesis favored by the base rates. Since failure to consider base rate information is of both theoretical and practical consequence, the phenomenon merits further attention, including attempts at replication in other situations.

Confirmation bias can be inferred from two different forms of behavior: (a) failure to change one's opinion in the face of nonsupporting or even contradictory evidence and (b) selection of data favoring one's hypothesis while ignoring sources of data which might be likely to contradict one's hypothesis. Confirmation bias seems to be a pervasive phenomenon. Pitz's "inertia effect" (Pitz et al. 1967; Geller and Pitz 1968) is an instance in a bookbag and poker chip task, as is 'freezing' of hypotheses in the clinical judgment setting (Wiggins 1973; see also Wallsten 1976). A tendency to look only for confirmatory evidence was found in Wason's four card selection task and 2-4-6 studies (Wason 1960, 1968). Mynatt et al. (1978) observed the effect in simulations of scientific research, and Mitroff (1974) noted that top-grade scientists not only admit to a confirmation bias but assert that it plays a necessary role in scientific progress. Confirmation bias thus seems to be a phenomenon meriting further exploration, especially given the logical asymmetry (Popper 1959) between confirmation and disconfirmation in inductive inference.

Much research on human inference has compared opinion revision with the degree of such revision prescribed by Bayes' theorem (Pitz 1975). One recent experiment is of special interest. Troutman and Shanteau (1977) explored a phenomenon called the "nondiagnosticity effect". They provided subjects with samples from one of two boxes of beads. One box had 70 red, 30 white and 50 blue beads. The other box had 30 red, 70 white, and 50 blue beads. Thus a blue bead was noninformative. Yet subjects significantly revised their subjective probabilities when they were given blue beads, in the direction of less extreme judgments. Wallsten's (1976) subjects were also influenced by neutral information. There is other evidence that subjects are not properly sensitive to diagnosticity (e.g., Steinmann and Doherty 1972). The fundamental issue we wish to address is not whether subjects misuse diagnosticity in some quantitative way, but whether diagnosticity is a cross-situational, behaviorally significant task variable.

In the experiments reported below, the subjects were asked to determine from which of two islands (corresponding to H_1 and H_2) an artifact had come. The artifact was described by a number of binary characteristics (D_1 and D_2 for each characteristic) and subjects sought information about the values of $P(D/H)$ of these characteristics from the two islands.

Base rates were manipulated to determine whether the phenomena described above would replicate in a concrete task with clearly defined statistical properties. The presence of confirmation bias in the task can be indirectly inferred by observing whether subjects sought information about a favored hypothesis rather than an alternative. Finally, and most importantly, the question of whether subjects comprehend diagnosticity can be addressed directly, by observing whether subjects choose data relevant to *both* hypotheses when given the opportunity to do so.

The experiment

Method

Subjects

The experiment was run in a large Introductory Psychology class. There were 152 participants.

The task

All instructions and manipulations were accomplished by varying the content of a 5-page booklet. The *S*s' responses were made directly in the booklet. The first paragraph on the first page provided the context for the *S*s:

Imagine that you are an undersea explorer who has just found a pot on one of your dives. The pot looks very valuable and you would like to return it to its homeland. Although there are many islands in the surrounding area from which the pot could have come, you know that pottery-making exists on only two of these islands, Coral Island and Shell Island. You would like to determine from which of these two islands the pot came. One method you could use to try to determine this would be to compare the characteristics of your pot with what is known about pots and pottery on each of the two islands. Luckily, on each island there is a museum that contains information about the pottery made on that island. From your ship, you can place phone calls to the museums on each of the two islands. The museum supervisors are very busy, however, so they will give you only one piece of information about their collections during each phone call.

You first make a careful examination of the pot you found, and list its characteristics as follows:

Smooth clay (not rough)
Curved handles (not straight)
(Six more binary characteristics were then listed.)

The next paragraph then indicated that a phone call to each island had produced information about the number of pots that had been lost at sea by each. Differences in these numbers provided the base rate manipulation.

Page 2 of the booklet instructed *S*s that they could now make two more 'phone calls', which were simulated by peeling off any two opaque stickers. One possible layout would be:

	Coral Island	Shell Island
Curved handles	21%	87%
Straight handles	79%	13%
Smooth clay	19%	91%
Rough clay	81%	9%

Each of the four pairs of percentages (e.g., 21% and 79%) was covered by a circular sticker.

The *S*s were then asked "which island do you think the pot came from?" They checked either Coral or Shell, and then wrote down the reasons for their choice.

On page 3 the *S*s were faced with an array of 12 stickers covering the proportions on each island of the other six binary characteristics. They were permitted to peel off any six of the 12. They were asked to number each choice and state a decision after each. Page 4 asked for a final choice, and asked two open ended questions about strategies. Page 5 provided a check on whether *S*s remembered the base rate, and asked more open ended questions about their strategies.

There were two levels of base rate. An equal base rate treatment was provided by instructing the *S* that 1000 pots had been lost at sea from each island. An unequal base rate treatment was provided by having 5000 pots lost by one island, and only 500 by the other. The name of the island with the high loss was Coral for some *S*s, Shell for others. All unequal base rate *S*s had data on page 2 which favored the opposite hypothesis than that favored by the base rate. However, since only two stickers could be peeled, *S*s could observe at most only one pair. In the example given above the pot was smooth and had curved handles. The data favored Shell Island whether the *S*s pulled the pair of stickers for texture or for handles. But the base rate was sufficiently extreme to outweigh either data set by a large margin. Suppose that an *S* who favored Coral Island after seeing the base rate peeled the stickers describing the distributions of handle characteristics. If so, equation 1 tells us that the probability that the pot came from Coral Island is

$$P\,(\text{Coral/curved handle}) = \frac{(0.909)\,(0.21)}{(0.909)\,(0.21) + (0.091)\,(0.87)} = 0.707.$$

The comparable value given information about the smooth-rough dichotomy is 0.676. Note also that an *S* might select stickers so as to yield, from a formal point of view, no information. That is, they could select one sticker from each pot characteristic, or both from one island. Such choice behavior would not give the *S* the values needed for the formally appropriate computations.

On page 3, the data associated with all six binary characteristics agreed in their implications for the hypotheses favored by the data on page 2. The ratios were of the same magnitudes as those shown above. For a control group (N = 20), the favored hypothesis was reversed.

There were six equal base rate conditions, which controlled for first order of mention of the island, the direction of the data on page 2 and the direction of the

data on page 3. Since, as expected, no interesting differences involving these groups emerged, subsequent analyses are collapsed across them, and they will not be further described. A total of 66 *S*s had unequal base rates, while 86 had equal base rates.

Procedure

The students were previously informed that an inclass experiment was to be performed that day. Verbal instructions were brief, and consisted only of a request for silence and a comment that since there were many different forms of the booklet, the students should not look at their neighbors' booklets. The students were given 35 minutes. Any who completed the task early had the booklet collected and took out a book or notes to study. The remainder of the class period was given over to a discussion of the optimal responses, and why they were optimal.

Results

The data will be discussed with respect to each of the three issues raised in the introduction.

Base rates

Of the 66 *S*s who had unequal base rates, 60 stated that they thought the pot had come from the island favored by the data on page 2 of the booklet. Thus, base rate was ignored by these *S*s in the face of individuating information. Of the 86 equal base rate *S*s, 75 chose the pot favored by the data on page 2, one did not respond and one wrote out an explanation which indicated a reversal in the reading of the percentages. Nine *S*s gave responses counter to the evidence, the reasons for which were uninterpretable.

Confirmation bias

Each *S* was categorized according to the number of pairs chosen on page 3 of the booklet and by the number of stickers peeled for each island. There were a maximum of three pairs per *S*, and six stickers for either island. The number of pairs chosen constrained the number of stickers for an island (for example an *S* who chose 3 pairs had to have the choices distributed equally). An *S* who chose no pairs could have any number of choices from 0 to 6 for a single island. Data were collapsed over island names. The instruction to record the choice sequence on page 3 was not sufficiently emphasized, and many *S*s did not follow it, so only *S*s who had the same final hypothesis on page 3 as they had on page 2 were counted. 11 *S*s were dropped for switching hypotheses somewhere on page 3. 20 others were excluded since they were in the control condition which led them to switch on page 3. Thus, 121 *S*s met the criterion of having the same initial hypothesis on page 2 as their final hypothesis. Their choice responses (i.e., stickers peeled) on page 3 can then be interpreted as being in the context of a favored hypothesis. Table 1 presents these responses. These are the crucial data for this and the next section, and will be illustrated by example. Suppose an *S* hypothesized Coral Island on page 2, selected 6 stickers revealing data about pots from Coral Island, and then concluded that the

Table 1
Number of subjects removing differing numbers of labels on page 3, categorized by number of pairs selected and number of labels selected relevant to favored island. A dash indicates that no entry in that cell was possible.

Number of pairs	Number of labels relevant to favored island							
	0	1	2	3	4	5	6	Σ
0	1	0	5	32	16	10	7	71
1	–	0	2	14	9	5	–	30
2	–	–	1	4	4	–	–	9
3	–	–	–	11	–	–	–	11
Σ	1	0	8	61	29	15	7	121

pot was from Coral Island. That *S* would be one of the 7 in the upper right cell in table 1. So would an *S* who hypothesized Shell, picked 6 Shell stickers, and concluded Shell. The middle data column represents *S*s who chose stickers equally from both islands. The right three columns indicate choices of data from the favored island [i.e., data to assess $P(D/H)$], while the left three columns are choices of data from the non-favored island [$P(D/{\sim}H)$].

61 *S*s chose equally from the two islands. For the remaining 60 *S*s, the data strongly suggest the operation of a bias to confirm. Of those 60, 51 chose more often from the favored island. A test of the significance of the obtained proportion for the asymmetric choices against a null hypothesis that $P = 0.5$ was highly significant ($z = 5.42, p < 0.01$).

Pseudodiagnosticity

The question of whether the *S*s behave as though they have any working understanding of the concept of diagnosticity is also addressed by table 1. Very few of the *S*s shown chose 3 pairs – the only rational strategy. The majority chose *no* pairs. Of the total number of 152 *S*s, only 19 chose 3 pairs. Again the raw data are impressive. *Everyone* should have picked three pairs, had they even a rudimentary grasp of how to make sense out of the data that were available. A χ^2 test for the data shown in table 1, testing the row totals against a null hypothesis that *S*s were selecting stickers by chance, is highly significant ($X^2 = 538.5, df = 3, p < 0.01$).

The data from page 2, especially from the unequal base rate *S*s, are even more compelling. Recall that the data for either pair of characteristics were not sufficiently diagnostic to lead an optimal Bayesian to depart from the base rate. The *S*s' responses were tabulated by the number of pairs they chose (zero or one) and by the hypothesis they stated after removing the two stickers. Table 2 reports the results of this categorization for the unequal base rate *S*s only, collapsed across island names. Note that if all our *S*s had behaved optimally with respect to the selection and interpretation of evidence, all of them would have been in the lower left cell.

Table 2
Number of subjects[a] with given hypotheses on page 2, categorized by number of pairs chosen on page 2.

	Hypothesis	
Number of pairs chosen	*H* favored by Bayes' rule	Incorrect *H*
0	1	49
1	3	11

[a] Two subjects were unclassified since they peeled only one sticker on page 2.

Three of the 64 *S*s were. If the limitation were only on the *S*s' ability to process the evidence presented, as is often assumed in experiments employing Bayes' theorem as a model, then all *S*s should have shown up in the bottom row, with the distribution over hypotheses determined by other factors. But only 14 of the 64 *S*s selected a pair of labels. Of the *S*s in the equal base rate conditions, only 20% selected a pair. It is clear that the diagnostic value of pairs of characteristics was not understood by *S*s. Not only did they not select pairs, but they changed their hypotheses based on the pseudodiagnostic data they did select.

Discussion

The hypothesis that people ignore base rates in the face of individuating information received very strong support from the present study. The tendency to ignore base rates is truly a robust phenomenon. Subjects abandoned the hypothesis strongly favored by the base rate, even though the data they selected were either insufficient to warrant such a change or, more often, worthless.

The hypothesis that people's inference behavior is limited by a strong bias to confirm is also supported by the data. That phenomenon, too, is robust. Choice responses on page 3 were clearly consistent with a confirmation bias hypothesis. There seems to be no reasonable alternative explanation for such an imbalance in the choice data.

The novel contribution of this study lies in the demonstration that people do not have a working knowledge of the concept of diagnosticity. When the subjects sought statistical data with which to evaluate the artifacts, the great majority did not ask for the data needed to form likelihood ratios. It is not a question of nonoptimal revision, given data.

The subjects actively chose irrelevant information and ignored relevant information which was equally easily available. Furthermore, table 2 shows that the subjects who chose *no* pairs were almost uniformly (98%) swayed by these data. The term 'pseudodiagnosticity' was chosen to highlight the fact that the data had a major impact on the decisions. In spite of the 10:1 prior probability (or base rate), the subjects revised their opinion, given worthless data.

Two other experiments dealing with the phenomenon were run, with a total of 66 subjects. These experiments did not permit the choice of pairs *vs* non-pairs, but called for a decision after only a single datum. The subjects revised their opinions based on that single, hence formally inadequate, $P(D/H)$. These two experiments serve primarily to achieve a modest cross-task generalization, since the operation used to demonstrate pseudodiagnosticity differed from the main study. The data of the two smaller studies were also consistent with the base rate and confirmation bias phenomena discussed above.

It might be instructive to speculate about how the subjects may have been attempting to solve the task. A theorist trying to rationalize the subjects' behavior would have to assume that the subjects were making explicit guesses about the distributional characteristics of the task. That is, one might posit that the subject who uncovered a high value of $P(D/H_1)$ might then guess that $P(D/H_2)$ is likely lower. Such a strategy would be rational in some circumstances, specifically those in which the subject has relevant experience to bring to bear on the estimates of unknown probabilities. However the present task was novel to the subjects, and there was no way in which they could reasonably have guessed the relative magnitude of $P(D/H_2)$, given $P(D/H_1)$. This is especially so since the pseudodiagnosticity effect appeared on the first opportunity for it to do so, before any observations at all were made of $P(D/H_2)$. Even if some model which posits that subjects are making assumptions about hidden frequencies does describe the underlying process, the fundamental point of the research is nonetheless valid: the relevant data were there for the asking – and the subjects did not ask for it. While we do not yet have sufficient data to propose a formal process model, we believe that the basic phenomenon operating in this context, underlying both pseudodiagnosticity and confirmation bias, is a cognitive one: it is hard to think about one datum in relation to two hypotheses. It is much easier to think about the relevance of two data to one hypothesis.

It is worthwhile to contrast the picture of inference behavior which emerges from these experiments with that which emerges from traditional bookbag and poker chip studies. The bookbag tasks typically constrain subjects sharply with respect to the response which can be made, permitting subjects only to revise their subjective probabilities given the prior probabilities, data sequences, and diagnosticities which have been imposed by the experimenters. The picture which emerges is either that the subject is or is not a conservative Bayesian (see Pitz 1975, for a recent discussion). From the present task a more articulated picture of the inference process is possible. The subject, given a chance to ask for data, seems more likely to ask for data perceived as relevant to the favored hypothesis than to seek information about alternatives. Given any individuating information, base rates are readily abandoned. Once the datum is 'in', the subject makes an immediate adjustment in the hypothesis state, even in the absence of the other data which form the necessary context in which to evaluate the obtained datum. This process, which we have labeled pseudodiagnosticity, can only abet premature closure with respect to the hypothesis, preclude consideration of the impact of the datum on other hypotheses, and inhibit the search for other evidence which may support alternative hypotheses.

There is an interesting parallel between our results and those observed in Wason's (1968) four card selection task. His subjects were required to seek hidden information to verify a general rule of the form 'if P then Q'. P referred to one side of a card, Q to the other side, and subjects were asked to choose between P, Q, not P, and not Q cards. Most chose only P and Q. In fact, however, only P and not Q are informative, since these are the only choices that can falsify the rule. All other combinations are logically consistent with the rule, and so are uninformative. Wason's results have generally been explained as the results of reliance on intuitive matching processes, rather than on logical relations (Evans and Lynch 1973; Wason and Evans 1975). As in Wason's task, pseudodiagnosticity represents failure to seek the only potentially 'falsifying' information available to our subjects. The choices actually made by subjects are consistent with the pot being from either island, and so are uninformative. This comparison may point toward understanding of the mechanisms which underlie pseudodiagnosticity. The phenomenon may be another example of what Wason and Evans (1975) have called "an 'intuitive' way of coping with conceptual difficulty". Our results may extend such processes to the domain of probabilistic reasoning.

Pseudodiagnosticity is clearly dysfunctional in the task presented our subjects. If it operates in more open problem spaces, as for example in scientific thinking, its consequences may be especially severe. Whereas practicing scientists often see confirmation bias as potentially valuable for its motivating effect and for its role in establishing 'good hypotheses', it is hard to see what positive benefits could result from processing data in a pseudodiagnostic fashion. Such data cannot provide disconfirmatory evidence for a hypothesis. More seriously, they cannot even serve to confirm a hypothesis, and are, in the fullest sense of the word, worthless.

References

Evans, J. St. B. T. and J. S. Lynch, 1973. Matching bias in the selection task. British Journal of Psychology 64, 391–397.

Geller, E. S. and G. F. Pitz, 1968. Confidence and decision speed in the revision of opinion. Organizational Behavior and Human Performance 3, 190–201.

Mitroff, I., 1974. The subjective side of science. Amsterdam: Elsevier.

Mynatt, C. R., M. E. Doherty and R. D. Tweney, 1977. Confirmation bias in a simulated research environment: an experimental study of scientific inference. Quarterly Journal of Experimental Psychology 29, 85–95.

Mynatt, C. R., M. E. Doherty and R. D. Tweney, 1978. Consequences of confirmation and disconfirmation in a simulated research environment. Quarterly Journal of Experimental Psychology 30, 395–406.

Pitz, G., 1975. Bayes' theorem: can a theory of judgment and inference do without it? In: F. Restle, R. M. Shiffrin, N. J. Castellan, J. R. Lindman and D. B. Pisoni (eds.), Cognitive theory, Vol. 1. Hillsdale, N. J.: Erlbaum.

Pitz, G. F., L. Downing and H. Reinhold, 1967. Sequential effects in the revision of subjective probabilities. Canadian Journal of Psychology 21, 381–393.

Popper, K. R., 1935. The logic of scientific discovery (original edition, Logik der Forschung, Vienna: Springer). London: Hutchinson, 1964.

Steinmann, D. O. and M. E. Doherty, 1972. A lens model analysis of a bookbag and poker chip experiment. A methodological note. Organizational Behavior and Human Performance 8, 450–455.

Troutman, C. M. and J. Shanteau, 1977. Inferences based on nondiagnostic information. Organizational Behavior and Human Performance 19, 43–55.

Tversky, A. and D. Kahnemann, 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1131.

Wallsten, T. S., 1976. Complex sequential judgments given neutral and disconfirming information. Paper presented at the Annual Meeting of the Psychonomic Society. St. Louis, Missouri, November, 1976.

Wason, P. C., 1960. On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology 12, 129–140.

Wason, P. C., 1968. Reasoning about a rule. Quarterly Journal of Experimental Psychology 23, 273–281.

Wason, P. C. and J. St. B. T. Evans, 1975. Dual processes in reasoning? Cognition 3, 141–154.

Wiggins, J. S., 1973. Personality and prediction: principles of personality assessment. Reading, Mass.: Addison-Wesley.