No Logic? No Problem! Using A Covariation Analysis On A Deductive Task

John B. Best
Department of Psychology
Eastern Illinois University Charleston, IL 61920
cfjbb@ux1.cts.eiu.edu

Abstract

Subjects were presented with previously played Mastermind games in the form of "Mastermind problems". Although each problem was formally deducible, and in some cases, overdetermined, subjects nevertheless usually failed to make more than a third of the potential deductions. A Bayesian model that treated the task as one of "probabilistic reasoning" rather than "logical deduction" accounted well for the performance of the lower performing subjects. It is argued that at least some of the reasoning failures seen on hypothesis evaluation tasks such as this one are produced in part by the solver's replacement of a "deduction" representation with a "probabilistic reasoning" representation.

Introduction

Studies using hypothesis evaluation paradigms suggest that the "confirmation bias" seen in the 2-4-6 task (Wason, 1960) results in part from subjects' inability to generate the hypotheses. That is, when the appropriate disconfirmatory hypotheses are already generated for them, then subjects who are instructed to do so can recognize disconfirmatory hypotheses, suggesting that the necessary logical operators are intact (Farris & Revlin, 1989). But there is at least one important proviso to this finding. In hypothesis evaluation paradigms (Farris & Revlin, 1989), or in rule discovery tasks in which the subjects are "debiased" by instructions (Gorman & Gorman, 1984), the context in which the materials are presented to the subjects almost always involves "logical deductions or scientific thinking". a contextual effect that seems very likely to be reflected in the subjects' representation of the task. This implies that the ability to generate or recognize disconfirmatory response may be a product of both the presence of the disconfirmatory hypotheses, and the "right" elements in the solver's representation. Whether people can routinely engage in disconfirmatory analyses when the context, and therefore perhaps the person's representation, are not so explicitly presented as "logical deduction" is the issue motivating this paper. I argue that, in such contexts, some failures on formally deductive tasks are produced in part by the subject's replacement of the concept of "logical necessity" with a concept of "probabilistic reasoning". What follows is some evidence to support this claim, as well as a model that duplicates some of the effects of "probabilistic reasoning" on a purely formal deductive task.

Method

Materials and Procedure

Four previously played Mastermind games of moderate complexity were presented as "Mastermind problems" in this study. Thus, in this form of Mastermind, subjects did not generate their own hypotheses. Rather, their task was to evaluate the hypotheses and feedback that had been produced in an effort to deduce the code. In each of the four problems, the set of hypotheses and feedback that were displayed contained information that was necessary and sufficient to permit the deduction of that problem's code.

Each of the 48 subjects was run individually. The experimenter explained the rules of Mastermind. also stating that in this form of the task, the subjects would not be generating their own hypotheses. The problems were presented in a way that simulated actual Mastermind play. That is, subjects saw each hypothesis and its associated feedback individually. Following the presentation of each hypothesis, the subject was asked to indicate the extent of his or her deductions on a response form. Subjects indicated two principal types of deductions: The subject marked an "assignment" when he or she was convinced that a particular letter was definitely a code member. The subject also indicated the purported location of the assigned letter. "Exclusions" were marked by the subject when he or she was positive that a particular letter was definitely not a code member. All previously presented hypotheses from that problem remained on view until after the problem's penultimate hypothesis was presented. The presentation order of the four problems was completely counterbalanced.

Results

The data were scored by counting the number of accurate assignments and exclusions marked by the subjects following the _n_th row of each of the four problems. Following the _n_th row of each problem, enough information was present to permit four assignments and two exclusions. One point was awarded for each such deduction. Mean performance across the four problems was 8.6 (maximum score = 24). Even though the specific assignments and exclusions were all logically deducible following the presentation of the _n_th hypothesis and feedback of each problem, the likelihood of the subject's correctly deducing assignments varied significantly both within and across problems. Of the 16 assignments in the four problems only 4 assignments were made accurately by a majority of the subjects. Five assignments were made correctly by 40-49% of the subjects. Three assignments were made correctly by 30-39% of the subjects, and four of the assignments were made correctly by only 10-29% of the subjects.

Moreover, it appears that the likelihood of a particular letter's being correctly assigned to a specific position was influenced by the number of times that the letter appeared in the problem at the same position. Specifically, in cases in which a letter appeared several times at the same location, and in which black feedback was given, subjects were likely to conclude that the letter must be correctly placed at the position where it appeared most of the time.

For example in Problem 2, 69% of the subjects correctly deduced that B must be placed at position 4. B appeared three times in the problem, each time at the same location (ratio of total appearances/different positions = 3/1), and was accompanied by a total of seven black feedback pins. But substantially fewer subjects (23%) correctly deduced that E must be placed at position 3 (ratio of total appearances/different positions = 3/3, or 1/1). This analysis suggests that the subjects were basing their deductions on the degree of covariation between a letter's placement and the occurrence of black feedback. Table 1 shows the results of similar computations carried for this and the other three problems used in the study, and it confirms the findings suggested by the initial problem. When a particular letter appeared in the array frequently and remained more or less stationary, the subjects were likely to conclude that this covariation enabled a necessary logical connection. But when a letter appeared infrequently, or appeared frequently in different locations, then subjects were less likely to deduce its assignment .

Table 1

Likelihood of Correct Assignment as a function of Appearance/Position Ratio

Correct Assignments Percentage Deducing	Total Appearances	Positions	Ratio
51% (or more), N=4	4.25	1.75	2.43
40-49%, N=5	3.40	2.40	1.42
30-39%, N=3	3.33	2.67	1.25
10-29%, N=4	2.00	2.00	1.00

Modeling the Deductions of Low-Performance Subjects

Subjects who accept the experimenter's depiction of the task may indeed represent the four problems as involving logical deduction. Such a representation would likely include logical operators whose function is to take various inputs from the problem array and produce deductions. While such operators may not always succeed in producing valid deductions, the situations in which they fail are presumably describable in terms of memory, or other exogenous demands on the cognitive system.

However. there are certainly other plausible ways of representing the problems, and these other forms of representation may predict outcomes that are more consistent with the observed findings than are the predictions of the "logical" model. For example, subjects who treat the logical problems as analogous to everyday problems in the real world may view covariation as a useful heuristic in making assignments. One way of conceptualizing the reasoner's task is to consider the reasoner as using the evidence that has accrued (the black feedback) in an effort to assess its effects on the likelihood of a particular hypothesis (namely, that a specific letter is assignable at a specific location) being true. Using a standard form of the Bayesian equation to represent this state of affairs we have:

$$p\,[L(P)/F = \frac{p\,[F/L(P)\;p\,[L(P)]}{p\,[\,F/L(P)\;p\,[L(P)] + p[F/{\sim}L(P)\;p[{\sim}L(P)]}$$

where p [L(P)/F] represents the probability of a particular letter (L) to position (P) assignment being true given that a certain feedback pattern (F) has been observed; P [F/L(P)] represents the likelihood of observing a certain pattern of feedback given that a letter to position assignment is true, and p [L(P)] represents the prior probability of any specific letter to position assignment being true. p [F/~L(P)] represents the probability of the feedback pattern being observed given that the letter to position assignment is not true, and p [~L(P)] represents the prior probability that the letter to position assignment is not true. Computing some of the equation's terms is straightforward: Given that any of the six available letters can be assigned to, any specific position, p[L(P)] can be estimated at .17, and p [~L(P)] = 1 - p [L(P)]. The estimation of p [F/L(P)] involves computing for any specific letter to position assignment (as in letter A in position 1) the proportion of all codes (of which there are 360) that would generate this particular feedback pattern through this hypothesis, if A were indeed correctly located at position 1. That is, of the 60 codes in which A is correctly located at position 1, what proportion would be followed by the specific feedback "1 Black, 1 White" if this hypothesis had been played? The same logic is used to estimate p [F/~L(P)]. That is, of the 300 codes that do not have A correctly located at position 1, what proportion would be followed by the specific feedback "1 Black, 1 White" if the hypothesis A B C D had been played?

To run the Bayesian model, each of the six letters was initialized as a four place vector with prior probabilities of .17 in each of the four slots, thus creating a 6 X 4 matrix. The probabilities for each letter to position assignment were updated after each hypothesis, treating the previous hypothesis's posterior likelihood as the current hypothesis's prior likelihood. A normalization procedure was applied after each updating cycle for any row vector whose probabilities exceeded unity.

Testing the model's predictions involved splitting the sample of subjects into two subgroups based on their overall performance. The logic here is that subjects who have access to logical operators, and who are motivated enough to use them, will be unlikely to rely on the covariation analysis to assign letters, and thus will be likely to correctly assign letters whose logical status is clear regardless of how such letters look to the covariation analysis. On the other hand, subjects who do not have access to such operators, or who are not motivated enough to engage in the fairly effortful analysis required to use them should be likely to rely on covariation analysis which the Bayesian model should pick up. The subjects were divided into two groups. High performance subjects (N = 25, M = 12.6) were those who scored 9 or better on the four problems, while low performance subjects (N =23, M = 3.7) were those whose score ranged from 0 to 8. Expected frequencies of letter to position assignments were computed for each letter in each problem by

multiplying the elements of the letter's row vector by the number of subjects in each subgroup who actually made assignments of that letter. Eight chi-squares (4 problems X 2 subgroups) were used to evaluate goodness of fit between expected and observed frequencies of letter to position assignments. For all four
problems in the high performance group, the chi-squares were significant at $p < .001$, indicating poor goodness of fit. But for all four problems in the low performance group, the chi-squares failed to reach statistical significance ($p > .05$), suggesting a reasonable conformity between the model's predicted assignments and those that the low performance subjects actually made. Moreover, the deviations from the model's predictions made by the high performance subjects were always in the direction indicated by a logical analysis rather than by a covariation analysis.

Discussion

These findings suggest that the deductive performance of subjects who do particularly poorly on this task is the result of an over-reliance on a covariation analysis, and this covariation analysis can be modeled effectively using principles of Bayes' theory. This is not to say that the lower-performing subjects are engaged in a sophisticated Bayesian analysis. Rather, such subjects seem to be engaged in a type of probability estimation, and these estimations seem to be capturable in a Bayesian model.

In addition, the findings are suggestive that the use of the covariation analysis is driven by the subject's representation of the problem as an example of "real life" reasoning, rather than as a problem in formal logic. That is, although few of the subjects in the study had studied logic formally, as college students they were well aware that logic is a formal discipline, and the topic of university coursework. And from what they know of all academic disciplines, such subject matter is approached with considerably more rigor and intensity than is the "corresponding" subject matter in day-to-day life. Thus for example, I may mix the ingredients in a cake recipe with considerably less precision than I would use to mix the ingredients in the chemistry lab, knowing that the cake will probably turn out regardless. When applied to the current situation, the typical subject may be well aware that to be "logical" might mean exercising greater precision in the reasoning process, including being more demanding of evidential standards, being more skeptical, being more alert to discrepancies of appearance and reality, and so on. Presented as it is in this context, that is, as a game, Mastermind might be not necessarily invite the more rigorous approach characteristic of subjects in studies of "logical deduction".

One of the issues in the literature on hypothesis evaluation concerns the ability of humans to recognize disconfirmatory hypotheses when such hypotheses have been generated for them. Researchers typically find that people are good at discerning disconfirmatory hypotheses in this situation. To the extent that the Mastermind problems used in this study can be seen as analogs to the reasoning vignettes used by Farris and Revlin (1989), these findings suggest that there are situations in which simply having the relevant hypotheses available does not mean that the subjects can engage in the modus tollens like reasoning of necessary logical operations. As we saw in this study, in an absolute sense the subjects did not make all the necessary logical operations that were available. Finally, these findings may be seen as an instantiation of the rational analysis approach (Anderson, 1990) that has proved useful in the areas of memory and categorization. As applied to reasoning, such an analysis suggests that, given that some individuals understand formal reasoning as equivalent to model of causation that might be derived from daily experiences, their deductions within that context are orderly and plausible.

References

Anderson , J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Farris, H. H., & Revlin, R. (1989). Sensible reasoning in tasks: Rule discovery and hypothesis evaluation. *Memory & Cognition, 17,* 221-232.

Gorman, M. E., & Gorman, M. E. (1984). Comparison of disconfirmatory confirmatory and control strategies on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology, 36A,* 629-648.

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