The Effect of Presentation Order in Clinical Decision Making

JOHN P. W. CUNNINGTON, JEFFREY M. TURNBULL,
GLENN REGEHR, MICHAEL MARRIOTT, and GEOFFREY R. NORMAN

Medical diagnosis is, at its heart, a categorization task in which the clinician must use available information to determine the most likely diagnostic category for the patient's problem. In most clinical situations the data are gathered sequentially with the order of the acquisition under the control of the clinician. This sequential nature of data acquisition in diagnostic tasks poses a potential problem.

Research in the domain of social cognition has shown that information received early in hypothesis testing frequently has an inappropriately powerful influence on the decision making process. This influence, referred to as the *primacy effect*, was demonstrated by Asch,[1] who showed that the same set of personality descriptors presented to subjects in different orders led to substantially different opinions about the individual being described. Thus, a description beginning with the traits "intelligent" and "industrious" led subjects to form a favorable impression that overpowered the presence of later traits such as "impulsive," "stubborn," and "critical." By contrast, when the description began with the more negative traits and ended with the more positive traits, the overall impressions of the individual were far less favorable. In a similar demonstration, Jones et al.[2] asked subjects to watch an individual try to solve 30 multiple-choice analogy problems of roughly equal difficulty. All subjects solved 15 of the 30 items; however, for one group of subjects a majority of correct responses occurred in the first half of the test, whereas for the other group a majority of correct responses were in the second half. Observers who saw the test taker score well in the first half of the test remembered that individual as having solved more problems, rated the individual as more intelligent, and predicted that the individual would solve more problems correctly in future than did the observers who saw the test taker score well in the second half of the test.

A similar phenomenon has been proposed in the field of medicine in the form of premature closure during a diagnostic task. Elstein[3] demonstrated the central role of an early hypothesis in the diagnostic process, and Barrows et al.[4] showed that this early hypothesis initiates a search for confirmatory information. More recently, Norman et al.[5] showed that the presence of a small amount of clinical information on a film bag influenced not only the diagnostic interpretation of radiologists but also the likelihood of their perceiving features on the radiograph. These demonstrations seem to indicate that the danger of premature closure arises from the tendency to search out, and therefore to *see* only data consistent with the early hypothesis.

While this tendency to notice only confirmatory evidence is undoubtedly important, the data from the social cognition literature suggests that a further danger exists. That is, even evidence presented in a clear and unambiguous manner will be given less weight if it is inconsistent with earlier data.[6] At least in generating impressions about an individual's personality, it appears that evidence that disconfirms an early hypothesis is not merely missed but is actively discounted. If this is true in medical diagnosis, then the dangers of generating an incorrect hypothesis early in the diagnostic process are further exacerbated. Using a design similar to that of Asch[1] and Jones et al.,[2] the present study explores the extent to which this discounting phenomenon is present in the diagnostic processes of novice, intermediate, and expert medical diagnosticians.

Hypothesis

When presented with history data for cases in which two diagnoses are of equal likelihood, subjects are more likely to select the diagnosis for which confirmatory data was presented early in the sequence rather than late.

Method

Case development. As the intent was to recreate the actual process of data acquisition at the bedside or in the office, the initial plan was to create scenarios for common clinical problems with both history and physical examination information. This, however, would have greatly complicated studying the order effects, as it would be necessary to look at the presentation order of the history information, the presentation order of the physical examination information, and at their interactions. This was thought to have too great a likelihood of obscuring whatever ordering effects might actually exist, and therefore the decision was made to look only at information presented in the history.

Two internists developed case histories for ten common presenting problems in internal medicine. The problems were selected such that the symptoms suggested two frequently competing diagnoses (for example, pulmonary embolism versus pneumonia, cardiac ischemia versus esophageal reflux, infectious colitis versus inflammatory colitis). The cases were designed in a series of discrete, short statements that replicate the items of information acquired by clinicians in the process of history taking. For example:

> A 43-year-old woman is brought to the emergency room by her husband at 2 a.m. because of acute shortness of breath.
>
> The dyspnea had occurred suddenly three hours earlier and had awoken her from sleep.
>
> She complained of retrosternal chest pain which was worse on deep breathing.

Typically these histories contained 12–15 statements. The cases were structured such that the totality of the information provided equally supported both competing diagnosis. For example, in a woman with pleuritic chest pain, information supporting the diagnosis of pulmonary embolism—such as oral-contraceptive use, sharp chest pain, and acute dyspnea—was provided along with information supporting a diagnosis of pneumonia, such as fever, cough, and purulent sputum. (The complete list of problems and history statements are available from the authors.)

Equivalency of diagnoses. We wanted each case to have two essentially equally probable, competing diagnostic possibilities. To accomplish this the complete set of statements for each case was circulated to six internists unfamiliar with the project. For each case, the internists were asked to give their top two diagnoses and the probability of each. This information was collated to determine whether the blinded internists arrived at the same two diagnoses as the question developers, and to determine whether the probabilities for the two diagnoses were about equal. The results were discussed at a subsequent meeting of the investigators and the internists, and for cases where the probabilities were not equal, changes in the histories were made to equalize the diagnostic ambiguity. The internists were also asked about the credibility of the cases, and all appeared realistic and credible.

Case sequences. Each individual history statement was identified as either being supportive of Diagnosis A, supportive of Diagnosis B, or neutral (not helpful in distinguishing between the diagnoses). Case histories were then reassembled in two formats, Format A, in which all the information consistent with Diagnosis A was presented at the beginning of the list, and the information consistent with Diagnosis B was presented at the end, and Format B, in which the order of the case information was reversed. The result was two sets of ten cases with identical information but opposing presentation orders.

The materials were set up to be presented by a computer that randomized the subject to receive either Format A or Format B. Each case was presented by giving the subject the opening statement on the screen. Subjects advanced to each subsequent statement at their own speed. Each statement appeared in isolation on the screen, i.e., lines did not accumulate. When the subjects had viewed all the information for the case they were asked to type in their top two diagnoses and the probabilities of each.

Subjects. Subjects were 16 final-year medical students, 36 first- and second-year residents in internal or family medicine, and 15 academic internists. Students and residents were paid for their participation. Although the subjects were volunteers who agreed to participate, the overall rate of agreement to participate was high in all groups.

Data handling and analysis. Because the diagnoses were entered by the subjects, there were numerous synonyms for each. Before analysis, all the diagnoses were listed; one internist, blinded to the data, grouped the material into diagnostic categories, i.e., synonymous with Diagnosis A, synonymous with Diagnosis B, or Other. The diagnosis with the highest probability was selected as the subject's primary diagnosis. The analysis was confined to this primary diagnosis, where it was coded as Diagnosis A, Diagnosis B, or Other. We examined the proportion of subjects who selected each diagnosis under each of the two presentation sequences (format A vs format B).

No formal analysis of reliability or validity of the measure was conducted for three reasons. First, we believed that accuracy of diagnosis has sufficient face validity in and of itself. Second, the nature of the design is such that we should show some improvement in accuracy with educational level (where accuracy, in these cases, amounts to selecting either of the target diagnoses). Finally, the intent of the study was *not* to identify good or poor clerks or residents; thus reliability, a measure of discrimination within an educational level, is actually irrelevant. In fact, as Cronbach[7] showed in 1957, heterogeneity of subjects, a prerequisite for reliability, is a liability for experimental designs.

These frequency counts were subjected to a log-linear analysis with three factors (Diagnosis × Education × Format). Log-linear analysis is a general form of non-parametric analysis suitable for multiway frequency tables, and it results in tests of significance for main effects and higher order interactions. Because we were not interested in the effect of sequence on Other diagnoses, these were omitted from the statistical analysis, so a $2 \times 3 \times 2$ log-linear model was tested.

It was expected from the primacy hypothesis that subjects presented first with data consistent with Diagnosis A followed by data consistent with Diagnosis B, would be more likely to select Diagnosis A for each case, and that those presented first with Diagnosis B data would be more likely to select Diagnosis B. This would be evidenced by a Format × Diagnosis interaction. If there was an effect of educational level, this would be evidenced by a three-way interaction (Format × Diagnosis × Education).

Results

The subject response to each case was used as the unit of analysis. Thus, 670 data points are represented in the analyzes. Table 1 displays the frequency of Diagnosis A vs Diagnosis B vs Other, separately for clerks, residents, and faculty under each of the two presentation formats.

The table shows that the case histories were sufficiently well designed that subjects overwhelmingly chose as their top two diagnoses the target diagnoses. Only 11% of clerks, 6% of residents, and 7% of faculty selected as one of their top two probabilities diagnoses that we did not expect. The table also shows that, as we had anticipated, the top two choices were about equally attractive, i.e., about 50% support for diagnosis A and 50% support for diagnosis B. This held up on a case-by-case basis where in nine of the ten cases the less popular of the two diagnoses was selected by more than 20% of subjects.

There was a clear tendency for the subjects to favor the diagnosis for which the information was presented first, although the magnitude of the effect varied considerably across educational level. For clerks, this amounted to a change of about 18% from one diagnosis to the other; for faculty, about 8%; and for residents, virtually none. The log-linear analysis showed a significant overall effect of sequence on diagnosis (chi-square(1) = 3.74, $p = .05$) and a marginal interaction with educational level (chi-square(2) = 4.26, $p = .11$).

Discussion

The primacy effect is a robust phenomenon in the cognitive psychology literature; Hogarth and Einhorn[8] document more than 30 studies in which this effect has been observed (primarily in subjective judgments of personality traits). Our study has shown that the sequence of data presentation can influence medical diagnosis. To our knowledge, this is the first time that this primacy effect has been observed to play a role in medical decision making. Perhaps this is not surprising; intuitively, it seems unlikely that physicians solve clinical problems in a strictly Bayesian manner, and that data elicited early in the encounter may introduce a systematic bias to clinical decision-making. Indeed, Hogarth and Einhorn point out that Anderson[9] suggests that the primacy effect can be explained by a process of *attention decrement* whereby people pay less attention to successive items of evidence. As we have demonstrated that the order of data presentation by the researcher may affect diagnosis, it is reasonable to assume that the order of data acquisition by the student (and possibly the clinician) may have a similar biasing effect.

Table 1. Frequencies of Diagnosis A, Diagnosis B, and Other by Three Groups for Two Presentation Formats

	Format	
Diagnosis	A then B	B then A
By clerk		
A	55 (69%)	41 (51%)
B	17 (21%)	30 (37%)
Other	8 (10%)	9 (12%)
By resident		
A	97 (54%)	91 (50%)
B	76 (42%)	74 (41%)
Other	5 (4%)	13 (7%)
By faculty		
A	53 (56%)	30 (50%)
B	28 (34%)	27 (45%)
Other	8 (9%)	3 (5%)

Although the primacy effect does not appear to have been previously documented in medical decision making, a *recency effect*, in which data acquired later in the history have a disproportionate effect, has been shown by Bergus et al.[10] In their study of a case of a patient with brain metastasis from lung cancer, they showed that subjects were more likely to come to the correct diagnosis if the past history of lung cancer was given late in the history. Their study may, however, have limited generalizability as it is limited by the context specificity inherent in studying a single problem. In our study of ten cases, we observed no evidence of a recency effect.

A limitation of our study is that we used subjects who were available and who agreed to participate. There is, however, no reason to believe that the decision-making behavior of subjects who were willing to participate would be different from randomly selected subjects. A further limitation is that we looked only at common medical problems and at the effects observable on the history. Although we have no reason to believe that the effect is not more widely generalizable, the demonstration that this is the case awaits further study. In addition, while it appears that there may be a relationship with educational level, we had too few subjects to demonstrate this. It is, nonetheless, tempting to speculate that with increased clinical experience the primacy effect may be moderated or disappear.

The primacy effect has significant implications for medical education and evaluation. In education, medical teachers need to be aware of the possibility of cognitive biases in decision making resulting from data presentation order. Students themselves may benefit from knowing that premature diagnosis may lead to false conclusions, an argument that medical faculty have long advocated, notwithstanding the observation that most experienced clinicians come to a probable diagnosis very early in the clinical encounter.

In evaluation of medical students and residents it may be worth considering that the response to a multiple-choice question (MCQ) may be affected not only by knowledge of the subject but also by the presentation order of the data, either in the stem of the question or perhaps even in the response. Future research in this area should address whether manipulation of presentation order of data items in MCQs actually alters the response, and whether the primacy effect can be modified either by increasing medical training or by awareness of the potential biasing effects of presentation order.

In summary, this study has demonstrated the effect of primacy as a cognitive bias in medical decision making with potentially important implications for teaching and testing.

This study supported by a Research in Medical Education grant from the Association of Canadian Medical Colleges/Medical Research Council.

The authors gratefully acknowledge the help of Robert Hayward, MD, and John Horsman in preparing the materials for computer presentation.

Correspondence: John Cunnington, MD, McMaster University HSC 3U2, 1200 Main St. W., Hamilton, Ontario, Canada L8N 3Z5. e-mail: <cunningt@mcmaster.ca>.

References

1. Asch S. Forming impressions of personality. J Abnorm Soc Psychol. 1946;41:258–90.
2. Jones EE, Rock L, Shaver KG, Goethals GR, Ward LM. Pattern of performance and ability attribution: an unexpected primacy effect. J Pers Soc Psychol. 1968;10:317–40.
3. Elstein AS, Kagan N, Shulman LS, Jason H, Loupe MJ. Method and theory in the study of medical inquiry. J Med Educ. 1972;47:85–92.
4. Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning of randomly selected physicians in general medical practice. Clin Invest Med. 1982;5:49–55.
5. Norman GR, Brooks LR, Coblentz CL, Babcook CJ. The correlation of feature identification and category judgments in diagnostic radiology. Mem Cogn. 1992;20:344–55.
6. Anderson NH. Averaging versus adding as a stimulus combination role in impression formation. J Exp Psychol. 1965;70:394–400.
7. Cronbach LJ. The two disciplines of scientific psychology. Am Psychol. 1957;12:671–684.
8. Hogarth RM, Einhorn HJ. Order effects in belief updating: the belief-adjustment model. Cogn. Psychol. 1992;24:1–55.
9. Anderson NH. Foundations of information integration theory. New York: Academic Press, 1981.
10. Bergus GR, Chapman GB, Gjerde C, Elstein AS. Clinical reasoning about new symptoms despite preexisting disease: sources of error and order effects. Fam Med. 1995;27:314–20.