

The influence of feedback and diagnostic data on pseudodiagnosticity

MICHAEL E. DOHERTY, MICHAEL B. SCHIAVO, RYAN D. TWENEY,
and CLIFFORD R. MYNATT

Bowling Green State University, Bowling Green, Ohio 43403

Pseudodiagnosticity is the tendency of people to select diagnostically worthless data in an opinion revision task and then to make a judgment based on those data. University students were presented four problems in which they selected data and then made a judgment based on those data. Knowledge of results was provided on each trial. Half the subjects were "force-fed" truly diagnostic data. Getting the correct answer virtually guaranteed that subjects repeated the data selection strategy that preceded it. Data selection following an incorrect answer was contingent upon whether the subjects had seen diagnostically relevant data.

The heuristics people use in trying to understand the world provide one of the major focuses of research in cognitive psychology (Einhorn & Hogarth, 1981; Hogarth, 1980; Nisbett & Ross, 1980; Slovic, Fischhoff, & Lichtenstein, 1977; Tversky & Kahneman, 1974; Wimsatt, 1980). While most of the heuristics and biases are functional in some contexts, they are dysfunctional and perhaps irrational in others. Whether such instances can be considered truly irrational is questioned by Einhorn and Hogarth (1981) and vigorously denied by Cohen (in press). In a recent paper, however, we argued that a particular bias, pseudodiagnosticity, was wholly dysfunctional and irrational (Doherty, Mynatt, Tweney, & Schiavo, 1979).

Pseudodiagnosticity refers to a strong tendency on the part of subjects to select and use diagnostically worthless information in preference to equally available diagnostic data. Subjects in the Doherty et al. (1979) study were given two data (D_1 and D_2) and two hypotheses (H_1 and H_2) to test, and then they selected any two of $p(D_1/H_1)$, $p(D_2/H_1)$, $p(D_1/H_2)$, and $p(D_2/H_2)$. The only appropriate strategy in such a task is to select either $p(D_1/H_1)$ and $p(D_1/H_2)$ or $p(D_2/H_1)$ and $p(D_2/H_2)$, that is, the likelihoods of one datum given both hypotheses. The overwhelming majority of subjects chose diagnostically worthless data, and then proceeded to draw unwarranted conclusions. These findings have been replicated with advanced medical students (Kern & Doherty, 1981) and in unpublished research with graduate students and faculty.

In none of these studies have subjects been provided feedback about the correctness or incorrectness of their decisions. Subjects have not been systematically exposed to truly diagnostic information either, since their own behavior precluded most of them from ever seeing appropriate probabilities. Pseudodiagnostic selection strategies may be attenuated if subjects see that they frequently lead to erroneous decisions and/or see diagnostic data. This paper tests these conjectures by provid-

ing subjects both feedback and "force-fed" diagnostic data.

METHOD

Subjects

The experiment was run in an introductory psychology class. There were 176 subjects, of whom 167 provided complete data.

Materials and Procedure

All instructions and manipulations were accomplished by varying the content of an eight-page booklet, patterned after the one used in Doherty et al. (1979). Since the first page describes the task relatively completely, it will be reproduced verbatim.

"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 pottery made on that island. From your ship, you can place phone calls to the museums on each of the two islands and find out what percentage of pots have the characteristics of the pot you found. 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 examine your pot and note that it has a rather wide mouth and a curved handle. On your first call to Coral Island, the supervisor tells you that pottery-making has always been popular on his island. All pots shipped to other islands are first registered at the museum. All told, about 1,000 pots have been registered. On your first call to Shell Island, you learn that pottery-making is popular on this island also. And, as is the policy on Coral Island, all pots leaving the island must be registered with the museum. All told, about 1,000 pots have been registered at the museum on Shell Island.

"Now we would like you to determine which island the pot came from by making more phone calls. Remember, since the supervisors are busy, you can ask for information about only one pot characteristic on each call. To simulate this, we have placed the information you could receive from one phone call underneath a peel-off label. To make a phone call, decide

Table 1
Percentages Under the Stickers on Page 2

	Island	
	Coral	Shell
Wide Mouth	61	66
Narrow Mouth	39	34
Curved Handles	52	58
Straight Handles	48	42

which pot characteristic you would like to have information about. Then decide which island you want to call (since you can't talk to both supervisors at the same time). Then simply peel off that label corresponding to the characteristic and island you chose, and read the information."

Note that the instructions established equal prior probabilities. On Page 2, subjects were given two binary pot characteristics and four pairs of percentages covered by opaque stickers. They then made two "calls" (peeled two stickers), the sequence being recorded by subjects' numbering each call. They were then asked to decide whether the pot had come from either Coral or Shell Island.

The percentages under the stickers on Page 2 are shown in Table 1. Normatively, subjects should select either row, each of which favors the hypothesis that the pot had come from Shell Island. Any subject who peeled both stickers from the first column was expected to conclude pseudodiagnostically that the pot had come from Coral Island, and those who chose both in Column 2, Shell. Either diagonal provided mixed "evidence." All subjects received feedback, the name of the island with the higher value of p(H/D), after making their choice. In the above example, the word "Shell" would be under a peel-off label.

After receiving feedback, subjects were told either to think about their choices (control group) or to peel a third sticker (experimental, of "third-call," group) from the array. Note that all subjects in the third-call group saw exactly one properly diagnostic pair of data per trial. Thus, third-call subjects saw that the datum apparently diagnostic of Coral Island (e.g., the wide mouth) actually favored Shell, since a still higher percentage of Shell's pots had a wide mouth. Pages 3, 4, and 5 were identical in form to Page 2, each with a new pot differing in features and percentages. The final three pages asked the subjects to evaluate the strategies of three other divers. Subjects were shown problems in which two "calls" had been made. One hypothetical diver had requested data on a diagonal, the second a column, the third a row. Subjects were asked to decide from which island the pot had come, but they were given for the first time a third choice, "can't tell." They were also asked on each of the three pages whether the diver had asked for the right data and to explain why or why not.

RESULTS

Table 2 shows the percentages of subjects choosing diagnostic data across trials. There was a weak tendency for control subjects to adopt a normative strategy, compared with a dramatic increase in normative behavior for the experimental subjects. Considerable understanding of the reasons for the dramatic shift can be gained by examining choice behavior as a function of experience on the previous trial. Table 3 presents these data, broken down by third-call vs. control group, by whether the choice on the previous trial had been diagnostic (row) or nondiagnostic (column or diagonal), and whether or not the subjects' conclusions had been correct.

Knowledge of Results

The data reveal a powerful effect of knowledge of results on data selection. Note that by the fourth trial, 65 of 66 subjects in the control condition repeated the data selection strategy that had led them to be told their prior conclusion had been correct (24 of the 25 subjects who chose diagnostically on Trial 3 and got positive feedback chose diagnostically on Trial 4, and 41 of 41 subjects who chose nondiagnostically on Trial 3 and got positive feedback chose nondiagnostically on Trial 4). The effect of being told that one's conclusion is wrong was less predictable. Of the 44 control subjects who chose nonoptimally on Trial 1 and were wrong, 35 chose nonoptimally again on Trial 2. Examination of individual subjects' data shows that of those 35, 25 were correct on Trial 2, all of whom chose nonoptimally again on Trial 3 and Trial 4. Of the 10 nonoptimal choosers (from the 44) on Trial 2 who were wrong, 4 chose optimally in Trial 3, and 6 stayed with the wrong strategy. These "reinforcement effects," with 98% of controls on Trial 4 sticking with the reinforced strategy and the "punished" subjects behaving relatively unpredictably, indicate that subjects were attending to the task and provide baseline data for the third-call group.

Table 2
Percentages of Subjects Choosing the Diagnostic Data Necessary to Form Likelihood Ratios

Group	N	Trial			
		1	2	3	4
Control	87	20	30	32	33
Third Call	80	26	55	65	71

Table 3
Frequencies of Data Choices on Trials 2, 3, and 4 as a Function of Immediately Previous Data Choice and Conclusion

Prior Event*	Trial		
	2	3	4
Control Group			
D+	13/16	22/24	24/25
D-	1/1	0/2	2/3
N+	3/26	1/46	0/41
N-	9/44	5/15	3/18
Third-Call Group			
D+	19/21	40/43	46/47
D-	0/0	1/1	2/5
N+	3/20	4/22	3/16
N-	22/39	7/14	6/12

Note—Cell entries refer to number of subjects choosing diagnostic data (numerator) as a fraction of total number within each cell (denominator). For example, the upper left entry, 13/16 means that, of the 16 subjects who chose diagnostic data on Trial 1 and made a correct choice, 13 chose diagnostic data on Trial 2. *D and N refer to whether diagnostic or nondiagnostic data, respectively, had been chosen on the previous trial. The + and - refer to the correctness of the subjects' conclusions about the island.

Table 4
Percentages of Subjects Choosing Diagnostic Pairs,
Given the Events on the Previous Trial

Group	Prior Event	Trial		
		2	3	4
Third Call	N-	56	50	50
Third Call	N+	15	18	19
Control	N-	20	33	17

Effects of Exposure to Diagnostic Data

Comparison of data selection behavior between control and third-call subjects given a previously correct conclusion based on optimal choice is precluded by ceiling effects. On Trial 4, for example, the probabilities were .96 (24/25) and .98 (46/47) for the control and third-call groups, respectively. Thus, any effects of exposure to diagnostic data must be shown either in competition with or in interaction with reinforcement effects. Several comparisons are of interest. Compare third-call subjects and controls for data choices after having chosen a nondiagnostic pair of data and having been correct (the rows labeled N+ in Table 3). Over the three trials, the percentage of these subjects then selecting diagnostic pairs went from 12% to 0% for the controls, but from 15% to 19% for the experimental subjects. This suggests an effect, although a weak one, of the third-call exposure to diagnostic data when it is pitted against reinforcement. The above statement is chancy, since the frequencies of inappropriate choices in the third-call group get small as trials progress.

A more striking effect can be observed by comparing experimental and control groups' choices following selection of nondiagnostic data and having been wrong (N- in Table 3). Table 4 extracts these data from Table 3 and presents them as percentages, along with the percentages of third-call subjects who chose normatively after having chosen nondiagnostic data and having been right. Notice that the event that induces substantial change is the conjunction of diagnostic information (the third call) and feedback that the previous conclusion had been in error. Whether there is some sort of interaction, in the technical sense, between negative feedback and exposure to diagnostic information, or whether these factors summate is not determinable from these data.

Thus, negative feedback and exposure to diagnostic data coacted to induce subjects to change to a normative strategy. What happened after diagnostic data were correctly selected? Collapsing over all the rows in Table 3, we find that only a tiny minority of subjects ever chose nondiagnostic data after having chosen diagnostic data. There were five such subjects on Trial 2, seven on Trial 3, and six on Trial 4. The comparable frequencies for optimal choices following optimal choices were 33, 63, and 74. Thus, once a subject had chosen normatively, whether serendipitously or because of the experimental manipulation, that subject tended

to continue selecting normatively. The tendency to stick with the proper strategy when once used accounts for the growing separation of percentages over trials in Table 2. Of 21 third-call subjects who chose normatively and were correct on Trial 1, 19 did so on all three subsequent trials. There were 11 of 17 such subjects in the control group. Subjects who selected nondiagnostic data also tended to switch to diagnostic data after having observed the third datum. Of the 39 third-call subjects who chose nondiagnostic data and were wrong on Trial 1, 22 (compared with 9 of 44 controls) switched to normative selection on Trial 2. Of these, 21 made the correct conclusion, and 17 of these repeated the appropriate strategy on Trials 3 and 4. Thus, subjects were being absorbed into the D+ cells of Table 3.

The Assessment of Understanding

All subjects were classified as to whether they showed a clear understanding of the concept of diagnosticity by their conclusions and explanations on the last three pages. These pages were separated from the first five and coded independently by two of the authors without knowledge of performance on the data selection task. Subjects were credited with understanding only if they (1) concluded "can't tell" for the column and diagonal pages, (2) concluded correctly on the last page, and (3) made reference to needing data about the same pot characteristic from both islands on all three pages.

The data selection behavior choices were also classified into overall strategies. Subjects who selected diagnostic data and named the correct island on both of the last two trials were categorized as having an appropriate strategy. All others were classed as having a nonoptimal strategy. Table 5 shows the relation of data selection strategy and understanding. Collapsing over experimental and control, there was a relationship between strategy and understanding ($\chi^2 = 37.97$, $p < .01$). Table 2 suggests that the experimental manipulation led to an appropriate choice strategy, but Table 5 indicates that there was probably no influence of the experimental treatment on understanding over and above that mediated by the observation of diagnostic data. Note that 18 of 45 (40%) of third-call subjects and 8 of 23 (35%) controls who had the appropriate strategy understood the task. It seems that having chosen diagnostic

Table 5
Classification of Subjects by Strategies of Data Selection on
Trials 1-4 and By Whether Understanding Was Evidenced
on Pages 6-8

Group	Strategy	Understanding	
		Absent	Present
Third Call	Appropriate	27	18
	Inappropriate	33	2
Control	Appropriate	15	8
	Inappropriate	64	0

data was the critical factor in mediating understanding, whereas having seen a third sticker seems to have functioned to induce the choice of diagnostic data.

DISCUSSION

This study serves, first, as a replication of the pseudodiagnosticity phenomenon reported in Doherty et al. (1979); 67% of the students in the control group on Trial 4 selected formally worthless data, every one of whom drew conclusions about the island from which the pot had come. The task requires no complex mental arithmetic. It requires a qualitative understanding of the concept of diagnosticity and the ordinal comparison of the percentages uncovered, if the appropriate ones are indeed uncovered. The results on the selection of data again show a lack of that understanding.

The analysis of the last three pages also shows that few subjects understand diagnosticity. Forty-one percent of the subjects were classed as having appropriate choice strategies, and of those, only 38% were scored as having a verbalizable understanding of the normative principle. What led to the optimal strategies and to understanding? Reinforcement effects were powerful, but once subjects saw diagnostic data, a primarily cognitive influence was operative. When greater degrees of control over choice strategies are achieved, and more subjects demonstrate understanding, the question of the extent to which the subjects' understanding generalizes to other task domains will be assessed.

Why bother? Data selection is implicated in many forms of cognitive activity. If the bias that we have labeled pseudodiagnosticity occurs in this laboratory environment, then it likely occurs in other, more complex, cognitive tasks. If we can change it in the laboratory, and if the change generalizes, then perhaps we will be able to influence diagnostic (in the broadest sense) practice in a salutary fashion.

The other answer to "Why bother?" is that the pseudodiagnosticity phenomenon is interesting human inference behavior. It may be related to Wason's (1960) finding that subjects fail to consider that the data sequences they produce in the 2-4-6 task are as relevant to the hypothesis "increasing numbers" as they are to the hypothesis "numbers increasing by two." It may be tied to behavior in the concept-formation literature and the problem solving literature in which subjects tend to produce and test hypotheses singly and seriatim, rather than to test hypotheses against each other (Newell & Simon, 1972). If our speculation in the Doherty et al. (1979) pseudodiagnosticity paper was correct (i.e., that it is far easier for subjects to think about two data in relation to one hypothesis than one datum in relation to two hypotheses), this would be a

central fact about human inference. In this regard, note that Miller (1956), discussing how information was combined in multidimensional tasks, concluded that "it [is] better to have a little information about a lot of things than to have a lot of information about a small segment of the environment" (p. 88). This is true in the present situation, in which the "things" are alternative possibilities. Information in the form of many properties of one possibility may well be worse than useless: Such information is pseudodiagnostic. That outcome feedback can lead subjects away from this error is reassuring, but it is less reassuring that accidentally correct outcomes fortuitously strengthen erroneous strategies of information selection.

REFERENCES

- COHEN, L. J. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Science*, in press.
- DOHERTY, M. E., MYNATT, C. R., TWENEY, R. D., & SCHIAVO, M. D. Pseudodiagnosticity. *Acta Psychologica*, 1979, **49**, 111-121.
- EINHORN, H. J., & HOGARTH, R. M. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 1981, **32**, 53-88.
- HOGARTH, R. M. *Judgement and choice: The psychology of decision*. New York: Wiley, 1980.
- KERN, L., & DOHERTY, M. E. Pseudodiagnosticity in an idealized medical problem-solving environment. *Journal of Medical Education*, in press.
- MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81-97.
- NEWELL, A., & SIMON, H. A. *Human problem solving*. Englewood Cliffs, N.J: Prentice-Hall, 1972.
- NISBETT, R., & ROSS, L. *Human inference: Strategies and shortcomings of human judgment*. Englewood Cliffs, N.J: Prentice-Hall, 1980.
- SLOVIC, P., FISCHOFF, B., & LICHTENSTEIN, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, **28**, 1-39.
- TVERSKY, A., & KAHNEMAN, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, **185**, 1124-1131.
- WASON, P. C. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 129-140.
- WIMSATT, W. C. Reductionistic research strategies and their biases in the units of selection controversy. In T. Nickles (Ed.), *Scientific discovery: Case studies*. Boston: Reidel, 1980.

(Received for publication August 19, 1981.)