Believing Is Seeing: The Influence of a Diagnostic Hypothesis on the Interpretation of Clinical Features

VICKI R. LEBLANC, LEE R. BROOKS, and GEOFFREY R. NORMAN

Many clinical decisions are made on the basis of information gathered from patients, whether in the form of chief complaints, patient appearance, or physical findings. Errors made in gathering and interpreting such clinical information could result in serious consequences in patient care, such as delays or errors in treatment.[1] Little research has been devoted to understanding the mechanisms underlying such errors.

In psychology, it has long been known that context influences perception. Phenomena such as visual illusions and the word superiority effect, in which letters are more easily recognized when presented in words rather than in unrelated letter strings,[2] are pertinent examples of how the context influences feature interpretation. Similarly, research in radiology demonstrates that information such as the location of tenderness and swelling[3] and tentative diagnoses[4] increases the likelihood that physicians will detect fractures and lesions. Even with less ambiguous visual stimuli, such as patient appearance[5] or electrocardiograms,[6] the consideration of the correct diagnosis leads to an increase in detecting clinical signs compared with having no diagnosis in mind.

These studies demonstrate that the identification of clinical signs is influenced by the diagnostic hypotheses held by diagnosticians. However, by measuring only performance in reporting the depicted correct features from the clinical stimuli, these researchers do not allow for conclusions to be drawn regarding the specific influence of a diagnostic hypothesis on feature identification. Once a diagnosis is considered, it can activate a representation of the disease presentation, bringing to mind the possible features that can be visible on a patient suffering from the given condition. The diagnosticians can then run through this list of features, checking for the presence or absence of each. This diagnosis would thus serve as a focus of attention, determining which features to look for and where to look for them. Additionally, the diagnostic hypothesis might have a stronger impact of inducing a bias in the identification of the observed physical characteristics, e.g., knowing that a moon-shaped face is a feature of Cushing's disease and believing the diagnosis to be Cushing's might lead diagnosticians to interpret a slightly obese face as moon-shaped. This interaction between the diagnosis and feature identification, if present, might be sufficient to lead clinicians to report features that are not present in patients. Given that there are a number of clinical situations where potentially biasing effects on diagnostic suggestions can occur, such as in referral letters or patient charts, it is important to understand the impact of such suggestions on the interpretation of clinical information.

In the present study, we investigated whether the influence of the diagnosis (and accompanying brief case history) is strong enough to bias the interpretation of clinical information. Medical students and residents are suggested either the correct diagnosis or an alternate but plausible diagnosis prior to reporting all clinically important features from photographs of patients. If a diagnosis simply focuses attention on the relevant features, participants who are biased toward an alternate diagnosis should report fewer of the correct features than participants biased toward a correct diagnosis. However, the two groups should not differ in terms of reporting features that are consistent with the alternate diagnosis but not present in the photograph. Alternatively, if a diagnostic hypothesis does change the interpretation provided to the clinical data, participants biased toward an alternate diagnosis should be more likely to misinterpret correct features or normal variations in appearance as features supporting the alternate diagnosis. Therefore, they should report more alternate features than participants biased toward the correct diagnosis.

Method

Participants. Twenty medical students and 20 family medicine residents from McMaster University's Medical School volunteered to participate in this study. The students were recruited as they began their clerkship training (19th to the 23rd month of a three-year program), and the residents had completed a minimum of six months to a maximum of 18 months of a 24-month residency program.

Materials. Ten head-and-shoulders photographs were selected from textbooks or physician slide libraries and were considered to be classic, or prototypical, representations of a given diagnosis. Eight of those photographs served as the test stimuli and were selected because plausible alternate diagnoses could be generated for them. The first two photographs were used in practice scenarios to set up the atmosphere that the diagnoses suggested to the participants were plausible.

Case histories biasing to the correct and to the alternate diagnoses were generated for each of the eight test cases. The correct case histories were the same as those used in the Brooks et al.[5] study. The alternate diagnoses were generated by choosing a "feature" in each photograph around which an alternate but plausible diagnosis could be generated. These "features" were due either to normal variations (i.e., tanned skin interpreted as jaundice despite white sclerae), or were created by the reinterpretation of a cardinal feature of the correct diagnosis (i.e., a moon-shaped face reinterpreted as facial edema). The case histories were generated by a general internist and independently verified by another general internist to assure that they were appropriate and plausible, given the information contained in the photographs.

For each of the correct and alternate diagnoses, a list of its features was generated from a leading medical textbook.[7] Two expert general internists were then asked to independently indicate which of the features of the correct diagnoses were present in the photographs, as well as which features of the alternate diagnoses could be reported if students misidentified either a correct feature or a characteristic caused by normal variation. Only those features for which both experts agreed on their presence (or potential presence) were recorded. This resulted in a sum of 21 features of the correct diagnoses and 25 features of the alternate diagnoses. The correct and alternate features were mutually exclusive; the identification of a specific feature provided support for either the correct or the alternate diagnosis, but not both.

Procedure. Participants saw all ten scenarios, including the two practice scenarios presented first. The eight test scenarios were randomly divided into two groups of four, labeled A and B. Half of the participants were biased toward the correct diagnoses on the scenarios of group A, and biased toward the alternate diagnoses on the scenarios of group B. The other half of the participants were biased toward the correct diagnoses on the scenarios of group B, and biased toward the alternate diagnoses on the scenarios of group

A. The ordering of the eight test scenarios was randomized and kept the same for both groups.

The procedure for each of the scenarios was as follows. Participants were presented with the head-and-shoulders photograph, below which they read the brief case history (one to two lines) and the tentative diagnosis (correct or alternate). They were then asked to write down all the clinically important features present in the photographs. They received further instructions to include anything that could rule in or rule out a diagnosis, and any feature that was abnormal, even if it could not be linked to any specific diagnosis. Once they had written down the features, the students were asked to rate the likelihood (scale of 0% to 100%) of the suggested diagnosis, and any other self-generated diagnosis(es).

Measures. To assess the diagnostic decisions of the participants, there were two measures of interest: (1) how often they concluded for the correct diagnosis; for each scenario, they were given a 1 if they gave the highest likelihood rating to the correct diagnosis and a 0 if they gave the highest likelihood rating to another diagnosis; and (2) how often they concluded for the alternate diagnosis; for each scenario they were scored a 1 if they gave the highest likelihood rating to the alternate diagnosis and a 0 if they gave the highest likelihood rating to another diagnosis. There were also two measures of feature identification: (1) the percentage of correct features and (2) the percentage of alternate features reported. These four measures were submitted to separate $2 \times 2 \times 8$ mixed-design analyses of variance (ANOVA) with the condition (correct diagnosis and alternate diagnosis) and level (students, residents) as between-subject variables and the scenarios as a repeated-measures variable. The scenarios were anticipated to be variable in difficulty and were included in the analyses to account for some of the variance in the measures.

Results

Predictably, the diagnostic decisions of both residents and students were influenced by the case history and the tentative diagnosis presented. Overall, the participants decided in favor of the correct diagnoses more often when they were suggested the correct diagnoses, 77.2% versus 8.8%, $F_{(1,36)} = 239.36$, MSE = .160, $p < .05$. Alternatively, the participants decided in favor of the alternate diagnoses more often when it was suggested to them, 65.8% versus 9.8%, $F_{(1,36)} = 238.89$, MSE = .140, $p < .05$. As shown in 1, the performances of the residents and the students were remarkably similar. This was supported by a lack of significant main effects of level for concluding for the correct diagnoses, $F_{(1,36)} = .881$, MSE = .160, $p = .35$, and for concluding for the alternate diagnoses, $F_{(1,36)} = .036$, MSE = .140, $p = .85$, as well as a lack of significant level-by-condition interactions for concluding for the correct diagnoses, $F_{(1,36)} = .003$, MSE = .160, $p = .96$, and concluding for the alternate diagnoses, $F_{(1,36)} = .066$, MSE = .140, $p = .80$.

Consistent with previous research, participants identified more correct features when biased toward the correct diagnoses than when biased toward the alternate diagnoses, 48.9% vs. 36.9%, $F_{(1,36)} = 11.44$, MSE = .591, $p < .05$. As shown in Table 1, the performances of the students and the residents were similar, as supported by a lack of main effect of level, $F_{(1,36)} = .779$, MSE = .591, $p = .38$, and a lack of significant level-by-condition interaction, $F_{(1,36)} = .022$, MSE = .591, $p = .88$.

These results support the hypothesis that the diagnosis serves to alter the interpretation given to physical characteristics. When biased toward the alternate diagnoses, the participants identified more alternate features than when biased toward the correct diagnoses, 24.7% vs. 7.4%, $F_{(1,36)} = 100.09$, MSE = .237, $p < .05$ (see Table 1). Again, the performance of the residents was similar to that of the students, as supported by a lack of main effect of level, $F_{(1,36)} = .983$, MSE = .237, $p = .98$, and a lack of significant level-by-condition interaction, $F_{(1,36)} = 2.57$, MSE = .237, $p = .12$.

TABLE 1. Diagnostic Conclusions and Feature Identification of Participants as a Function of the Biasing Condition

	Bias			
	Correct Diagnosis (%)		Alternate Diagnosis (%)	
	Mean	SE	Mean	SE
Scenarios				
Concluded for correct diagnoses				
Students	80.0	3.6	6.9	2.6
Residents	80.6	4.1	12.4	3.5
Concluded for alternate diagnoses				
Students	3.1	1.9	66.9	5.0
Residents	0	0	69.4	4.7
Features				
Of correct diagnoses				
Students	48.8	3.9	35.1	4.3
Residents	52.7	3.9	40.0	4.4
Of alternate diagnoses				
Students	5.7	1.5	26.5	2.8
Residents	8.4	1.6	23.8	2.4

Discussion

The influence of a hypothesized diagnosis had an impact on both the diagnostic ratings and the interpretation of features. The strong influence on the diagnostic ratings was not surprising given the anecdotal reports from clinicians who state that they rely strongly on case histories when generating diagnostic hypotheses. As the case histories used in this study were strongly supporting of the suggested diagnosis, it is not surprising that our biasing manipulations had a strong impact on diagnostic decisions. It does not follow that such effect would also be observed with the interpretation of features, as this later process is generally assumed to be independent of other sources of information. However, as demonstrated in this study, the interpretation of features is not independent of other sources of information. Further, the influence of the hypothesized diagnosis is twofold. First, it serves to focus the attention of students and residents on relevant features. For example, they were more likely to correctly interpret a superclavicular lymph node when considering stomach cancer than when considering liver cancer. Additionally, the diagnosis serves to change the interpretation given to the physical characteristics. In one scenario, participants misinterpreted the tanned skin as jaundice more often when biased toward liver cancer than when biased toward stomach cancer, despite the patient's white sclerae. In another scenario, they were more likely to misinterpret the parotid swelling of the boy with mumps as a moon-shaped face when biased toward Cushing's disease. Such findings are remarkable, given that the selected photographs are considered by experts to be prototypical, or "textbook case" examples of the disease presentations. In a separate study,[8] we investigated the robustness of this phenomenon and observed that these effects are also present when the credibility of the suggested diagnosis is reduced by matching it with a case history suggestive of another diagnosis.

The results of this study further indicate that the biasing effects on feature interpretation are not the result of a lack of clinical experience. Residents having between six and 18 months of training in the extraction of features from patients' appearance are as susceptible to the biasing effects of the tentative diagnosis as were second-year medical students. Such findings pose a challenge to researchers who argue that clinicians shift from using backward reasoning to forward reasoning as they acquire experience.[9] Had the

residents been using more forward reasoning than the students, their identification of the features should have been less strongly influenced by the diagnosis. Rather, these results suggest that the reasoning process of clinicians is an interactive process between knowledge and incoming clinical information. Models of such interactive processes have been proposed to explain language processing.[10] In such models, individuals bring with them knowledge about the general properties of objects whenever they perceive such objects. The incoming "bottom-up" information from the features interacts with the "top-down" information, what people know about their environment, to determine what is seen. An alternative explanation, which does not refute the argument of a shift to forward reasoning with expertise, is that the gap in experience between the students and the residents might not have been large enough to allow for the development of forward reasoning. The remarkable similarity between the performance of the students and the residents suggests that residents might be closer to being novice diagnosticians than originally thought. To test the two possibilities, a similar study needs to be conducted with experienced family physicians. Although these results do not provide definitive evidence against theories of differential clinical reasoning, they do refute the possibility that a diagnostic hypothesis will influence the identification of feature only in novices who have little clinical experience.

In addition to the relatively narrow range of clinical experience of our participants, a limitation of the present study was the use of highly visual clinical stimuli. As mentioned previously, anecdotal reports from clinicians suggest that many clinical decisions are made on the basis of the verbal reports of patients. This suggests that errors made in the interpretation of clinical signs may play a relatively minor role in the final clinical decisions. To address this possibility, similar studies are planned with different clinical stimuli to determine whether diagnostic hypotheses will also influence the interpretation of the verbal information received from patients.

Regardless of whether the findings of this study generalize to verbal reports, the present paper documents a potentially important source of clinical errors. If, as argued by some researchers,[11] novices and experts generate diagnostic hypotheses early in the clinical encounter, the subsequent gathering and interpretation of clinical signs is likely to be guided by these hypotheses. This poses a challenge to medical educators and researchers to devise studies or educational interventions aimed at investigating how to reduce clinical errors. Possible interventions are to increase exposure to visual signs and their variable presentations when teaching diagnostic categories or to teach clinicians to consider alternative diagnostic hypotheses during the clinical encounter.

Correspondence: Vicki LeBlanc, PhD, Center for Research in Education at the University Health Network, University of Toronto Faculty of Medicine, 200 Elizabeth Street, Eaton South 1-604, Toronto, Ontario, Canada, M5G 2C4.

References

1. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. Acad Med. 1999;14(10 suppl):S138-S140.
2. Reicher GM. Perceptual recognition as a function of meaningfulness of stimulus material. J Exp Psychol. 1969;81:274–80.
3. Berbaum KS, Franken EA, Dorfman DD, et al. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Invest Radiol. 1986;21:532–9.
4. Berbaum KS, El-Khoury GY, Franken EA, Kathol M, Montgomery WJ, Hesson W. Impact of clinical history on fracture detection with radiography. Radiology. 1988;168:507–11.
5. Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. Psychol Sci. 2000;11(2):112–7.
6. Norman GR, Brooks LR, Colle C, Hatala R. The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. Cogn Instruction. 1999;17:433–48.
7. Fauci AS, Braunwald E, Isselbacher KJ, et al. (eds). Harrison's Principles of Internal Medicine. 14th ed. New York: McGraw–Hill, 1998.
8. LeBlanc VR, Brooks LR, Norman GR. What are the relative influences of case history and tentative diagnosis on feature interpretation? Poster presented at the annual meeting of the American Association of Medical Colleges (Research in Medical Education), Chicago, IL, November 2000.
9. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. Cogn Sci. 1986;10:91–116.
10. McClelland JL, Rumelhart DE. An interactive activation model of context effect in letter perception: part 1. an account of basic findings. Psychol Rev. 1981;88: 375–407.
11. Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. Med Educ. 1981; 15:315–22.