INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

UMI®

THE INFLUENCE OF A TENTATIVE DIAGNOSIS ON THE

IDENTIFICATION OF FEATURES FROM PATIENT

APPEARANCE

By

VICKI R. LEBLANC, B.Ps., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requiements

for the Degree

Doctor of Philosophy

IDENTIFICATION OF FEATURES FROM PATIENT

APPEARANCE

McMaster University

DOCTOR OF PHILOSOPHY (2000)   McMaster University

(Psychology)   Hamilton, Ontario

TITLE : The Influence of a Tentative Diagnosis on the Identification

of Features from Patient Appearance.

AUTHOR : Vicki R. LeBlanc,   B.Ps. (Universite de Moncton),

M.Sc. (McMaster University)

SUPERVISOR : Professor Lee R. Brooks

NUMBER OF PAGES : x, 117

Abstract

The clinical signs that a physician can identify from the appearance of a patient represent an important source of information, upon which the diagnostic decision is nominally based. Most of the research in medical education emphasizes the organization of medical knowledge or the reasoning processes based on these signs. This emphasis carries the implicit assumptions that identifying features is not the major problem and that evaluation of the clinical signs occurs largely independently of consideration of the diagnosis. However, there is accumulating evidence to suggest that the identification of these clinical signs can be influenced by the diagnosis being evaluated. The studies in this thesis contribute to this body of research by investigating the underlying processes by which the diagnosis being considered influences feature identification. Participants in these experiments were asked to identify the clinical signs from photographs of patients or electrocardiogram strips after having been biased towards the correct or an alternate diagnosis. It was found that the availability of a diagnosis served both to change the probability of reporting relevant clinical signs as well as to influence the identification of ambiguous signs. Manipulating the credibility of the suggested diagnosis, subsequently suggesting a second diagnosis, or decreasing the size of the pool of alternatives available to the diagnostician had a large impact on diagnostic conclusions, but produced relatively small effects on the features reported. These results suggest that changing the degree of focus that a clinician places on the suggested diagnosis has a small effect on the identification of the features by comparison to the substantial

effect of merely suggesting a diagnosis. Furthermore, it was found that the subsequent suggestion of a competing diagnosis did not lead to a reinterpretation of the data. This indicates that once clinicians have seen the evidence one way, they are unlikely to see and label it differently. The implication of these findings for research on medical decision making, the mental organization of medical categories, as well as medical education are discussed.

Acknowledgements

I would like to express my gratitude to Dr. Lee R. Brooks, my primary surpervisor. He has taught me to be a thouthful, rigorous and thorough researcher. I would also like to thank Dr. Geoffrey R. Norman for having supported me both professionally and as a friend. He also provided me with the necessary experience in the field of medical education, my chosen career area. I would also like to thank Dr. Bruce Milliken, for having served on my supervisory committee and providing me with excellent constructive advice over the years.

I would also like to acknowledge my colleagues and my fellow graduate students from the Brooks laboratory. Dr. John Cunnington, Dr. Rose Hatala, Dr. Alan Neville, and Dr. Steve Puchalski were a wonderful addition to the group. Their medical expertise and interest in the area of cognitive psychology made for interesting and challenging discussions regarding medical education. Dr Tim Wood, Kevin Eva, Chan Kulatunga-Moruzi, and Sam Hannah were some of the best lab mates a person could ask for. These four individuals are intelligent, interested and thoughtful. They always served as great people with whom to discuss research ideas...and attending conferences with them was always an adventure.

On a more personal note, I would like to thank the many great friends that I met at McMaster University : Nicole Anderson, Amy Branscombe, Nicole

Table of Contents

List of Tables

List of Appendices

Chapter 1: Introduction

When seeing a patient for the first time, clinicians are faced with a wealth of information that can help them in determining the diagnosis and treatment for the patient. How clinicians attend to this information, combine it, and match it with a mental representation of diagnostic categories is of great interest to medical researchers and educators. The goal in studying these processes is to understand the components involved in the development of expertise and to identify major sources of diagnostic errors in order to design interventions and curricula to prevent these errors. As will be discussed in Chapter 2, researchers in both the fields of medical decision making and medical categorization tend to underemphasize the ambiguity present in the clinical signs of patients. Researchers in medical decision making focus on determining the heuristics and biases underlying neglect of base-rates of disorders and improper weighting of the evidence acquired during a case. Their aim is to design computer-aided diagnostic tools to assist clinicians in reducing errors resulting from these heuristics (Elstein, 1999; Friedman et al., 1999; Tversky & Kahneman, 1974). Researchers in medical categorization place emphasis on describing the way in which clinicians match the features of a case with the mental representations of diagnoses to determine diagnostic categorization (Bordage & Lemieux, 1991; Bordage & Zacks, 1984; Regehr & Norman, 1996). Few researchers in medical decision making or categorization place much emphasis on the identification of clinical signs from patients (for a notable exception, see Bordage, 1999). This lack of emphasis leads to the implicit assumptions that the identification of features is not a major source of errors in diagnosing a patient and that this identification occurs independent of the diagnosis being considered.

1

Evidence from cognitive psychology and medical diagnosis suggests that visual features are often ambiguous and that their identification is influenced by the hypothesis being tested (Berbaum et al., 1986, 1988; Norman, LeBlanc, & Brooks, 2000). The studies in the present thesis are aimed at investigating the processes underlying the influence of a diagnostic hypothesis on feature identification. Naive participants, novice medical students, and more experienced clinicians were asked to identify the clinical features from head-and-shoulder photographs or from electrocardiographs of actual patients after either the correct or an alternate incorrect diagnosis had been suggested to them. Two experiments demonstrate the influence of a diagnosis on feature identification for medical students and residents in family medicine. In measuring both the correct features and the features of the alternate diagnosis presented, the aim was to establish whether the suggested diagnosis simply served to guide the attention of clinicians to relevant features or whether it also influenced the identification of those features. Various techniques were used to manipulate the degree of focus placed on a diagnostic suggestion by the participants to establish whether the influence of the suggested diagnosis was a function of just its availability or also of the degree of focus placed on it. In a second experiment, changing the degree of focus placed on the suggested diagnosis was achieved by changing the case history accompanying a suggested diagnosis. In a third experiment, this was achieved by presenting medical students with a subsequent competing diagnosis. Finally, in a fourth experiment, the number of alternatives diagnoses was manipulated.

Chapter 2: Literature Review

The literature of how clinicians reach a diagnostic decision is divided into two areas; medical decision making research that aims to identify and reduce errors caused by reliance on heuristics when selecting and combining the information available from the patient, and categorization research that aims to describe differences in the mental organization of medical categories between expert and novice clinicians, or between competent and incompetent clinicians.

Medical Decision Making

When faced with uncertain and complex information from a patient, clinicians develop heuristics to compensate for the inherent limitations in the amount of information that is available and that can be processed and stored in memory (Elstein, 1999; Wallsten, 1981). Although these heuristics sometimes lead to inaccurate conclusions, they are used because they lead to adequate decisions most of the time. Proponents of decision theory attempt to determine the optimal processes for making a decision (normative view), and sometimes to describe how individuals actually make decisions (descriptive view). A main concern in this field is to understand why the normative and descriptive accounts of decision making do not always match. That is, the concern is to understand why people do not behave optimally from a normative point of view when making decisions. The principle espoused as the normative procedure of decision making is Bayes's theorem, which states that given an imperfect diagnostic test, the post-test probability of a diagnosis is a function of its pretest probability and the strength of the evidence (Elstein, 1999; Fischoff & Beyth-Marom, 1983). The pretest, or *a priori*, probability is the known

3

prevalence of the disease (its base-rate), while the strength of the evidence is measured as the ratio of the probability of observing a clinical sign in patients with the disease versus the probability of observing the clinical sign in patients without the disease (Elstein, 1999). Errors in diagnostic conclusions have been found to result both from errors in estimating the prior probability and from inaccurate assessments of the strength of the evidence.

Clinicians generate inaccurate assessments of the prior probability because they rely on heuristics of availability and representativeness in making these assessments. When relying on an availability heuristic to estimate the prevalence of a disease, clinicians rely on the ease with which particular instances of that disease come to mind (Detmer, Fryback & Gassner, 1978; Tversky & Kahneman, 1974). As frequent events will come to mind more readily than infrequent ones, this heuristic will typically lead to adequate estimates. However, the ease with which instances come to mind can be determined by factors other than frequency, such as salience. Unusual cases tend to be more salient and thus more memorable than ordinary cases. As people are more prone to overestimate the prevalence of easily recalled events and to underestimate the prevalence of ordinary events or events that are hard to recall, the prevalence of rare events tends to be overestimated (Elstein, 1999; Tversky & Kahneman, 1974). An example of the availability heuristic leading to inaccurate estimates of base-rates was shown by Detmer et al. (1978), who asked surgeons from various specialties to estimate the mortality rates in an entire surgical service. Detmer and collaborators (1978) observed that surgeons working in specialties with relatively high mortality rates (cardiovascular, neurosurgery, general surgery) overestimated the overall mortality rate of the service much more than surgeons in specialties with low mortality rates (plastics, orthopedics, urology). The actual mortality rate was 1.44 % of all cases, and the surgeons from the high mortality

specialties estimated the rate to be 4.79% while those from the low mortality specialties estimated it to be 2.1%. In estimating the mortality rate, clinicians based their judgments on the ease with which they could recall fatal events. As the surgeons in the high mortality specialties are likely to recall more instances than will those in specialties with low mortality rates, they overestimated the overall rate for the entire surgical service.

Judgments will also be influenced by the degree to which a case is representative of its category. In this case, clinicians will compare a particular case to several disease categories and will select the one with the best fit. They will base their judgments on how well the case is representative of the disease, on its apparent similarity to the category. In doing so, they neglect the base rate of the alternatives being considered, as all the alternatives are judged to be equally likely (Elstein, 1999; Heller, Saltzstein, & Caspe, 1992; Tversky & Kahneman, 1974).

Errors in medical decision making can also result from heuristics and biases in the assessment of the strength of the evidence. One example of this is the occurrence of conservatism, where clinicians fail to revise their diagnostic probabilities when presented with new information as much as would be suggested by Bayes' theorem. Related to conservatism is the heuristic of seeking information that confirms a diagnostic hypothesis rather than information that would facilitate the testing of competing hypotheses. This phenomenon is often called pseudodiagnostic testing or confirmation bias (Elstein, 1999) and results in clinicians ordering additional tests that, although they will bolster their confidence in the focal diagnosis, will not alter the Bayesian probability of the diagnosis.

A third demonstration of non-normative assessment of the evidence is the phenomenon of anchoring and adjustment (Elstein, 1999). When estimating the probability of occurrence of an event, individuals start from an initial estimate point and adjust it according to new evidence to reach a final answer. In most cases, the adjustment

is typically insufficient and the final estimate is strongly influenced by the starting value (Elstein, 1999; Tversky & Kahneman, 1974). In a classic demonstration of the phenomenon, groups of subjects were asked to estimate the percentage of African countries in the United Nations. A starting number between 0 and 100 was determined by spinning a wheel in front of the participants. The participants were then asked to say whether they thought the number was higher or lower than the true value and to estimate the true value by adjusting the number up or down from the initial value. The groups of subjects who received the initial value of 10 gave a median estimate of the percentage of African countries in the United Nations of 25%, while the group who received the initial value of 65 gave a median estimate of 45% (Tversky & Kahneman, 1974). Thus, different starting points resulted in different estimates.

As a result of the identification of such heuristics and biases, there is an emphasis on educating clinicians about these biases in the hope that they can learn to avoid them (Elstein, 1999). Another strategy to deal with these biases has been the implementation of computer-assisted diagnostic tools based on algorithms designed to be less vulnerable to biases in estimation of base rates and evaluation of the strength of evidence (Regehr & Norman, 1996). Such medical decision support systems can serve as tools for the management of information, for reminding clinicians of diagnoses or problems that might otherwise be overlooked, and for providing customized assessments or advice based on patient-specific data (Shortliffe, 1987). Typically these systems accept lists of clinical manifestations and then propose diagnostic hypotheses in order of likelihood, indicate additional information that may narrow the range of diagnostic possibilities, and suggest alternative work-up strategies (Barnett, Cimino, Hupp & Hoffer, 1987; Berner et al., 1994; Friedman et al., 1999; Shortliffe, 1987). The goal of such systems is to reduce the neglect of base-rates and insufficient revision of belief in the face of new evidence.

Friedman et al. (1999) reported that consultation with Quick Medical Reference (QMR) or ILIAD led to a 6% increase in the number of cases in which the correct diagnosis was generated, with students benefiting more from the decision support systems (increase in presence of correct diagnosis in 9% of the cases) than faculty.

A major problem with these diagnostic tools is that they assume that the data necessary for the generation of the appropriate conclusions are available to the clinician independently of the diagnosis (Regehr & Norman, 1996). As will be shown in this thesis, the diagnostic hypothesis(es) considered by clinicians influences the identification of the clinical features. Thus, the utility of diagnostic support systems appears to be limited by the information provided by the clinician.

Organization of knowledge of clinicians

Initially, research in medical education focused on the identification of the skills characterizing expert clinicians in the hope of designing medical curricula to teach these skills. Both novices and experts were shown to generate several hypotheses early in a clinical encounter and then seek information that will either confirm or refute each hypothesis. The consistent finding in these studies was that the accuracy of the hypotheses generated was related to expertise, which led researchers to suggest that a major component of expertise was the organization of knowledge (Bordage & Lemieux, 1991, Neufeld, Norman, Feightner & Barrows, 1981; Norman, Trott, Brooks & Smith, 1994).

An early attempt to elucidate the organization of knowledge in general, called the classical view, stated that people have what resembles dictionary definitions of concepts, in which a logical combination of features defines the categories (Bordage & Zacks, 1984; Medin, 1989; Regehr & Norman, 1996). According to this view, there is a list of

features for each category that are individually necessary for inclusion to the category and collectively sufficient to determine category membership (Komatsu, 1992; Medin, 1989; Murphy & Spalding, 1995). Thus, membership in a category is clear-cut (Komatsu, 1992). If an item has all of the defining features, it is a member of the category. If it does not have all of the defining features, it is not a member of the category. In addition, membership is discrete. All items in a category are equally good members because they all possess the defining features of that category. In medicine, this would imply that a patient can be diagnosed with a disease on the basis of a combination of symptoms, *all* of which would have to be present for the diagnosis to be applied (Custers et al., 1996).

An important challenge to the classical view of categorization was the finding that participants rarely generated features shared by *all* members of the category when asked to list the features of various items from the same category (Rosch & Mervis, 1975). This difficulty in generating lists of necessary and sufficient features of a category called into question the notion of defining features of categories. A second important challenge to the classical view came from the observation of typicality effects, in which items were judged to be better or worse examples of a category. When asked to rate items in terms of their goodness of membership to a category, participants in Rosch and Mervis' (1975) study rated as better members of the category those items that shared the most features with other members of the category. This finding is inconsistent with the idea that concepts are discrete. If all members of a category contain all of the defining features, no item should be judged as a better member than others (Komatsu, 1992). A third problem with the classical view was the presence of unclear examples, that is, items for which membership was difficult to determine such as a rug being a piece of furniture. Such unclear examples should not exist according the classical view, as people should be able to use the list of defining features to make decisions (Medin, 1989).

To account for the findings that challenged the classical view of categorization, Rosch and her colleagues proposed a second theory of categorization, one based on family resemblance (Rosch, 1978; Rosch & Mervis, 1975; for reviews see Komatsu, 1992; Regehr & Norman, 1996). According to this view, each item in a category has at least one attribute in common with at least one other member of the category, but there is no single attribute common to all items in a category. While members of a category share features, there is no single feature or combination of features that is necessary for category inclusion. Based on experiences with examples of a category, people will abstract the central tendencies (or prototype) that will become the summary representation of the category (Medin, 1989). Decisions about whether or not an item belongs to a category will be determined based on its resemblance to the prototype. The family resemblance view is capable of explaining the typicality effect, that is, that some items are viewed as being better or worse members of the category. Those items that contain more features of the category will be judged to be more typical of the category. Thus, categories are ill-defined and organized around a set of properties that are characteristic of the category rather than defining (Komatsu, 1992; Medin, 1989; Rosch, 1978; Rosch & Mervis, 1975).

Bordage and Zacks (1984) advanced a theory of medical knowledge being organized around prototypes. They hypothesized that if medical knowledge is organized around prototypes, then all category members should not be seen as equal members of the category. Category members that share many features with the prototype should be rated most typical of the disease. To test this prediction, Bordage and Zacks (1984) asked subjects to generate disorders of different medical categories, such as organ-system related disorders. They found that the disorders listed earliest were the ones rated as being more typical of the category. These diseases were also the ones that had the most features

in common with the other disorders in the category, that is, they had the highest family resemblance. Furthermore, when asked to verify category membership statements, clinicians showed both faster reaction times and fewer errors for the more typical examples of the category than for the less typical ones. This finding provides further support for the prototype view over the classical view, as the former predicts that typical instances of a category should be more easily accessible, while the latter holds that all instances of a category have full and equal membership and should therefore be equally accessible. Additional evidence that medical categories are represented by abstracted prototypes comes from the finding that patients have a preference for mentioning symptoms in order of typicality (most typical to least typical), as well as the finding that patients who present with a typical form of a disease will be diagnosed faster, more accurately, and with more confidence than patients who present with a less typical form of the disease (Custers et al., 1996).

In an extension to the prototype theory, Lemieux and Bordage (1992) argue that medical knowledge is organized along underlying semantic structures which allow clinicians to abstract and organize the information contained in factual evidence. The information is organized along abstract semantic axes that ascribe meaning in terms of binary opposition (e.g. acute vs. gradual, large vs. small) and to organize the diagnoses according to each opposing term (diseases with an acute onset of symptoms vs. those with gradual onset of symptoms). The semantic axes allow clinicians to compare symptoms and signs, and to contrast diagnoses according to precise perspectives. In support of their theory of semantic axes, Bordage and Lemieux (1992) report that the clinicians who correctly solve difficult clinical cases are those who evoke the greatest number of distinct semantic axes and who organize the clinical signs into a pertinent system of abstract semantic properties. The ones with diagnostic difficulties have a more

empirical view of the case and rely more on the material facts of the case. Chang, Bordage and Connell (1998) reported similar results, showing that clinicians who made the correct diagnosis after having read a chief complaint used three times as many semantic attributes than did those clinicians who made an incorrect diagnosis. Bordage and his colleagues (Bordage & Lemieux, 1992; Chang et al., 1998) argue that such findings support the notion that deeply abstracted representations play an important role in the understanding of medical problems and in the generation and elimination of diagnostic hypotheses.

The family resemblance or prototype theory of categorization has been applied extensively to medical education. Medical textbooks provide verbal descriptions of the features typically associated with a given diagnosis. When case examples are provided, they consist of the prototypical or "average" representation (the classic case) and little emphasis is placed on the variability around that prototypical case (Custers et al., 1996; Regehr & Norman, 1999). Rather, the emphasis is placed on the typical characteristics of diseases, and consequently, on students acquiring prototypical representations of diseases (Heller et al., 1992). An important assumption of this approach for medical education is that if students gain a clear representation of a prototypical case, it will then be easier for them to classify less ideal cases which differ due to normal variation (Bordage & Zacks, 1984; Custers et al., 1996; Mervis & Pani, 1980).

One problem with the prototype theory is that it treats categories as if they are context independent. It implies that the only abstracted information is the central tendency and that information regarding category size, variability of examples, and the correlation of features is discarded (Custers et al., 1996; Medin, 1989). For these reasons, researchers have questioned whether prototypes are actually used to make diagnostic categorizations on a regular basis. Regehr and Norman (1996) argued that, although they

may be useful for tasks requiring a systematic search of the clinical findings when confirming or refuting a hypothesis, prototypes may be less useful in the early phases of diagnosis, in which the generation of hypotheses occurs very rapidly and with limited information. Thus, they propose that a third theory of categorization, the instance-based theory, is a better description of the organization of knowledge during early hypothesis generation. This instance-based theory differs from the prototype theory in that a category is represented as a collection of instances of that category, with the integrity of each instance being preserved (Brooks, 1990; Medin, 1988). Instead of being compared to an abstracted prototype, a new instance is compared to the most similar instance in memory to determine its membership in a category (Komatsu, 1992; Nosofsky, Clark, & Shin, 1989; Regehr & Norman, 1996). A major difference between the family-resemblance and the instance-based theories is that while the family resemblance view requires that every attribute in the abstracted representation must characterize more than one instance, the instance-based theory holds that an attribute of a category does not need to be present for more than one instance of that category (Komatsu, 1992). In such an organization, information about the various members is maintained, thus allowing variability among instances of a concept.

To demonstrate that categories are composed of representations of specific instances, Allen, Norman and Brooks (1992) examined the impact of practice examples on the accuracy of diagnosing subsequent dermatological cases. They taught the rules of 6 dermatological diagnoses to a group of naive medical students, followed by a review of the diagnosis descriptions. The students then underwent a biasing practice session in which they saw one of two groups of 24 photographs of dermatological lesions. During the test session, all the students were asked to diagnose five items from each of six diagnostic categories and to describe the features in the item that led to their diagnosis.

Allen et al. (1992) observed that the practice "biasing" slides had an impact on performance. The students who had seen a practice item that was both similar to one of the test items and an example of the same disease correctly diagnosed the slide 89% of the time. Students who had seen a visually similar item that was an example of a different diagnostic category than the test slide correctly diagnosed the test slide only 42% of the time. These results show that the characteristics of specific learned examples can have an impact on diagnosis independent of the learned rule. Although both groups had learned the same diagnostic rules, the effect of these rules was modified on the basis of similarity to a previous example. Allen et al. (1992) argued, on the basis of these results, that the rules for diagnosis are insufficient to accurately learn the variations contained in clinical cases, and that prior instances are stored and used to categorize new cases.

However, there is also evidence against an instance-based theory. In one study, subjects who took part in concept formation experiments demonstrated knowledge about the central tendencies of a concept even if they had not been presented with specific instances matching those central tendencies. Posner and Keele (1968) had subjects learn a set of patterns by being shown distortions of a prototype. When asked to classify new examples, they were more accurate at classifying the prototype than new examples of the same pattern. Such results suggest that an abstract representation of the category is constructed around the average of the features to create a prototype. It is the prototype rather than the specific instances that is stored in memory. Also, some theorists argued that if new items are categorized on the basis of their similarity to prior instances, then the classification of new items should depend on the ability to recognize items presented previously. However, this relationship between recognition and classification is sometimes absent, as shown by Hayes-Roth and Hayes-Roth (1977) in which subjects were more confident about their classification of prototypes than of previously presented

old instances. However, they were more confident of having seen some of the old instances than the prototypes. Thus, the first finding suggests storage of central tendency information while the latter suggests storage of instance information. Because of the lack of correlation between recognition and classification, some researchers argued that concepts include both abstracted (prototypical) and particular (instance-based) information (Komatsu, 1992).

A fourth class of theories of categorization that could account for both central tendencies and instance information are semantic-networks or schema-based theories. In these theories, both function and structure are closely related and the emphasis is placed on explaining how knowledge is used to solve problems (Custers et al., 1996; Komatsu, 1992). Knowledge is represented in small units of meaning, called nodes, that are connected by links to form a semantic network. The nodes in the network will become active in response to some form of stimulation and this activation will spread to other nodes of the network according to the links. Thus, nodes can become active as a response to stimulation from the environment or from the spreading activation of other active nodes in the network (Custers et al., 1996). As the result of repeated stimulation of two or more linked basic nodes, new complex nodes can be formed in a process called chunking. When this occurs on a larger scale, the chunks encompass parts of a semantic network and form scripts or schemata representing generalized events (Custers et al., 1996).

Schmidt and Boshuizen (1992, 1993) used the phenomenon of the "intermediate effect" as support for the notion that medical knowledge is organized in scripts. They asked medical students and expert clinicians to read a text describing a patient's history. When asked to recall the clinical case information, the participants who were at an intermediate level of expertise (advanced medical students) recalled more clinical information than the participants with more and less expertise. Thus, the intermediate

effect consists of an inverted U-shaped relation between expertise and recall of clinical information. To explain the intermediate effect, Schmidt and Bozhuizen (1992, 1993) argued that students and clinicians apply different kinds of knowledge to the task. The novices' medical knowledge consists of a limited understanding of biomedical knowledge, while that of the intermediates consists of more extensive biomedical knowledge with little or no knowledge of the manifestations of a disease in actual patients. In experts, this biomedical knowledge plays a minor role in their reasoning. Rather, with repeated practice and exposure to patients, biomedical concepts have become clustered together in chunks that allow the expert clinicians to have shortcuts in their reasoning and skip the intermediary concepts. Their knowledge thus becomes "encapsulated" into clinical and diagnostic labels. This encapsulated knowledge contains information about the clinical manifestation of the disease, its consequences, and the context under which it develops. The intermediate effect is thus explained by the fact that the intermediate clinicians report many more biomedical concepts than experts, as the latter group's recall consists of encapsulated interpretations of the clinical information.

An advantage of semantic networks and of the schema-based theories is that they can explain the phenomena of family resemblance while also maintaining the particular information of instances (Komatsu, 1992). Because the schema view provides a consistent representation of both the abstracted information and the information about instances, it allows for flexibility in emphasizing either the particularized information or the abstracted information, depending on the context of use or the acquisition of the concepts (Komatsu, 1992).

Another advantage of semantic-network theories is that they are widely applicable, more so than the prototype or instance-based theories, as they address not only medical diagnosis, but are also capable of linking medical knowledge to medical

reasoning at various levels of expertise (Custers et al., 1996). They also allow for medical reasoning to occur even when no prototype or relevant instance can be generated. By definition, a clinician should misdiagnose the first case of any disease seen according to the prototype and instance based theories (Custers et al., 1996), however semantic networks can accommodate this situation. There are, however, serious limitations to semantic networks. The knowledge structures of experts show no overlap or consistency, which would be expected of experts in a domain who share a common stable representation of knowledge (McGaghie, Boerger, McCrimmon, & Ravitch, 1994).

An important limitation shared by these theories is that they fail to address the issue of possible interactions between diagnostic hypotheses and feature identification. As will be discussed in this thesis, the diagnosis being considered by a clinician has an important impact on the identification of clinical signs of patients.

Interaction of diagnosis and interpretation of the clinical signs

Studies of the organization of knowledge and reasoning in medical diagnosis, although having different emphasis, share a common limitation. They all have the underlying assumption that errors in reaching a diagnosis will be solely due to the combination of various sources of information or in the failure to correctly match a new case with the proper diagnostic category. Rarely do they acknowledge that the identification of features can be problematic and influenced by the diagnosis being considered. However, there is a growing body of evidence in psychology and medical education to suggest that the interpretation of these clinical signs is neither self-evident nor independent of the diagnosis being considered by a clinician.

Some of the more dramatic demonstrations of contextual influences on the perception and interpretation of the physical characteristics of stimuli are found in visual

illusions presented in basic psychology textbooks. One of the best known optical illusions is a drawing that, depending on the perspective taken by the viewer, depicts either a young lady or an older woman. The same characteristic, a dark line, is interpreted either as the necklace worn by the young lady or as the mouth of the old woman. The same characteristic is interpreted differently depending on the identity of the whole. Other demonstrations of the role of context in the interpretation of characteristics of stimuli can be found in Mooney's figures, where the interpretation of the features can be difficult unless the viewer knows the identity of the whole.

Empirically, this phenomenon has been studied in the Word Superiority Effect, where letters are more easily detected in words and orthographically regular pseudowords than when presented alone or in unrelated letter strings (Reicher, 1969; Whittlesea & Brooks, 1988). Following short presentation times, more letters will be recognized when presented in a word than when presented in an unrelated letter string. Similarly, letter strings constructed by deleting letters of a word or by replacing some of the letters of the word will often be perceived as the original word (Rumelhart, 1985). Thus, word-level perceptions affect letter perception, indicating that the letter information is preserved and that word level information provides constraints on the partial letter information (Rumelhart, 1985).

To account for the finding that people's comprehension of information at one level (letter) can be dependent on the comprehension of information at a higher level requires a model of information processing in which all sources of knowledge are applied simultaneously, and in which perception is a product of interactions among these levels (Rumelhart, 1985). A well-known model of this interactive process is found in McClelland and Rumelhart's (1981) interactive activation model of perception. The basic assumption of this model is that, in addition to receiving sensory information, individuals

bring in knowledge about the general properties of objects whenever they perceive objects. In this model, there is a visual feature level, a letter level, a word level, as well as higher levels of processing. These higher levels provide top-down information to the word level regarding what is expected. Therefore, the incoming "bottom-up" feature information interacts with the "top-down" information (what we know about words) to determine what is seen.

While these illusions and studies show that context plays a role in feature perception, the stimuli used were purposefully designed to be ambiguous or perceived under degraded conditions. There is nothing in these studies to guarantee that such context-dependent feature interpretation would also be observed with seemingly unambiguous materials, and under conditions in which the participants are allowed to inspect the data analytically and without time pressure, such as in the medical diagnostic task. In addition, few researchers in the field of psychology have studied the processes underlying this interaction between context and feature identification.

Even when the stimuli are seemingly unambiguous and the experimental conditions are not degraded, context-dependent feature identification can be observed. For example, studies in radiology have demonstrated that providing prior information, such as the location of tenderness and swelling for the detection of fractures (Berbaum et al., 1988) or the tentative diagnosis for the detection of radiographic abnormalities (Berbaum et al., 1986), leads to an increase in the true positive rate of detection of these fractures and lesions. Similarly, Norman, Brooks, Coblentz and Babcook (1992) observed that the case history of equivocal bronchiolitis cases had an impact on the detection of bronchiolitic features from chest radiographs. They conducted a cross-over study of history and radiographs, such that a normal history was not always matched with a normal radiograph and vice-versa. Participants were asked to interpret chest radiographs

that were either unambiguously normal or abnormal (obvious cases of bronchiolitis) or equivocally normal or abnormal. They found that clinical histories affected the ratings of features present in the equivocal radiographs.

These studies demonstrate that the interpretation of clinical signs can strongly influenced by the diagnosis being considered, even in medicine where the stimuli are not presented under degraded conditions. However, one could argue that radiologists are conditioned to function in a manner that would be strongly influenced by tentative diagnoses. Typically, they see patients after much of the clinical data has been accumulated and a tentative diagnosis has been made. Anecdotally, a number of radiologists tend to read the clinical information provided on the x-ray order-form prior to looking at the radiographs. Because of the wealth of information that can be gathered prior to their seeing the patients, these radiologists may have experienced few cases where the results of their analyses were significantly incongruent with the tentative diagnosis provided by the clinical information. In addition, it could be argued that radiographs are as ambiguous as stimuli used in psychological experiments. For these reasons, the tentative diagnoses presented in the radiology studies may have been fairly persuasive for the clinician. It would thus be of interest to study how clinicians reason on the first encounter with a patient, before any information has been gathered and in circumstances where the presenting features are fairly straightforward. Are the radiologists behaving in the observed manner because of a predisposition brought about by their field or because that is the usual way of reasoning of all clinicians?

One study that addresses this question was conducted by Brooks, LeBlanc and Norman (2000). Their study, with both medical students and academic general internists, was aimed at addressing the role of both verbal and visual information in the diagnostic process and at investigating whether or not the way features were described could affect

what was detected in patients. Participants were presented 15 cases in three passes, with each pass providing them with additional information (either visual or verbal). On the first pass, the participants read a short case history and were asked to list the diagnostic possibilities in order of likelihood. They did so for each of the 15 cases. In the second pass through the cases, the participants were shown head-and-shoulder pictures of each patient in addition to the case history. They were asked to list the diagnostic possibilities and to list any clinical features detected in the picture. On the last pass, the participants were given the case history and the picture along with an interpretative description of the clinical features present in the picture (e.g. ptosis). Again, they were asked to give their diagnoses. The results of the study showed a near linear increase in diagnostic accuracy across the three passes, for both students and experts. Although it was expected that both experts and novices would show an increase in accuracy upon seeing a picture of a patient, the fact that both groups showed an increase in diagnostic accuracy when the interpreted features were presented to them was of great interest. This result suggests that both groups, novices and experts, gained information when the critical features were made interpreted, implying that features are not self-evident. The participants may have seen the physical characteristics, but they, at the least, failed to recognize their clinical importance and to correctly identify them. The way the features were described was critical in invoking the disease.

The results from this last study suggest that considering a diagnosis will influence the identification of features in areas of medicine where the clinical features are believed to be less ambiguous than in radiology. However, the context provided by Brooks et al. (2000) was at the level of feature description. Although their results indicate that features are not self-evident and that their description can be critical to invoking a diagnosis, they do not address the issue of whether the identification of the features will be influenced by

the diagnosis being evaluated. Brooks et al. (2000) tested this assumption with their second experiment, by investigating whether students who were asked to evaluate a correct diagnosis would show greater accuracy in feature interpretation than those asked to avoid any diagnostic hypotheses. In addition, they investigated whether the number of diagnoses being considered, and thus the salience of the correct hypothesis, would have an effect on the students' identification of features. They observed that when the students had the correct diagnosis in mind, they reported 20% more of the depicted correct features than did students who were asked to avoid reaching any diagnostic conclusion. This effect was moderated by the number of diagnoses being considered. Students asked to evaluate five diagnoses (including the correct one) were in the middle of the two contrasting groups, reporting 10% fewer depicted correct features than those who were asked to evaluate only the correct diagnosis.

The results of this study also provided some clues as to the nature of the bias imposed by the context. The effect was not simply due to pushing the subjects towards reporting anything that would be consistent with the diagnosis regardless of the information provided by the picture. Although those subjects who did entertain the correct diagnosis reported more depicted correct features, they did so discriminantly, in that they did not report more of the potential features (consistent with the diagnosis but not visible in the picture). Brooks et al. (2000) also investigated the nature of the influence by asking students to rate, on a 5-point Likert scale, the likelihood that specific features were present (definitely not present to definitely present). The students who had been instructed to consider five diagnoses and those who had been instructed to avoid any diagnostic conclusion rated the depicted features as likely to be present as did the students who were asked to evaluate only the correct diagnosis. If the students who had considered only the correct diagnosis had rated the features as being more likely to be present than

the students considering five diagnoses or no diagnosis, it could then be argued that the diagnosis changed the way a feature was seen. However, these results suggest that when subjects are considering a diagnosis, it guides their search towards the relevant features in a manner similar to asking them directly about the presence of specific features.

This beneficial effect of having the diagnosis in mind during feature interpretation was also found by Norman, Brooks, Colle and Hatala (1998), who manipulated the instructional conditions when participants had to diagnose electrocardiograms (ECGs). After being taught to read ECGs, psychology undergraduate students read a series of test ECGs and were asked to list all of the important features, as well as to reach a diagnostic decision. When participants were encouraged to think of the diagnosis prior to searching for features, both diagnostic accuracy and performance of feature listing was improved compared to when they were asked to perform the feature search prior to generating any diagnostic hypotheses. These results held even when the subjects who were asked to do the feature search first had the ECGs visible to them when generating diagnostic hypotheses. When asked to perform the feature perception task first, the students listed additional irrelevant features that needed to be explained when trying to come up with the correct diagnosis. Thus, in cases where ambiguity might exist in the features, the focus that is provided by top-down processing could improve performance.

The results of these studies challenge the assumption in medicine that the identification of features is not problematic and that it occurs independently of the diagnosis considered. Both in the fields of psychology and medical education, relatively little is known about the processes underlying the interaction of the context and feature identification. The studies in this thesis are aimed at investigating whether this interaction can lead to errors, whether the initial interpretation is flexible, and whether the influence of the diagnosis on feature identification is the result of its availability or of the degree of

focus placed on the diagnosis. As this influence appears to be substantial and prevalent in various areas of medicine, it is crucial to understand under which situations it occurs, and to identify when it is beneficial or detrimental to clinicians. If, as argued by Neufeld et al. (1981) and Norman et al. (1994), clinicians generate diagnostic hypotheses early in the case and then proceed to gather information that either confirms or refutes the hypotheses, the influence of a diagnostic hypothesis could have a serious impact on the subsequent evaluation of the data. As the studies of this thesis were designed to investigate different underlying processes of the interaction between diagnosis and feature identification, they will be presented in separate chapters. The studies in the third chapter are aimed at investigating the nature of the influence of the diagnosis on feature identification. The studies in the fourth chapter were aimed at investigating the impact of manipulating the credibility of the suggested diagnosis, while the fifth chapter addresses the issue of the impact of the number of alternatives considered by the clinician. In the sixth chapter, the implication of the current studies will be discussed.

.

Chapter 3: The influence of a tentative diagnosis on identification of features from patients' appearance

Feature identification can be problematic and does not occur largely independent of the diagnosis being evaluated. Clinicians are more likely to correctly see and label a feature when they are thinking of the correct diagnosis than when thinking of an incorrect diagnosis or even avoiding thinking about any diagnosis at all. One limitation with the studies demonstrating the influence of a suggested diagnosis on feature identification is that they all do so by measuring the participants' accuracy in identifying the correct features. The question remains whether the increase in the accuracy of feature interpretation in the context of the correct diagnosis is the result of a more effective search strategy or the result of a bias in the identification of the features. The studies in the present chapter are designed to address the question of the nature of the influence of the diagnosis on feature identification. In the first experiment, this is done with medical students. As will be discussed in the introduction of Experiment 1B, the same study is subsequently run with medical residents to investigate whether clinical experience impacted on the influence on a suggested diagnosis on feature identification.

Experiment 1A. Initial Demonstration: Identifying the nature of the influence of the diagnostic context on medical students' identification of features

Once a diagnosis is being evoked, it might serve to activate a representation of the disease presentation, bringing to mind the possible features that can be present in a patient suffering from the condition. The clinicians can then run through the list of features, checking for the presence or absence of each. The diagnosis would thus serve as a focus of attention, determining which features to look for and where to look for them. In

addition, the diagnosis considered might have a stronger impact by inducing a bias in the identification of the observed clinical characteristics. Knowing that a moon-shaped face is a feature of Cushing's Disease and considering Cushing's as a possible diagnosis might lead the clinician to interpret a slightly obese face as moon-shaped. The interaction between diagnosis and feature identification, if present, might be sufficient to lead clinicians to identify features that are not present in the stimuli, according to experts. Studies in which the accuracy in detecting features is the sole dependent outcome measure do not permit conclusions to be drawn regarding the specific influence of the diagnosis on feature identification.

One way to detect the presence of a bias in feature identification is to measure whether clinicians who are biased towards an incorrect diagnosis identify features that are suggested to them by this diagnosis but not present in the stimuli. There is documented evidence from the reading literature that the context can lead people to incorrectly identify features of the stimuli. In a study by Rumelhart (1985), participants were asked to identify the letters from tachistoscopically presented letter strings. These letter strings consisted of an initial pair of consonants followed by a pair of vowels, and ending in a final pair of consonants. Each vowel and consonant pair was constructed such that one order could occur in the English language (legal cluster) and one order did not occur (illegal cluster). For example, a legal initial consonant cluster was "pr" while its illegal counterpart was "rp". Rumelhart (1985) observed that although subjects rarely transposed the initial consonant cluster, they transposed illegal vowel clusters approximately 25% of the time, and transposed illegal final consonant clusters about 14% of the time. Comparatively, they transposed legal vowel and final consonant clusters only about 3% of the time. Thus, their perception of the letter strings was influenced by the orthographic structure of the English language sufficiently to lead to errors in identifying the stimuli.

Hatala, Norman and Brooks (1998) found similar results using electrocardiograms under non-degraded perceptual conditions. They presented medical students, internal medicine residents, and cardiologists with 10 ECGs. For portions of the trials, the ECGs were accompanied by case histories. When a case history was presented, it was suggestive of either the correct diagnosis or a plausible alternative. All three groups showed the highest diagnostic accuracy when the ECGs were accompanied by the history suggestive of the correct diagnosis. When the history accompanying the ECGs was suggestive of the alternative diagnosis, diagnostic performance across the three groups was worse than when they were provided with no history. Therefore, the history had a bi-directional effect on diagnostic accuracy, showing that the biases observed in previous studies can also lead to errors. This bi-directional effect of the history was also observed for feature perception. When presented with a history consistent with the correct diagnosis, the clinicians listed more features consistent with the correct diagnosis than did the clinicians in the two other groups. As for the clinicians who received a history consistent with the alternative diagnosis, they listed more features consistent with the alternative diagnosis than did the participants in the two other conditions.

Thus, even in tasks where the stimuli are not presented under degraded conditions, the suggestion of a diagnosis can lead to the misidentification of features. However, the generalizability of the results from Hatala et al.'s (1998) study are limited, as ECGs are typically interpreted late in the process of diagnosing a patient, once much other clinical information has been gathered. The goal of the present study is to extend Hatala's (1998) findings to another area of medicine in which the diagnostic hypothesis initially generated by clinicians has a potentially important impact on the interpretation of visual information. Students will be presented with head-and-shoulder photographs of patients accompanied by short case histories and tentative diagnoses meant to bias them towards

the correct diagnosis or an alternate plausible diagnosis. They will be instructed to identify all the clinically relevant features in the photographs. If the diagnosis simply serves to guide the search to relevant features, it is hypothesized that the students biased towards the alternate diagnosis will identify fewer correct features, but will not misidentify more features from the alternate diagnosis, when compared to students biased towards the correct diagnosis. However, if the diagnosis also influences the identification of features, the students biased towards the alternate diagnosis will also identify more of its features, that is, features which are not present in the stimuli.

In the Brooks et al. (2000) study, students who were only considering the correct diagnosis did not identify more of the "potential" features, which lead to the possibility that the students in the present study will not identify more features of the alternate diagnosis when it is suggested to them. The potential features from the Brooks et al. (2000) study and the features of the alternate diagnosis in the present proposed study are similar in that in both studies, features are suggested by the diagnosis but not depicted in the photographs. However, the identification of the potential features in Brooks et al.'s (2000) study would have resulted from purely top-down processes as there was no information from the stimuli to support their identification. In contrast, the identification of features of the alternate diagnosis in the present study will result from an interaction between the bottom-up information from the stimuli and the top-down processes. There are some physical characteristics that could be misidentified as features of the alternate diagnosis. The misidentification could result from a similarity between the actual physical characteristics of the photographs and the features of the alternate diagnosis (tanned skin misidentified as jaundice) or from a similarity between the actual features of the correct and alternate diagnoses (parotid swelling misidentified as a moon-shaped face).

Methods

Participants

Twenty second-year medical students from McMaster University's Medical School participated in this study. They were recruited as they began their clerkship training (which begins in the 20th month of training at McMaster), when they move from primarily formal class-based education to ward-based experiences.

Approval for the study design and the use of photographs from medical textbooks was received from the ethics committee of McMaster University's medical program.

Materials

Ten head-and-shoulder photographs were used from a previous study (Brooks et al., 2000). All of the photographs came from textbooks or clinicians' slide libraries and were considered to be classic or prototypical representations of the given diagnosis; they were not marginal examples. Eight of the ten photographs served as the test stimuli and were selected because plausible alternate diagnoses could be constructed for them. The other two photographs were used as practice scenarios; these were scenarios for which the accuracy rate in the Brooks et al. (2000) study was greater than 90%. They were used to create the sense whereby the diagnoses suggested to the students were plausible.

Case histories biasing towards the correct and alternate diagnoses were generated for each of the eight test cases. The correct case histories were the same as those used in the Norman et al. (2000) study. The alternate diagnoses were generated by choosing a "feature" in the picture from which a plausible diagnosis could be constructed. For example, in some scenarios, these "features" were due to normal variations (for example, tanned skin interpreted as jaundice). In other scenarios, the chameleon "feature" was created by the reinterpretation of a cardinal feature of the correct diagnosis (for example, a moon-shaped face would be reinterpreted as facial edema). The biasing 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 diagnosis, a list of its features was generated from an authoritative internal medicine textbook (Fauci et al., 1998). Two experienced general internists were then asked to indicate which features of the correct diagnosis were present in the photograph as well as which features of the alternate diagnosis could be identified if students misidentified either a correct feature or a characteristic caused by normal variation. Only the actual or potential features for which both experts agreed on were recorded. This resulted in the identification of 21 features for the correct diagnosis and 25 features for the alternate diagnosis. The slightly higher number of features for the alternate diagnoses comes from the fact that one of the alternate diagnoses (respiratory failure) had seven potential features, as opposed to the one to four features for the other diagnoses. The correct and alternate features were mutually exclusive in that the identification of a specific feature provided support for either the correct or the alternate diagnosis, but not both. Appendix A contains the correct and alternate diagnoses and case histories for the eight test cases and the features of each diagnosis. Appendix B illustrates two of the photographs used in this study.

Procedure

Each participant saw all ten scenarios, including the two practice scenarios presented first. The eight test scenarios were randomly divided into two groups of 4, labeled A and B. Half of the subjects were biased towards the correct diagnosis for the scenarios in group A, and biased towards the alternate diagnosis for the scenarios in group B. The other half of the subjects were biased towards the correct diagnosis for the scenarios in group B, and biased towards the alternate diagnosis for the scenarios in

group A. The ordering of the eight test scenarios was randomized and kept the same for both groups. The sequence of the test scenarios was the same as presented in Appendix A.

The procedure for each scenario was as follows. On a first page, the subjects were presented with the head-and-shoulders photograph, the short case history, and a tentative diagnosis (correct or alternate). They were asked to look at the picture in order to detect the features that would have led to the tentative diagnosis. On a second page, they were asked to look at the photograph and list all the clinically important features that were present in the photographs. They received further instructions to include any feature that could rule in or rule out a diagnosis, as well as any feature that was abnormal, even if it could not be linked to a specific diagnosis. Once they had listed the features, the students were asked to rate the likelihood of the tentative diagnosis, and any other diagnosis(es) they considered. See Appendix C for an example of the one of the case presentations.

Measures

There were four measures of interest in this study. The first two were measures of diagnostic decisions aimed at measuring the effect of biasing the students towards the desired diagnosis. Two outcome variables were computed: (1) the percent of cases in which they gave the highest likelihood rating to the correct diagnosis, and (2) the percent of cases in which they gave the highest likelihood rating to the alternate diagnosis were calculated. Two other measures included were (3) the percent of the correct and (4) alternate features identified by the students.

Each of the four measures were analyzed using a 2x8 mixed design analyses of variance (ANOVA) with the diagnosis (correct and alternate dx.) as a between subject variable and the scenarios (scenarios 3V, 7V, 9V, 12V, 13V, 16V, 18V, and 27V) as a repeated measures variable. Further analyses of the data consisted of identifying and

quantifying the types of errors that were made in feature identification. The hypothesis was that errors could consist of failing to notice the correct features, mislabelling a characteristic of normal variation, or mislabeling a correct feature.

Results

It was hypothesized that the suggestion of a tentative diagnosis and case history would be sufficient to bias the students either towards a correct or an alternate diagnosis. The students opted for the correct diagnosis in the majority of the cases for which they received the correct case history and tentative diagnosis. Conversely, they opted for the alternate diagnosis in the majority of the cases for which they received the alternate diagnosis and case history (see Table 1). The results of the ANOVA indicate a main effect of diagnosis for both the percent of cases in which the students opted for the correct diagnosis, $F(1, 18) = 226.06$, $\underline{MSE} = .095$, $\underline{p} < .01$, and the percent of cases in which they opted for the alternate diagnosis, $F(1, 18) = 200.94$, $\underline{MSE} = .081$, $\underline{p} < .01$.

There was also some variability in the degree of difficulty of the cases as indicated by a main effect of scenarios for the percent of cases in which they opted for the correct diagnosis, $F(7, 126) = 4.81$, $\underline{MSE} = .065$, $\underline{p} < .01$, and for the percent of cases in which they opted for the alternate diagnosis, $F(7, 126) = 3.84$, $\underline{MSE} = .086$, $\underline{p} < .01$. The scenarios by diagnosis interaction was significant for the percent of cases in which they opted for the alternate diagnosis, $F(7, 126) = 5.73$, $\underline{MSE} = .086$, $\underline{p} < .01$, but not for the percent of cases in which they opted for the correct diagnosis, $F(7, 126) = 1.20$, $\underline{MSE} = .065$, $\underline{p} = .31$. This result indicates some variability in the credibility of the alternate diagnoses generated. For five of the cases, the students went from concluding for the alternate diagnosis in less than 10% of the cases when presented with the correct case history and diagnosis, to more than 75% of the cases when presented with the alternate

case history and diagnosis. However, even when presented with the alternate cases history and diagnosis, the students opted for it in less than 50% of the scenarios, for scenarios 9V, 18V and 27V (Turner's syndrome/cretinism, hypothyroidism/ respiratory failure, and acute glomerulonephritis/mumps).

The analyses of the percent of correct and alternate features identified tested whether biasing students towards a diagnosis also influenced how they interpreted the features from the patient's appearance. As can be observed in Table 2, the students identified more of the correct features when biased towards the correct diagnosis (49%) than when biased towards the alternate diagnosis (35%), $F(1, 18) = 7.64$, MSE = .099, p < .05. This difference translates itself into the students identifying an average of 10.3 features across the eight test scenarios when biased towards the correct diagnosis, and an average of 7.3 features when biased towards the alternate diagnosis.

The analysis of the alternate features identified by the students provided the test of the hypothesis that the influence of a diagnosis was strong enough to change the identification of features. If the diagnostic context simply served to direct attention to relevant features, the suggestion of an alternate diagnosis should not lead students to detect features that are not present according to our expert informants. However, if the diagnostic context served to influence the interpretation of the physical characteristics, then the suggestion of an alternate diagnosis should lead the students to detect features that are not present; features they would not identify when biased towards the correct diagnosis. The students identified more alternate features when the alternate diagnosis and case history were presented (26.5%) than when the correct case history and diagnosis were presented (5.7%) (see Table 2). A main effect of diagnosis indicates that this difference was statistically significant, $F(1, 18) = 60.153$, MSE = .029, p < .01. This translates into the students identifying approximately seven features across the eight

scenarios wen biased towards the alternate diagnosis and approximately 1.5 features across the eight scenarios when biased towards the correct diagnosis. The influence of the diagnosis on identifying alternate features was variable, as indicated by a significant diagnosis by scenario interaction, $F(7, 126) = 6.17$, $\underline{MSE} = .03$, $\underline{p} < .01$. The suggestion of a diagnosis had an impact on the detection of alternate features in all but two scenarios (3V: stomach cancer/liver cancer and 27V: acute glomerulonephritis /mumps).

The identification of an alternate feature could result from either mislabelling a correct feature or by mislabelling a normal variant, because there were no overlapping features (features supporting both the correct and the alternate diagnosis). Exploratory analyses were conducted to quantify the prevalence of each type of error. Of the 25 alternate features, 4 chameleon features were the result of mislabelling the correct feature (Scenario 12V: SLE= malar rash / fever; Scenario 13V: Cushing's=moon shaped face / facial edema, Scenario 16V: mumps= parotid swelling / moon-shaped face; and Scenario 27V: acute glomerulonephritis = edema/ parotid swelling), and the remaining 21 resulted from mislabelling a characteristic associated with normal variation. When the correct diagnosis was suggested, the students correctly identified the four chameleon features 82.5% of the time, misidentified them as alternate features 2.5% of the time, and omitted them 15% of the time. When the alternate diagnosis was suggested, they correctly identified the chameleon features 40% of the time, mislabeled them as alternate features 32.5% of the time, and omitted them 27.5% of the time. As for the 21 features consisting of misidentification of normal variation. the students identified them in 6.3% of the cased when the correct diagnosis was suggested and in 25.4% of the cases when the alternate diagnosis was suggested.

Discussion

The pattern of the results for identification of features indicates that the influence of the suggested diagnosis is bi-directional. First, it served to focus the attention of the students on the relevant features. For example, they were more likely to identify the superclavicular lymph node when considering stomach cancer than when considering liver cancer. In addition, the diagnosis served to change the identification given to the physical characteristics. This latter influence was sufficient to lead to misidentifications of both normal variation and of correct features. For example, in scenario 3V, the students misidentified the tanned skin as jaundice more often when biased towards liver cancer than when biased towards stomach cancer, despite the fact that the sclera of the man's eyes were white. Also, they were more likely to misidentify the parotid swelling of the boy with mumps as a moon-shaped face when biased towards Cushing's disease in scenario 16V. The majority of the features in the present study were cardinal features of the diagnoses. In most cases, the identification of one of these features (e.g moon-shaped face or parotid swelling) will be an important cue of the diagnosis. The incorrect identification or the failure to detect even one of these features could have a crucial impact on the diagnostic accuracy of a clinician.

Although the students did identify more alternate features when biased towards the alternate diagnosis than when biased towards the correct one, they showed some sensitivity to the information present in the stimuli. In both conditions, the students identified a higher percentage of the correct features than of the alternate features. Even when biased towards the alternate diagnosis, they identified 33% percent of the correct features, but only 26% of the alternate ones. This result shows that while the suggested diagnosis did influence students' interpretation of the stimuli, this was not the only influence of feature identification. However, this influence is nonetheless remarkable

given the fact that the test cases were mostly textbook pictures selected to illustrate the correct diagnosis.

Experiment 1B. Initial Demonstration: Identifying the nature of the influence of the diagnostic context on medical residents' identification of features

Although Experiment 1A provides some evidence regarding the processes governing the influence of a suggested diagnosis in feature identification, the generalizability of the results is limited by the fact that the participants were medical students in their second year. The findings may have been driven by their lack of experience in extracting clinical features from the appearance of patients. Although they did show some discrimination by generating more of the depicted correct features than the potentially present alternate features, they may have been excessively biased by the diagnostic context due to their lack of experience. With clinical experience, clinicians may become better at interpreting features and less susceptible to the influence of a suggested diagnosis. Such a possibility is of pedagogic interest. If the observation of such diagnosis-specific feature identification is a phenomenon of novices only, educational interventions could be aimed at increasing clinical experience in novice clinicians. However, if diagnosis-specific feature identification is also observed with clinicians who have acquired clinical experience, then any intervention to reduce such susceptibility to the suggested diagnosis should be aimed at the reasoning process involved during clinical interviews rather than at increasing exposure to patients, as mere exposure and experience would not be sufficient to reduce or eliminate it.

Research into the reasoning processes of clinicians provides conflicting predictions as to whether or not medical residents' identification of features will be

influenced by the suggestion of a diagnosis to the same extent as the students in Experiment 1A. Some theories of the reasoning processes of problem solvers (Larkin, McDermott, Simon, & Simon, 1981; Patel & Groen, 1986) suggest that novices and experts differ in the direction of their reasoning. For example, while students reason in a backward direction (from theory to data), experts proceed in a forward direction (from data to theory) (Patel & Groen, 1986). Should such theories be accurate descriptions of the clinical reasoning process and its development in the acquisition of expertise, residents who have gained more clinical experience and who are presumably more expert than the medical students should be less susceptible to the influence of a diagnosis when identifying features, since they would gather the data prior to generating a diagnosis. A study by Myles-Worsley, Johnston and Simons (1988) also provides evidence suggesting that the impact of a suggested diagnosis might be smaller in clinicians with additional clinical experience. They showed both normal and abnormal (containing a radiological abnormality) x-ray films to observers with four levels of radiographic experience (novices to senior radiologists) and tested their recognition memory. They found that as experience increased, clinicians were better able to recognize the abnormal films while performance in recognizing the normal films decreased. Based on these results, Myles-Worsley et al. (1988) argued that the knowledge acquired with expertise includes knowledge about what is considered normal in a radiographic film. When viewing a new radiograph, the normal features are processed with minimal effort and attention is captured by the unexpected elements in the radiograph. This process allows the expert radiologist to allocate resources more efficiently and to selectively attend to the distinguishing features of the radiograph. As the normal, unperturbed version of the radiograph is represented in knowledge, perturbations of this representation become easier to identify. If such an explanation can be generalized to the recognition of visual features from patients'

appearance, then residents, having had more exposure to patients, should have a better representation of a normal patient appearance and be better able to recognize abnormalities. This would lead to their identification of features from photographs to be both more accurate and less susceptible to the influence of the considered diagnosis than that of the students in Experiment 1A.

Another line of evidence relating to changes with expertise is provided by a number of studies by Brooks, Norman and colleagues, that suggest that the influence of a diagnosis on feature identification is not a phenomenon limited to novices. As discussed in the introduction to Experiment 1A, Hatala et al. (1998) showed that residents and experts in cardiology were influenced by the case history when identifying the clinical features from electrocardiograms. Similarly, Brooks et al. (2000) showed that the interpreted description of features present in photographs of patients led to an increase in diagnostic accuracy for novices and experts. These results suggest that the identification of clinical features out of context of a given diagnosis can be problematic for experts as well as novices.

The goal of the present study is to test the hypothesis that the influence of a suggested diagnosis on feature identification will decrease as clinicians gain experience at recognizing these features. To that end, Experiment 1A will be replicated with residents having completed a minimum of 6 months of training in family medicine. This specialty was selected because it is an area of medicine in which clinicians are most likely to be exposed to a variety of disorders where little clinical evidence has been gathered prior to a first encounter with a patient. Residents from this specialty should have the most experience at identifying a wide range of clinical features under a variety of diagnostic contexts.

Methods

Subjects

The participants were 20 Family Medicine residents from McMaster University, who had completed a minimum of 6 months to a maximum or 18 months of their residency training. Approximately two thirds of the residents had completed between 16 and 18 months of their training, and the remainder had completed between 6 and 8 months of their training. Approval for conducting the study was granted by the ethics committee of McMaster University's medical program.

Design

The methods, procedure and analyses were identical to those of Experiment 1A

Results

The analyses of the diagnostic ratings demonstrated that the suggestion of a diagnosis and case history were sufficient to bias the residents towards one diagnosis or another (see Table 3). A main effect of diagnosis was observed for both the percent of cases in which the correct diagnosis was rated the most likely, $F(1,18)=84.26$, $\underline{MSE}=.220$, $\underline{p}<.01$, and the percent of cases in which the alternate diagnosis was rated the most likely, $F(1,18)=105.71$, $\underline{MSE}=.182$, $p<.01$. The correct diagnosis was rated most likely in 81% of the cases in which the residents received the correct case history and tentative diagnosis. As for the alternate diagnosis, it was rated the most likely in 69% of the cases in which they were presented with the alternate case history and tentative diagnosis. As with the students, there were main effects of scenario for the percent of cases in which they opted for the correct diagnosis, $F(7, 126)=3.99$, $\underline{MSE}=.090$, $\underline{p}<.01$, as well as the percent of cases in which they opted for the alternate diagnosis, $F(7, 126)=4.44$, $\underline{MSE}=.058$, $\underline{p}<.01$, with the residents showing higher accuracy for scenarios 9V and 12V

(Turner's syndrome/ cretinism and SLE / Steven-Johnson's disease). In addition, the analysis of conclusions for the alternate diagnosis indicated a diagnosis by scenario interaction, $F(7, 126)=4.44$, $\underline{MSE}=.058$, $\underline{p}<.01$. Although the residents opted for the alternate diagnosis more often when presented with the alternate diagnosis and case history, the effect was variable across the scenarios, showing the largest difference for scenarios 7V and 13V (polymyositis/hypothyroidism and Cushing's disease / nephrotic syndrome), and the smallest difference for scenarios 9V and 27V (Turner's syndrome / cretinism and acute glomerulonephritis / mumps).

Consistent with previous studies, there was a main effect for diagnosis on the percent of correct features generated by the residents, $F(1,18)=14.30$, $\underline{MSE}=.045$, $\underline{p}<.01$. The residents identified more of the correct features when presented with the correct diagnosis and case history (52.7%) compared to when presented with the alternate diagnosis and case history (40.0%), (see Table 4). There was also a main effect of scenario, $F(7, 126)=25.24$, $\underline{MSE}=.065$, $\underline{p}<.01$, with the residents detecting a higher percentage of the correct features for two scenarios (9V: Turner's and 12V: SLE) and the smallest percent for two scenarios (3V: stomach cancer and 13V: Cushing's). A significant diagnosis by scenario interaction, $F(2,126)=2.51$, $\underline{MSE}=.065$, $\underline{p}<.05$, indicated that although the residents generated more correct features in the bias correct condition for most scenarios, the suggestion of a diagnosis had no effect on the identification of the correct features of scenarios 9V and 27V (Turner's and acute glomerulonephritis). This result was consistent with the finding that, for these scenarios, the suggestion of the alternate diagnosis and case history did not lead to an increase in the percent of cases in which they opted for the alternate diagnosis.

The main question of interest in this study was whether the residents' identification of alternate features would be influenced by the suggestion of the alternate

diagnosis and case history in a manner comparable to the students in Experiment 1A. The results of the ANOVA indicate a main effect for diagnosis, $F(1,18)=43.04$, $\underline{MSE}=.022$, $\underline{p}<.01$, with the residents identifying a significantly higher percent of the alternate features when biased towards the alternate diagnosis (23.8%) than when biased towards the correct diagnosis (8.4%). This translates into the residents identifying approximately six features across the eight scenarios when biased towards the alternate diagnosis and identifying approximately two features across the eight test scenarios when biased towards the correct diagnosis. Again, there is a main effect of scenario, $F(7,126)=8.94$, $\underline{MSE}=.020$, $\underline{p}<.01$, with a higher percentage of alternate features generated for scenarios 3V, 7V, and 16V ("liver cancer", 'hypothyroidism", "Cushing's disease") and the smallest percentage for scenarios 9V and 27V ("cretinism' and "mumps"). The effect of the suggestion of a diagnosis on the generation of alternate features was variable, as indicated by a diagnosis by scenario interaction, $F(7, 126)=8.61$, $\underline{MSE}=.02$, $\underline{p}<.01$. For scenarios 3V and 18V (liver cancer, respiratory failure), the suggestion of a diagnosis apparently had no effect on the number of alternate features generated.

To test whether the residents performed similar to the students on this tasks, additional analyses of variance were conducted, adding expertise (students, residents) as a between subject variable. The main effects of expertise were non significant for the proportion of cases in which the correct diagnosis was rated most likely, $F(1, 36)=.496$, $\underline{MSE}=.157$, $\underline{p}=.49$, the proportion of cases in which the alternate diagnosis was rated most likely, $F(1, 36)=.006$, $\underline{MSE}=.132$, $\underline{p}=.94$, as well as the percent of correct, $F(1, 36)=2.15$, $\underline{MSE}=.072$, $\underline{p}<.15$, and alternate features identified $F(1, 36)=.0001$, $\underline{MSE}=.025$, $\underline{p}=.99$. The lack of significance of these analyses was not surprising, given the remarkable similarity in the diagnostic decisions and feature identification of the students and the residents. Similarly, there were no significant expertise by diagnosis interactions for the

percent of cases the correct diagnosis was rated most likely, $F(1, 36)=.38$, $\underline{MSE}=.157$, $\underline{p}=.58$, the percent of cases the alternate diagnosis was rated most likely, $F(1, 36)=.481$, $\underline{MSE}=.132$, $\underline{p}=.49$, the percent of correct features identified, $F(1,36)=.028$, $\underline{MSE}=.072$, $\underline{p}=.87$, nor the percent of alternate features identified, $F(1,36)=2.24$, $\underline{MSE}=.025$, $\underline{p}<.14$.

As in Experiment 1A, exploratory analyses were conducted to quantify the prevalence of each type of error. When the correct diagnosis was suggested, the residents correctly identified the four chameleon features in 85.0% of the cases, misidentified them as alternate features in 2.5% of the cases, and omitted them in 12.5% of the cases. When the alternate diagnosis was suggested, they correctly identified the chameleon features in 50% of the cases, mislabeled them as alternate features in 45% of the cases, and omitted them in 5% of the cases. As for the 21 features consisting of misidentification of normal variation, the residents identified them in 9.4% of the cases when the correct diagnosis was suggested to them and in 20.1% of the cases when the alternate diagnosis was suggested to them.

Discussion

The results of this study with residents indicate that the biasing effects on feature interpretation observed with students in Experiment 1A were not the result of a lack of clinical experience on the part of the students. Residents with 6 to 18 months of training in the recognition of features from patients' appearance are susceptible to the biasing effects of the suggested diagnosis to virtually the same extent as second-year medical students. Both students and residents opted for the correct diagnosis in 80% of the cases in which they were presented with the correct diagnosis and case history, and opted for the alternate diagnosis in two thirds of the cases in which the alternate diagnosis and case history were suggested. Similarly, both groups detected about half of the correct features

when biased towards the correct diagnosis and approximately one quarter of the alternate features when biased towards the alternate diagnosis.

The pattern of results for both the students and residents further indicate that the influence of the suggested diagnosis on feature identification is not independent of the information present in the stimuli. Both groups of participants identified more of the correct features present in the photographs than the alternate features not present in the stimuli. Even when biased towards the alternate diagnosis, the participants identified more than a third of the correct features, but only about a quarter of the alternate features. This result demonstrates some sensitivity to the constraints provided by the information contained in the photographs. Again, the strong effect of a diagnostic suggestion on feature interpretation is remarkable given that the alternate diagnosis was generated from pictures that were taken as clear illustrations of the correct diagnosis.

The finding that more experienced clinicians are as susceptible as novices to the suggestion of a diagnosis when identifying features from a patient's appearance poses a challenge to Patel and Groen's (1986) argument that clinicians shift from using backward reasoning to forward reasoning as they acquire experience. 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. An alternative explanation, which does not refute the argument of a shift towards 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 the residents might be closer to being novice clinicians than originally thought. A similar study is needed with more experienced clinicians to test whether the finding of a strong influence of the diagnosis of residents' feature identification resulted from the fact that backward

reasoning occurs even in experts or whether only clinicians with considerable experience will be immune to the influence of the suggested diagnosis. Therefore, although these results do not provide strong evidence against theories of clinical reasoning, they do refute the possibility that the results observed in Experiment 1A were the result of the very limited clinical experience of medical students.

The social hypothesis literature provides some potential insights into the processes underlying the influence of a suggested diagnosis on feature identification. When testing social hypotheses, individuals tend to engage in pseudodiagnostic, or selective, hypothesis testing (Mynatt, Doherty, & Dragan, 1993; Sanbonmatsu, Posovac, Kardes, & Mantel, 1998; Trope & Liberman, 1996). The hypotheses generated by individuals are thought to guide the gathering and integration of information in a top-down approach. People will test the sufficiency of the focal hypothesis (likelihood of obtaining the evidence when the hypothesis is true) rather than its necessity. This results in a neglect of alternatives as well as in a phenomenon called confirmation bias, due to the assumption that what is consistent with one hypothesis will be inconsistent with its alternatives. Thus, individuals will search for information that will support the focal hypothesis while neglecting the evidence supportive of the alternate diagnoses. A similar phenomenon of confirmation bias might be occurring in the present studies. In both the fields of medical decision making and social psychology, the tasks are typically those of testing hypotheses in the face of uncertainty, where the evidence can be ambiguous. Having a diagnosis is mind would lead students and residents to seek confirming evidence, and might also change their interpretation of the data towards one that is confirming of the suggested diagnosis. To reduce the confirmation bias, researchers in social hypothesis testing instruct their participants to consider alternatives, and report that

this reduces the degree to which they engage in pseudodiagnostic hypothesis testing (Sanbonmatsu et al., 1998; Trope & Liberman, 1996). If the results observed in the present study are due to a similar phenomenon, then reducing the degree of focus on the suggested diagnosis, which will presumably lead the participants to consider alternative diagnoses and search for features supporting these competing diagnoses, should reduce the influence of the suggested diagnosis on feature interpretation. This hypothesis is investigated in the following studies, using various techniques aimed at reducing the degree of focus placed on the suggested diagnosis.

If the Bayesian perspective is reconceptualized as a theory of subjective probability rather than a theory of objective probability, it becomes possible to predict that decreasing the focus placed on the suggested diagnosis should have an impact on features identified by the participants. Presumably, reducing the focus that is placed on the suggested diagnosis will serve to lower its revised probability (its probability, given the accompanying case history). A reconceptualization that accounts for the interrelation of diagnosis and feature identification would predict that the higher the revised probability of a diagnosis, the higher the conditional probability of its features. For clinicians, the more they believe that a patient is suffering from Cushing's disease, the more they believe that they will observe a moon-shaped face. The higher the subjective conditional probability of the feature, less perceptual evidence will be necessary for clinicians to report this feature. In other words, if clinicians strongly believe in a diagnostic possibility, they will have high a priori expectations that its features will be present, and be more likely to interpret the visual characteristics as those features. Therefore, any manipulation that reduces the revised probability of the suggested diagnosis should also result in a reduction in the conditional probability of its features,

leading clinicians to be less likely to identify these features as they would need more perceptual evidence to do so.

Chapter 4: Manipulating the credibility of the tentative diagnosis

Experiment 2A. Credibility Study: Impact of reducing the credibility of the provisional diagnosis with students

In Experiments 1A and 1B, the presence of a tentative diagnosis served both to guide the clinicians' search for features and to change the identification given to the features from a patient's appearance. One criticism of those studies that arose from informal discussion with educators and clinicians was the concern that the bias may have been excessively powerful, as the case histories were strongly biasing towards the suggested diagnosis. This raises the question of the frequency of occurrence of such feature misidentifications in clinical practice, and the conditions under which they occur. The goal of the present study is to investigate whether such effects on feature identification are mediated by the credibility of the diagnosis, that is, by the amount of supporting independent evidence accompanying the diagnosis. If a diagnosis is rendered less credible because it is accompanied by conflicting independent information, will its availability nevertheless guide the search for features as well as their identification? It is predicted that reducing the subjective revised probability of the diagnosis will also reduce the subjective conditional probability of its features, thereby leading clinicians to report fewer of its features because they will require more perceptual evidence to identify them.

The subjective revised probability of the suggested diagnosis is manipulated by varying the case history accompanying it. In the two proposed studies of the current chapter, students will suggested either the correct or the alternate diagnosis prior to being

asked to identify the clinical features from head-and-shoulder photographs of patients. For half of the students, the case history accompanying the diagnosis will the one supporting the opposite diagnosis. For example, they will be suggested the correct diagnosis but will read the case history suggestive of the alternate diagnosis (and vice-versa). If, as argued by clinicians, a diagnosis will only influence feature identification when it is very credible, then students should only list more features of the suggested diagnosis when it is accompanied by a case history supporting it. However, if the influence of the diagnosis is determined by its availability rather than by its subjective revised probability, students should identify more of its features when it is accompanied by a supporting case history and more features compatible with another diagnosis when it is accompanied by a case history supportive of that diagnosis.

Methods

Participants

The participants were 40 medical students, some attending medical school at McMaster University (n=29) and others attending medical school at the University of Illinois at Chicago (n=11). All students had completed between 19 and 22 months of formal class-based education, and were paid for taking part in the study. The students at McMaster University were clinical clerks, and those at the University of Illinois at Chicago (UIC) were second-year medical students. Approval to conduct the study was granted by the Internal Review Board at UIC and by the ethics committee of the McMaster University medical program.

Materials

The photographs, case histories and diagnoses used in this study were the same as those used in Experiment 1A and 1B.

Procedure

The procedure in the present experiment was almost identical to that in Experiment 1, except that the scenarios were presented to the participants in four conditions. In the first condition, the photographs were presented with the correct case history and tentative diagnosis. In the second, they were presented with the alternate case history and diagnosis. In the third condition, the photographs were presented with the correct case history and the alternate diagnosis; while in the fourth, they were presented with the alternate case history and the correct diagnosis. Four counterbalancing orders were created (10 students in each); thus each scenario was presented in each of the four conditions, and each participant saw two scenarios from each condition. As in Experiment 1, the order of presentation was fixed with the two practice scenarios presented first.

Measures

The measures of interest were the percent of cases in which the students opted for the correct diagnosis (gave it the highest probability rating), the percent of cases in which they opted for the alternate diagnosis, and the percent of correct and alternate features identified by the students. These four measures were analyzed with separate 2x2x8 mixed-design analyses of variance (ANOVA's) with the diagnosis (correct, alternate) and the history (correct case history, alternate case history) as between-subject variables and the scenarios (scenarios 3V, 7V, 9V, 12V, 13V, 16V, 18V, 27V) as a repeated-measures variable. Post-hoc analyses were conducted using Scheffe's test.

Results

The analysis of the diagnostic conclusions reached by the students was aimed at investigating whether the credibility of the suggested diagnosis was influenced by the accompanying case history. Each diagnosis was favored more often when it was the

suggested diagnosis, but the case history also played a role. The students favored the correct diagnosis in a higher percent of the cases when it was accompanied by the correct case history (68.8%) than when it was accompanied by the alternate case history (41.3%). Conversely, the alternate diagnosis was favored more often when accompanied by the alternate case history (23.7%) than the correct one (1.9%) (see Table 5). The ANOVA of the percent cases in which the correct diagnosis was rated most likely show that there was a main effect for diagnosis, $F(1, 36) = 84.5$, $\underline{MSE}= .168$, $\underline{p}< .01$, indicating that students were likely to opt for the correct diagnosis if it was the tentative diagnosis. The main effect of history, $F(1, 36) = 28.9$, $\underline{MSE}= .168$, $\underline{p}< .01$, indicates that the students were more likely to opt for the correct diagnosis when presented with the correct case history than when presented with the alternate case history. The results of the ANOVA show no diagnosis by history interaction, $F(1, 36) = .376$, $\underline{MSE}= .168$, $\underline{p}=.54$, indicating the contributions of each are additive rather than multiplicative. As expected, there was variance in the difficulty of the scenarios, as reflected by a main effect of scenario, $F(7, 252) = 2.76$, $\underline{MSE}= .136$, $\underline{p}< .01$. There was a significant scenario by history interaction, $F(7, 252) = 2.48$, $\underline{MSE}= .136$, $\underline{p}< .05$, with some scenarioss being more influenced by the history than others (scenarios 9, 18 and 27, in particular).

Similar findings were observed with the analysis of the percent of cases in which the alternate diagnosis was rated the most likely. A main effect of diagnosis, $F(1, 35) = 86.72$, $\underline{MSE}= .183$, $\underline{p}<.01$ indicates that the alternate diagnosis was rated most likely more often when it was suggested as the tentative diagnosis (see Table 5). There was also a main effect of history, $F(1, 35) = 29.14$, $\underline{MSE}= .183$, $\underline{p}<.01$, with the alternate diagnosis rated most likely more often when it was suggested by the case history. The diagnosis by history interaction, $F(1, 35) = 10.55$, $\underline{MSE}= .183$, $\underline{p}<.01$, suggests that the two influences are multiplicative rather than additive. As with the analysis of the conclusion in favor of

the correct diagnosis, there was evidence for variability in the difficulty of the scenarios, with a main effect for scenarios, $F(1, 245) = 2.07$, MSE= .098, p<.05.

An exploratory analysis of the data suggests that the tentative diagnosis had a larger impact on the diagnostic conclusions than the case history. The difference in concluding for the correct diagnosis (42.2% absolute difference) and the alternate diagnosis (43.9% absolute difference) was larger when the suggested diagnosis was manipulated than when the case history was manipulated (24.6% and 24.9% absolute difference, respectively for the correct and the alternate diagnoses)

The finding that the diagnostic conclusions were influenced both by the availability of the diagnosis as well as by the case history were expected and treated as manipulation checks to ensure that the credibility of the diagnosis was affected. The main purpose of the study was to observe the effects of these manipulations on the features identified by the students. Table 6 shows the percent of correct and alternate features generated by the students. When compared with the condition where the students were provided with the alternate diagnosis and case history (41.2%), the addition of either a correct case history or tentative diagnosis was not sufficient to lead to an increase in the detection of correct features (41.7% and 42.5% respectively). It was only when both the correct case history and tentative diagnosis were provided that an increase in the detection of correct features occurred (53.7%). This was confirmed by the results of the ANOVA, which shows a diagnosis by history interaction, $F(1, 36) = 2.11$, MSE= .501, p< .05, but no main effect for diagnosis, $F(1, 36) = 1.80$, MSE= .501, p=.07, and no main effect of history, $F(1, 36) = 1.80$, MSE= .501, p=.07. There was a main effect for scenarios, $F(7, 252) = 25.05$, MSE= .444, p<.00, indicating that there was variability in the number of correct features listed for each scenario, but this did not interact with either the history,

$F_{(7, 252)} = 1.87$, MSE= .444, $p=.08$, or the diagnosis, $F_{(7, 252)} = .773$, MSE= .444, $p=.61$.

The pattern of results is somewhat different for the percent of alternate features identified by the students. Compared to the condition in which they considered the correct diagnosis (15.4%), the suggestion of the alternate diagnosis was sufficient to lead to an increase in the identification of alternate features (26.3%), as confirmed by a main effect for diagnosis, $F_{(1, 36)} = 19.22$, MSE=.049, $p<.01$. However, the presentation of the alternate case history was not sufficient to lead to an increase in generating alternate features (22.8% alternate case history versus 18.9% correct case history) , regardless of the diagnosis that was suggested, as evidenced by no significant main effect for history, $F_{(1, 36)} = 2.40$, MSE= .049, $p=.13$, nor a diagnosis by history interaction, $F_{(1, 36)} =$ .379, MSE= .049, $p=.54$. There was a main effect for scenarios, as well as a significant scenario by diagnosis interaction, $F_{(7, 252)} = 19.08$, MSE= .040, $p<.01$, scenario by history interaction, $F_{(7, 252)} = 2.67$, MSE= .040, $p<.05$, and scenario by history by diagnosis interaction, $F_{(7, 252)} = 3.31$, MSE= .040, $p< .01$.

An exploratory analysis of the data revealed that the suggested diagnosis appeared to have a larger impact on the identification of the features than the case history had . Although the magnitude of the differences is not as large as those observed for the diagnostic conclusions, the differences in identifying the correct features (6.7% absolute difference) and the alternate features (10.9% absolute difference) appeared larger when the suggested diagnosis was manipulated than when the case history was manipulated (0.8% and 3.9% absolute difference, respectively for the correct and the alternate features).

Discussion

The most interesting finding from this experiment was that changing the credibility of the diagnosis had a larger impact on the diagnosis than on the identified features. Compared to the condition in which the students were provided with the alternate diagnosis and case history, the correct diagnosis had to be accompanied with the correct case history in order to lead to an increase in the percent of correct features detected by the students. Thus, the correct diagnosis had to be credible in order to influence feature interpretation. Simply considering a diagnosis or having information that was consistent with it was not sufficient to lead to an increase in the detection of the correct features.

However, varying the credibility of the alternate diagnosis had little impact on the identification of alternate features. Compared to the students receiving the correct diagnosis and case history, the availability of the alternate diagnosis was sufficient to lead students to identify more alternate features. Once the alternate diagnosis was considered, the provision of a consistent case history, and consequently an increase in its credibility, did not lead to an increase in the identification of the alternate features (25.2% vs 27.5%). Thus, although the correct diagnosis needed to be credible for students to detect more correct features, the same was not true for the alternate diagnosis.

Another interesting finding in this study was that the manipulation of the suggested diagnosis appeared to have a stronger impact on the diagnostic conclusions and on feature interpretation than the case histories did. This suggests that the availability of a diagnosis plays a larger role in diagnostic conclusions and feature identification than does the independent evidence in favor of it.

Experiment 2B. Credibility Study: Impact of reducing the credibility of the provisional diagnosis with residents

As with Experiment 1A, the participants in the Experiment 2A were second-year medical students. Although it was shown in Experiments 1A and 1B that students and residents are equally susceptible to the suggestion of a diagnosis when identifying features, the importance of the credibility of the diagnosis might differ between the two groups. In the formal education of medical students, the emphasis is placed on the relationship between diseases and their features. With clinical experience, anecdotal evidence from clinicians suggests that more emphasis is placed on the case histories when generating a diagnosis from patient information If this is true, the credibility of the diagnosis might play a larger role in mediating the influence of a diagnosis on feature identification by residents.

Methods

Participants

The participants in the study were 28 family medicine residents from the University of Illinois at Chicago (n=15) and from the University of Toronto (n=13), who had completed between 20 months and 32 months of residency training. They were paid for taking part in the study. Approval to conduct the study was granted by the ethics committees of McMaster University and the University of Toronto's medical programs.

Design

The materials, procedure and measures of the present study are the same as those used in Experiment 2A.

Results

Similar to the students in Experiment 2A, the residents opted for a diagnosis most often when it was the suggested diagnosis. Also, the case history had an impact on

diagnostic conclusions. The residents opted for the correct diagnosis in a higher percent of cases when it was accompanied by the correct case history than when it was accompanied by the alternate case history. Conversely, the alternate diagnosis was favored more often when accompanied by the alternate case history than the correct one (see Table 7). The ANOVA of the percent cases in which the correct diagnosis was rated most likely show that there was a main effect for diagnosis, $F(1, 24) = 46.26$, MSE= .242, $p< .01$, indicating that students were likely to favor the correct diagnosis if it was the tentative diagnosis. The main effect of history, $F(1, 24) = 14.02$, MSE= .242, $p< .01$, shows that the correct diagnosis was more likely to be rated most likely if it was suggested by the case history. The results of the ANOVA show no diagnosis by history interaction, $F(1, 24) = .382$, MSE= .242, $p=.54$, indicating the contributions of each are additive rather than multiplicative. As expected, there was variance in the difficulty of the scenarios, as reflected by a main effect of scenarios, $F(7, 168) = 2.21$, MSE= .137, $p< .05$.

Similar findings were observed with the analysis of the percent of cases in which the alternate diagnosis was rated the most likely. A main effect of diagnosis, $F(1, 22) = 33.62$, MSE= .238, $p<.01$ indicates that the residents opted for the alternate diagnosis most often when it was the suggested diagnosis. There was also a main effect of history, $F(1, 22) = 10.88$, MSE= .238, $p<.01$, with the residents concluding for the alternate diagnosis most often when it was suggested by the case history.

An exploratory analysis of the data revealed that the suggested diagnosis had a larger impact on the diagnostic conclusions than did the case history. The difference in concluding for the correct diagnosis (44.9% absolute difference) and the alternate diagnosis (39.3% absolute difference) was larger when the suggested diagnosis was manipulated than when the case history was manipulated (24.5% and 22.4% absolute difference, respectively for the correct and the alternate diagnoses)

The identification of the features was influenced differently by the diagnostic context than were the diagnostic conclusions. Table 8 shows the mean percent of correct and alternate features identified by the residents. The credibility of the diagnosis did not have an impact on the percent of correct features generated by the residents. The nature of the suggested diagnosis was the only contributing factor. This is supported by the finding of a significant main effect of diagnosis, $F(1, 24) = 5.35$, MSE=.083, p<.05, but no main effect of history, $F(1, 24) = .855$, MSE=.083, p=.36, and no significant diagnosis by history interaction, $F(1, 24) = .855$, MSE=.083, p=.36. Overall, the residents who were suggested the correct diagnosis identified more correct features than those who were suggested the alternate diagnosis (45.2% vs. 36.3%, Scheffe's test, p<.05). However, when the correct diagnosis was suggested, the residents did not identify fewer correct features when given the alternate history than when given the correct one (45.2% vs. 45.2%, Scheffe's test, p=1.0).

The identification of the alternate features was influenced by the credibility of the alternate diagnosis (see Table 8). There was a main effect for diagnosis, $F(1, 24) = 24.22$, MSE=.052, p<.01, and of history, $F(1, 24) = 6.76$, MSE=.052, p<.05, but no significant diagnosis by history interaction, $F(1, 24) = 2.88$, MSE=.052, p=.10. The lack of interaction suggests that the two sources of influence are additive rather than multiplicative. Scheffe's tests were conducted to explore marginal trends towards significance of the interaction. Overall, the residents who were presented with the correct diagnosis identified fewer alternate features than those presented with the alternate diagnosis (13.5% vs. 28.5%), and those who read the correct history identified less alternate features than those who read the alternate history (17% vs. 25%). Of those who were presented with the correct diagnosis, there was no difference in the identification of the alternate features depending on whether they read the correct or the alternate case

history (12% vs. 15%, Scheffe's test, p=.45). However, of those who were presented with the alternate diagnosis, those who read the alternate case history identified more alternate features than did those who read the correct case history (35% vs. 22%, Scheffe's test, p<. 05).

An exploratory analysis of the data revealed that the suggested diagnosis had a larger impact on the identification of features than did the case history. The difference in the proportion of correct features (8.9% absolute difference) and alternate features (15.1% absolute difference) identified was larger when the suggested diagnosis was manipulated than when the case history was manipulated (3.6% and 7.9% absolute difference, respectively for the correct and the alternate diagnoses)

The pattern of results of the residents appeared to differ from that of the students in 2A. To test whether these differences were significant, additional analyses of variance were done on the diagnostic conclusions and identification of the features by adding expertise (students, residents) as a grouping factor. For the percent of cases in which the participants opted for the correct diagnosis, the main effect of expertise was not significant, $F(1, 60) = 1.46$, MSE=.266, p=.23, neither were the expertise by diagnosis interaction, $F(1, 60) = .073$, MSE=.266, p=.79, the expertise by history interaction, $F(1, 60) = .01$, MSE=.266, p=.92, nor the expertise by diagnosis by history interaction, $F(1, 60) = 073$, MSE=.266, p=.79. Similarly, the analysis of the percent of cases in which the participants opted for the alternate diagnosis revealed no main effect of expertise, $F(1, 57) = 2.18$, MSE=.204, p=.15, as well as no significant expertise by diagnosis, $F(1, 57) = .503$, MSE=.204, p=.48, expertise by history, $F(1, 57) = .214$, MSE=.204, p=.65, or expertise by diagnosis by history, $F(1, 57) = .112$, MSE=.204, p=.74, interactions.

The identification of the correct features did not differ between the students and the residents, as indicated by a lack of significant main effect of expertise, $F(1, 60) = 1.24$,

$\underline{MSE}$=.502, $\underline{p}$=.27, as well as non-significant expertise by diagnosis interaction, F(1, 60) =.008, $\underline{MSE}$=.502, $\underline{p}$=.93, expertise by history interaction, F(1, 60) =.405, $\underline{MSE}$=.502, $\underline{p}$=.53, or an expertise by diagnosis by history interaction, F(1, 60) =3.59, $\underline{MSE}$=.502, $\underline{p}$=.07. Similarly, the analysis the percent of alternate features identified revealed no main effect of level of expertise, F(1, 60) =.287, $\underline{MSE}$=.387, $\underline{p}$=.59, no level of expertise by diagnosis interaction, F(1, 60) =.083, $\underline{MSE}$=.387, $\underline{p}$=.77, no level of expertise by history interaction, F(1, 60) =.867, $\underline{MSE}$=.387, $\underline{p}$=.36, as well as no level of expertise by diagnosis by history interaction, F(1, 60) =1.67, $\underline{MSE}$=.387, $\underline{p}$=.20.

Discussion

As observed in Experiment 2A with medical students, the suggestion of a diagnosis and a case history had an impact on the credibility of both the correct and alternate diagnoses. Both students and residents opted for the suggested diagnosis more often when it was accompanied by a supporting case history than when it was accompanied by a case history suggestive of the opposite diagnosis.

A main contribution of the present experiment, along with Experiment 2A, is to provide evidence of the robustness of the findings from Experiments 1A and 1B, in which the suggested diagnosis has a strong influence on feature identification. In both sets of studies, clinicians identified more correct features when the correct diagnosis was suggested to them. When the alternate diagnosis was suggested, they identified fewer correct features and misidentified more alternate features. Sensitivity to the information provided by the stimuli also appears to be a robust finding. In all four studies, the clinicians identified more correct features than alternate features, indicating that the influence of the diagnosis was mediated by the information provided by the data.

As in Experiment 2A, the credibility of the suggested diagnosis had a different impact on the identification of correct features than on the identification of alternate

features. The credibility of the correct diagnosis did not have an influence on the identification of the correct features. The availability of the correct diagnosis, regardless of whether the case history supported it or not, led to an increase in the detection of correct features over the cases in which the residents were suggested the alternate diagnosis. However, the identification of the alternate features was influenced by both the availability and the credibility of the alternate diagnosis. The residents identified more alternate features when the tentative diagnosis was the alternate one than when it was the correct one. In addition, of the residents who were suggested the alternate diagnosis, those that read the alternate case history identified more of the alternate features than did those who read the correct case history.

These two studies provide mixed results regarding the role of the credibility of the diagnosis in influencing feature identification. For the students, the credibility of the correct diagnosis mediated its influence on feature identification. For the residents, the credibility of the alternate diagnosis mediated its effect on feature identification. However, when the results of the students and residents were analyzed together, the level of expertise did not interact significantly with the effect of suggesting a diagnosis, the effect of a case history, nor with the interaction of the diagnosis and case history (its credibility). Thus the apparent differences between the residents and students result from a small effect, and are not statistically significant. This lack of significance for the effect of expertise were not surprising given that for both students and residents, feature identification was strongly influenced by the suggestion of a diagnosis. There was no consistent pattern for either the students or the residents for when the credibility of the diagnosis has an impact on feature identification. Together, these results suggest that the influence of the availability of a diagnosis on feature identification is not strongly or consistently affected by the independent evidence in favor of it. Although the credibility

of the diagnosis appeared to have an effect on feature identification in some situations, the effects were variable and small with respect to the major results of both sets of studies; that is, the finding that the availability of a diagnosis serves to guide feature search and identification. In addition, across all measures, the impact of the suggested diagnosis was stronger than the impact of the case history. These results address some of the concerns of educators regarding the results of Experiments 1A and 1B, suggesting that, in clinical practice, a suggested diagnosis will have a strong impact on feature identification, even when the diagnosis is not judged to be very credible. Such results do not support the predictions that were generated from the extension of the Bayesian perspective, as discussed in Chapter 3. Changing the conditional probability of the diagnosis did not appear to impact on the conditional probability of the presence of its features, in as much as it had no impact on the identification of these features.

Chapter 5: Considering alternatives

The finding that feature identification is influenced by the availability of a diagnosis provides a problem for medical educators. Although considering the correct diagnosis will increase students' and residents' accuracy in identifying features from a patient's appearance, considering an incorrect diagnosis will lead to errors in the direction of providing additional support for this erroneous diagnosis. The challenge now lies in identifying interventions to help students and residents reinterpret the evidence when initially biased in the wrong direction, especially since the results from Experiment 2 show that a diagnostic hypothesis will influence feature identification even when its credibility is low.

The influence of a suggested diagnosis on feature identification might result from the clinicians engaging in pseudodiagnostic or selective hypothesis testing. By focusing solely on a leading hypothesis and neglecting the alternatives, they engage in testing that is biased towards confirming evidence and assesses only the sufficiency of the hypothesis. A number of researchers suggest that "instructing participants to consider the alternatives" engages them in more diagnostic or comparative hypothesis testing (Arkes, 1981; Klayman, 1995; Trope & Liberman, 1996). This results in a more exhaustive analysis of both the focal hypothesis and its alternatives, leading to a search for evidence that will be more effective in testing the likelihood of the various alternatives. Rather than simply testing the sufficiency of the leading hypothesis, participants test the necessity of the hypothesis, that is, whether the hypothesis is necessary for obtaining the evidence. Thus, diagnostic testing might reduce the impact of a suggested diagnosis on feature

identification. The two studies in this chapter consist of attempts to induce diagnostic hypothesis testing in students by reducing the degree of focus placed on the suggested diagnosis. The hypothesis is that reducing the degree of focus placed on the suggested diagnosis will lead students to consider the alternative diagnoses and reduce the impact of the suggested diagnosis on feature identification.

Experiment 3. Impact of a competing diagnosis: Does the subsequent suggesting of a differential diagnosis reduce the effect of an initial diagnosis on feature identification?

The goal of the present study is to investigate whether getting students to consider a differential diagnosis, once they are already biased towards an initial diagnosis, will lead to the reduction of the influence of the initial diagnosis on feature identification. Before any educational intervention is to be developed to get students to correct themselves when initially biased towards an incorrect diagnosis, we need to establish whether the initial feature interpretation is flexible. Once they are biased towards an incorrect diagnosis and have improperly interpreted the clinical signs in light of this diagnosis, will the suggestion of the correct diagnosis reduce their errors? There are two possible ways that this could occur. Reminding them of the correct diagnosis could remind them of features they might otherwise overlook or it could promote the reinterpretation of a clinical sign that has been misinterpreted as a feature of the incorrect diagnosis.

The literature on order of information presentation provides conflicting predictions on whether such a manipulation will be effective in reducing

misidentifications of features. Although there is some research suggesting that the presence of an alternative diagnostic hypothesis will lead clinicians to engage in more diagnostic and comparative hypothesis testing, there also exist findings suggesting that clinicians do not reinterpret initial information in light of new information. In the field of medical education, there are two classic studies aimed at investigating the effect of order of information on the integration of information.

Bergus, Chapman, Gjerde, and Elstein (1995) found that clinicians gave more weight to information regarding prior medical history when it was presented last in a clinical case than when it was presented first. They presented clinicians with a real case of a patient with a history of lung cancer who presented with a new neurological disturbance and a normal computerized tomographic (CT) scan. All participants received the same information. One group was informed about the prior history of lung cancer upon initial contact with the case, while the other group was informed of the history of lung cancer towards the end of the case. The clinicians who received the history of lung cancer last were more likely to decide that the patient was suffering from metastatic cancer of the brain (correct diagnosis) than were those who received the same information first. Bergus et al (1995) argued that receiving the cancer history last led clinicians to give it more weight. This latter information was then used to reinterpret the earlier findings from the case, leading clinicians to generate an alternative explanation for the findings.

Conversely, Cunnington, Turnbull, Regehr, Marriott and Norman (1997) found that clinicians in their study showed a tendency to favor the diagnosis for which information was presented first. The clinicians worked through ten cases that were designed such that the symptoms were suggestive of two competing diagnoses (e.g. pneumonia versus pulmonary embolism). These cases were designed in a series of discrete statements replicating the information acquired by a clinician when taking a

history from a patient, and the totality of the information equally supported the two competing hypotheses. One group of clinicians received all the information for diagnosis A at the beginning of the list of statements and all the information consistent with diagnosis B at the end of the list of statements, while the other group received the same information in the reversed order. The clinicians tended to favor the diagnosis for which the evidence has been provided in the beginning of the list.

One difference between the Cunnington et al. (1997) study and the Bergus et al. (1995) study that might account for the different results lies in the nature of the evidence presented to participants. In the Cunnington et al. (1997) study, each item of evidence was supportive of only one of the two diagnoses, and could not be reinterpreted to be supportive of the competing diagnosis. In contrast, the presence of a new neurological disturbance accompanied by a normal CT scan could be consistent with both a diagnosis of a transient ischemic attack or a diagnosis of metastatic cancer, depending on how it was integrated with the other clinical information. In this case, information about a prior history of lung cancer changed the interpretation provided to the clinical findings. Providing this information late in a case might serve to remind clinicians of an alternate interpretation of the data. In contrast, the information in the Cunnington et al. (1997) study would not cause the reinterpretation of the earlier presented data. In this case, if the earlier presented information clearly supported one diagnosis and lead clinicians to think that there was sufficient information to support their diagnostic hypothesis, they might be more susceptible to discounting new information. The information in the Bergus (1995) study, by not clearly pointing to one diagnosis, might have left the clinicians unsatisfied with the sufficiency of the diagnosis, leaving them more susceptible to subsequent information causing the reinterpretation of the initial information to be strongly consistent with one of the diagnoses.

The materials in this thesis provide for an interesting extension in which all of the evidence was available at once and only the availability of the diagnosis was varied. A fundamental difference between the previous studies and the present one were that while Bergus et al. (1995) and Cunnington et al. (1997) added information as the case progressed, all of the information was present from the beginning in the present study. It was the time at which a diagnosis was suggested that was manipulated. However, all three studies were similar in that they dealt with the reinterpretation of the information. Therefore, the results of the Cunnington et al. (1997) and the Bergus et al. (1995) studies suggested different predictions for the present study. The clinical histories, by being strongly suggestive of one diagnosis resembled the data from the Cunnington et al. (1997) study, while the information in the photographs resembled the more ambiguous data of the Bergus et al. (1995) study in that they could be interpreted as consistent with the correct diagnosis or as consistent with the alternate diagnosis. Thus, there were two possible patterns of results that could follow from the subsequent suggestion of a competing diagnosis. If the clinical case history and initial suggestion of a diagnosis (and the resulting initial interpretation of feature) leave the students thinking that there is sufficient support for the initial diagnosis, the subsequent suggestion of a competing diagnosis will be discounted and should not impact on their identification of features. They should thus show a pattern of feature identification resembling that of Experiment 1A. However, if the subsequently suggested diagnosis is attended to sufficiently to provide an alternate interpretation to the data, the groups should identify fewer features of the initial diagnostic suggestion than did students in Experiment 1A

Methods

Participants

Twenty medical students from McMaster University Medical School took part in this experiment. The students were in the 17th to the 27th month of medical school, and were paid for their participation. Approval for the study was granted by the ethics committee of the McMaster University medical program.

Materials

The case histories, photographs and diagnoses were the same as those used in Experiments 1 and 2.

Procedure

The students worked through 10 scenarios on an individual basis. These scenarios were the same 2 practice scenarios and 8 test scenarios used in Experiments 1 and 2. For each case, the procedure was as follows. On a first page, the students were shown a head-and-shoulders photograph of a patient accompanied by either the correct or alternate case history and diagnosis. They were instructed to read the case history and to look at the photograph until they felt they were comfortable with the patient's condition. They then turned to a second page, where they read that the differential diagnosis was the opposite of the tentative diagnosis (alternate diagnosis when the tentative diagnosis had been the correct one, and vice-versa). They were then asked to write down *any and all* clinically important features observed in the photograph and to give a likelihood rating (on a scale of 0 to 100%) to the tentative diagnosis, the differential diagnoses, to any self-generated diagnosis(es), and to a residual category. They were instructed that these represented all possible options, and that their ratings should sum to 100%. See Appendix D for an example of the answer sheets used in this experiment.

The students were divided into two counterbalancing orders, such that each order had four scenarios in which the initial bias was towards the correct diagnosis and four scenarios in which the initial bias was towards the alternate diagnosis. The order of presentation of the scenarios was held constant for all participants.

<u>Measures</u>

As in Experiments 1 and 2, the measures of interest were the percent of cases in which the students opted for the correct diagnosis and the percent of cases in which they opted for the alternate diagnosis. To measure the impact of the experimental manipulations on the detection of features, the percent of correct and the percent of alternate features identified by the students were recorded. These four measures were submitted to separate 2x8 mixed-design ANOVAs with the condition (Correct diagnosis and history first with alternate diagnosis second; and Alternate diagnosis and case history first with correct diagnosis second) as the between subject variable, and the scenarios as a repeated measures variable.

<div align="center">Results</div>

The suggestion of a subsequent diagnosis was hypothesized to have an effect on both the diagnostic conclusions and the feature perception of the students. The analysis of the percent of cases in which the students opted for the correct diagnosis showed a main effect of condition, $F(1,20)=73.67$, $\underline{MSE}=.11$, $p<.01$. They opted for the correct diagnosis in 78% of the cases when the initial bias was in the direction of the correct diagnosis, and in 35% of the cases when they were initially biased towards the alternate diagnosis (see Table 9). As with the previous studies, there was variability in the difficulty level of the cases, $F(7, 140)=1.76$, $\underline{MSE}=.129$, $p<.01$, with the diagnostic accuracy being highest for cases 9V, 12V and 16V (Turner's, SLE, and mumps) and the lowest for cases 3V, 7V, and 18V (stomach cancer, polymyositis, and hypothyroidism).

There was also a main effect of condition in the percent of cases in which the students opted for the alternate diagnosis, $F(1,20)=66.53$, MSE=.185, p<.01. They opted for the alternate diagnosis in 60% of the cases that they were initially biased towards the alternate diagnosis and in 7% of the cases in which they were initially biased towards the correct diagnosis (see Table 9). There was variability in the level of difficulty of the scenarios, $F(7, 140)=2.31$, MSE=.105, p<.05. There was also a significant condition by scenario interaction, $F(7, 140)=2.45$, MSE=.105, p<.02. For all of the scenarios, the students opted for the alternate diagnosis more often when initially biased towards it, but the size of the effect varied across the scenarios.

The analysis of the percent of correct features identified by the students revealed no main effect of condition, $F(1,20)=.687$, MSE=.059, p=.42. Regardless of whether the correct diagnosis was considered first or after they were initially biased towards an alternate diagnosis, the students identified an equal percent of correct features (43.7% vs. 40.6% respectively; see Table 10). The analysis of the percent of alternate features identified revealed a main effect of condition, $F(1,20)=5.93$, MSE=.043, p<.05. The students who were initially biased towards the alternate diagnosis still identified more alternate features than those who were initially biased towards the correct diagnosis. There was also a main effect of scenarios, $F(7, 140)=6.36$, MSE=.085, p<.01, with the students identifying a higher percentage of the correct features in scenarios 9V, 12V, and 18V (Turner's, SLE and hypothyroidism), and the least in scenarios 3V, 13V, and 16V (stomach cancer, Cushing's disease, mumps). This higher percent of alternate features listed in the bias to alternate condition was not consistent for all scenarios, as shown by a scenarios by bias interaction, $F(7, 140)=2.19$, MSE=.044, p<.01. In scenarios 3V, 12V, 18V and 27V ("liver cancer", "Steven-Johnson's", "respiratory failure", and "mumps"),

the students listed the same percent of alternate features, regardless of whether the alternate diagnosis was considered first or second.

An exploratory analysis was conducted to quantify the prevalence of each type of error. When initially suggested the correct diagnosis, the students correctly identified the four chameleon features in 67.5% of the cases, misidentified them as alternate features in 15% of the cases, and omitted them in 17.5% of the cases. When initially suggested the alternate diagnosis, they correctly identified the chameleon features in 57.5% of the cases, mislabeled them as alternate features in 22.5% of the cases, and omitted them in 20% of the cases. As for the 23 features consisting of misidentification of normal variation, the students identified them in 13.7% of the cases when the correct diagnosis was suggested to them and in 21.2% of the cases when the alternate diagnosis was suggested to them.

To test whether the subsequent presentation of a competing diagnosis had an impact on the influence of a suggested diagnosis on diagnostic conclusions and feature interpretation, the results of this study were compared to those in Experiment 1A. The task in Experiment 1A was similar to the present task, with the exception that those students were not subsequently presented with a competing alternative after having been initially biased towards the correct or the alternate diagnosis.

The results of a 2x2x8 analysis of variance (study x diagnosis x scenarios) indicate that students in the present study opted for the correct diagnosis in a higher percent of cases when subsequently presented with a competing diagnosis (56.8%) than did students in Experiment 1A (43.8%), as confirmed by a main effect of study, $F_{(1, 38)}=14.51$, $\underline{MSE}=.103$, $p<.01$. In addition, a significant study by diagnosis interaction, $F_{(1, 38)}=18.16$, $\underline{MSE}=.103$, $p<.01$, indicates a smaller influence of the initial diagnostic suggestion on opting for the correct diagnosis in the present study, as compared to the influence of the initial diagnostic suggestion in Experiment 1A. There was a difference of

43.2 percentage points between the biasing conditions in the present study, compared to a difference of 73.9 percentage points in Experiment 1A.

A similar analysis of the percent of cases in which the students in each study opted for the alternate diagnosis revealed no main effect of study, $F(1, 38)=.067$, $MSE=.178$, $p=.80$. However, the initial suggestion of a diagnosis had less of an impact on the percent of cases in which the students opted for the alternate diagnosis in the present study than in Experiment 1A, as confirmed by a study by diagnosis interaction, $F(1, 38)=9.94$, $MSE=.178$, $p<.01$. There was a difference of 52.8 percentage points between the biasing conditions in the present study, compared to a difference of 63.8 percentage points in Experiment 1A.

The subsequent presentation of a competing diagnosis had no impact on the overall percent of correct and alternate features identified by the students. Students in Experiment 1 who received only the initial suggestion of either the correct of alternate diagnosis identified the same percent of correct (41.9%), $F(1, 38)=.002$, $MSE=.078$, $p=.96$, and alternate features (16.1%), $F(1, 38)=.469$, $MSE=.036$, $p=.50$, as did students who received the subsequent presentation of a competing diagnosis in the present study (42.1% of correct features and 17.6% of alternate features). The subsequent presentation of the competing diagnosis did not reduce the influence of the initial diagnostic suggestion on the percent of correct features identified in the present study, compared to the influence of the suggested diagnosis in Experiment 1A. This was confirmed by a non-significant study by diagnosis interaction, $F(1, 38)=3.10$, $MSE=.078$, $p=.09$. The subsequent presentation of a competing diagnosis did reduce the impact of the initially suggested diagnosis on the identification of alternate features, as indicated by a significant study by diagnosis interaction, $F(1, 38)=10.39$, $MSE=.036$, $p<.01$.

Discussion

The effect of suggesting a subsequent diagnosis when students were initially biased towards another diagnosis provides a mixed answer to the question of whether clinicians' identification of features from patient appearance is strongly affected by the information considered first and is inflexible to reinterpretation, or whether they place sufficient weight on later information and reinterpret the evidence based on it.

In looking at the diagnostic decisions of the students, it can be observed that students attended both to the initial suggested diagnosis and, to a certain extent, to the subsequent suggested diagnosis. Overall, the students opted for the correct diagnosis more often in the present study (56.8%) than did the students in Experiment 1A (43.8%), but both groups opted for the alternate diagnosis in an equal percent of the cases (33.8% in the present experiment and 35.0% in Experiment 1A). Also, the impact of the initial suggestion of a diagnosis was lessened by the presence of a subsequent diagnosis. Thus, the students did appear to be attending to the information presented later in the case, even if they placed the most weight on the initial information when drawing diagnostic conclusions.

The order in which the correct diagnosis was presented did not seem to impact on the identification of the correct features. The students identified as many of these features when initially biased towards the correct diagnosis as when initially biased towards the alternate one (43.7% vs. 40.6% respectively). Thus, consistent with the previous studies in this thesis, the availability of the correct diagnosis appears to suggest features to look for. The results of this study add that the time at which the diagnosis is considered is not important.

The pattern of results for the identification of the alternate features differed from that of the correct features. Again, there was evidence that the availability of the

diagnosis leads to more identification of its features. The students who were initially biased towards the correct diagnosis and received the subsequent suggestion of the alternate diagnosis identified more alternate features (14%) than did the students in Experiment 1A who were biased towards the correct diagnosis (6%). The significant study by diagnosis interaction suggests that the influence of the initially suggested diagnosis on the identification of alternate features was larger in Experiment 1A than in the present experiment. In looking at the data, it appears that this result is due to the subsequent presentation of the alternate diagnosis leading students to identify more alternate features than did students in Experiment 1A, who only received the suggestion of the correct diagnosis. However, the subsequent presentation of the correct diagnosis did not appear to result in students identifying fewer alternate features than did the students in Experiment 1A who only received the suggestion of the correct diagnosis (21.4% versus 26.5% respectively)

This study replicated the findings of the previous experiments of this thesis in demonstrating that the availability of a diagnosis appears to influence diagnostic decisions as well as feature identification. This influence of a provisional diagnosis does not appear to be substantially changed by the time at which a diagnosis is considered. Even when the competing diagnoses were presented subsequent to the students being biased towards an initial diagnosis, the initially presented correct and alternate diagnosis led to the students identifying more of each diagnosis' features than students in Experiment 1A who did not receive these subsequent suggestions. However, information that is presented later in a case does not lead to a reinterpretation of features already identified. The subsequent suggestion of a correct diagnosis did not lead to a reduction in the errors consisting of identifying alternate features. In other words, once the students had seen and labeled a feature, it was difficult for them to see it as something else. Thus,

once the "jaundice" in case 3 or the "moon-shaped face" of case 16 had been identified, the students did not reinterpret them as tanned skin or parotid swelling, even when suggested a diagnosis that would provide this latter interpretation of the data. Anecdotal support for this conclusion comes from the comments given by the students upon receiving feedback regarding the features present in the photographs. When the clinical sign had been identified as an alternate feature, the students expressed surprised at being told that the sign was the correct feature. For example, when the students had identified the facial edema for case 13V, they expressed initial disbelief and surprise at finding out that the clinical sign was the moon-shaped face of Cushing's disease. The overall lack of effect of order of the suggested diagnosis possibly reflects a low proportion of these chameleon features.

In terms of contributing to the order-of-information literature, this study provides mixed results. The students did attend to information presented later in the case in terms of identifying features of the subsequently suggested diagnosis. However, when features had been identified in light of the initial diagnosis, little weight seemed to be placed on the subsequently presented information and they did not reinterpret the evidence. Furthermore, they placed less weight on the latter information when reaching a diagnostic conclusion. Rather, they opted for the initially suggested diagnosis in a pattern similar to the one of the students in the corresponding conditions of Experiment 1A.

Experiment 4. Does decreasing the number of diagnoses available at test reduce the degree of incorrect diagnosis-specific feature identification?

In Experiment 3, the subsequent suggestion of a competing diagnosis was not effective in reducing feature identification errors resulting from the availability of an

incorrect provisional diagnosis. The aim of the present study was to investigate whether making competing diagnoses available at the time that students were initially suggested a diagnosis would lead to the identification being less susceptible to the influence of this suggested diagnosis. In Experiment 3, the finding that the suggestion of a subsequent diagnosis to had a small impact on diagnostic conclusions or on feature identification might have resulted from the fact that it was presented after the students had settled on one diagnosis, judged it to be sufficient, and thus discounted additional information. If competing diagnoses had been made more salient before the students began to interpret the data, an alternate interpretation of the evidence might be suggested. This would serve to reduce the degree of focus placed on the suggested diagnosis and decrease its influence on feature identification. This idea is tested in the proposed study by reducing the pool of alternatives at the time of test. The prediction is that by making the most competitive diagnoses more salient, their increase in availability will provide an alternative interpretation of the evidence, thereby reducing the influence of the suggested diagnosis on feature identification. If participants only have three diagnostic possibilities available to them, the hypothesis is that they will attend to and evaluate all three diagnoses. By comparison, students who have ten diagnostic possibilities will not be able to attend to and evaluate all of them due to memory overload. The suggestion of a diagnosis to the participants considering all ten diagnostic possibilities is likely to allow them to focus their attention to the suggested diagnosis. For the students considering three diagnostic possibilities, the suggestion of a diagnosis is less likely to focus their attention to this suggestion, as they will still be able to evaluate all three possibilities. To test this idea, naive students will be taught 10 diagnoses and be tested on a task similar to that in the previous studies of this thesis. At the time of test, half of the students will be instructed that the cases can be an example of all 10 diagnoses, while the other half will be

instructed that it can only be an example of one of three diagnoses. This paired-down pool of alternatives consists of the most confusable diagnoses for each case.

The participants who will take part in the study differ from those in the previous experiments of this thesis. The recruitment and running of medical students in the studies of this thesis is very time consuming. In addition, the availability of the medical students is limited. Therefore, the present study will be run with readily available Psychology undergraduate students. The use of these naive students requires different materials from those used to date in this thesis, in order to obtain enough cases to teach the students how to diagnose various medical conditions. Although it is difficult to obtain sufficient head-and-shoulder photographs of patients to support teaching and testing, it is relatively easy to obtain electrocardiograms of patients from instructional textbooks and from the patient files of patients. A second advantage in using electrocardiograms in the present study is that each diagnosis could be diagnosed on the basis of few features (at the most, 5 features), many of which do not need to be present in each case. This limited number of features makes it easy to record the potential features of all diagnoses (features suggested by the diagnosis but not present in the stimuli) in order to measure the potential impact of an authoritative suggestion on the identification of the features. The presence of responding to authoritative suggestions can be identified if the students identify as many potential features as depicted features.

With the use of different materials and participants, this study also serves as a test of the robustness of the findings that the availability of a diagnosis has an important influence on the interpretation of clinical signs. The hypothesis is that the interaction of a diagnosis and feature identification will not be limited to medically trained individuals nor to the visual appearance of patients, but will also generalize to individuals with virtually no medical education and different visual materials.

Method

Participants

Forty-six undergraduate Psychology students from McMaster University took part in the study in exchange for course credit and payment. None of the students had any training or experience in ECG diagnosis prior to participating in the study.

Materials

The materials consisted of training booklets detailing the basics of ECGs (placement of electrodes, basic waveform, descriptions and examples of each wave and segment), the features of 10 diagnoses (Normal, Right Ventricular Hypertrophy, Left Ventricular Hypertrophy with Strain, Right Bundle Branch Block, Left Bundle Branch Block, Acute Anterior Myocardial Infarction, Acute Lateral Myocardial Infarction, Ischemia, Pericarditis, Hyperkalemia), and teaching and test stimuli ECG strips. In addition, students received individual answer booklets in which they listed the features detected as well as their diagnostic conclusions for each case. All stimuli were presented to the students in individual booklets, and were selected from a larger sample used by Norman et al. (1999).

Procedure

During the training session, the students were taught the basics of electrocardiograms, the features seen on a normal ECG, and taught the features of 10 diagnoses, organized by chapters. In each chapter, they were shown a list of features of the diagnosis and the experimenter went over two examples, pointing out the features that were present. They were then asked to go through two examples on their own, writing down which features of the given diagnosis were present in the ECG. Feedback was given immediately after each chapter. This procedure was followed until all 10 conditions were taught.

In the test phase, students were presented with 10 new ECGs, and were asked to identify all the features present and to make a diagnostic decision. At this point, the subjects were divided into 2 test conditions: the 3 diagnosis (3dx) condition (n=24) and the 10 diagnosis (10dx) condition (n=22). In the 3dx test condition, the students received two booklets of 5 ECGs. For each booklet, they were instructed that the 5 ECGs could only be one of three possible diagnoses. In booklet 1, the possible diagnoses were Left Ventricular Hypertrophy with Strain, Hyperkalemia, and Ischemia. In booklet 2, the possible diagnoses were Right Ventricular Hypertrophy, Right Bundle Branch Block and Acute Anterior Myocardial Infarction. In the 10dx test condition, the students received the same 10 ECGs in one booklet, and were told that any of the 10 diagnoses was possible for each ECG.

The students were suggested a diagnosis (either the correct one or a plausible alternate one) prior to listing the features observed in the EGCs. They were also asked to rate how confident they were (on a scale of 0-100%) that the suggested diagnosis was the correct one. In addition, they were asked to write down and rate any other self-generated diagnosis, as well as the residual category ("something else"). They were instructed that the sum of the ratings given to the suggested diagnosis, to the self-generated diagnosis (es), as well as to the residual category should sum to 100%. See Appendices E, F, G for examples of the description of the rules of the 10 diagnoses, of the ECGs used in the study, as well as one of the answer sheets.

The students were divided into two counterbalancing conditions, with the correct diagnosis being suggested for half of the ECGs and the plausible alternate diagnosis being suggested on the other half. The order of test ECGs was randomized and held constant for all participants in each of the test conditions.

<u>Measures</u>

The percent of cases in which the students gave the highest likelihood ratings to the correct diagnosis and the percent of cases in which they gave the highest likelihood rating to the alternate diagnosis were submitted to separate 2x2x10 mixed-design ANOVAs, with the test condition (3dx, 10dx) and the diagnosis (correct, alternate) as between-subject variables and the ECGs as a repeated measures variable.

The percent of the depicted correct, alternate and irrelevant features as well as the percent of potential correct, alternate and irrelevant features listed by the students were calculated and submitted to separate mixed-design ANOVAs, with the test condition (3dx, 10dx) and the diagnosis (correct, alternate) as the between-subject variables and the ECGs as a repeated measures variable. The potential features were features that were suggested by the diagnoses, but were not present in the ECGs. Their identification would result from a misinterpretation of a normal variant (identifying an elevated ST segment when the segment was not the required 1 mm above the pre QRS level).

Results

The first hypothesis was that the suggestion of a diagnosis would be sufficient to bias students either towards the correct or the alternate diagnosis. The finding of a main effect of diagnosis for both the percent of cases the students opted for the correct diagnosis, $F(1, 42)= 96.98$, $\underline{MSE}=.154$, $\underline{p}<.01$, and the alternate one, $F(1, 42)= 89.77$, $\underline{MSE}=.191$, $\underline{p}<.01$ confirm this hypothesis. Table 11 shows the diagnostic conclusions of the students, as a function of the suggested diagnosis and of the number of alternatives available at the time of test. When biased towards the correct diagnosis, the students opted for it in a higher percent of cases than when biased towards the alternate diagnosis (84% vs. 48%, p<.01). Conversely, they opted for the alternate diagnosis in a higher percent of cases when biased towards the alternate diagnosis (46%) than when biased

towards the correct one (7.4%). The effect of the diagnosis was variable depending on the ECGs, as evidenced by a significant case by diagnosis interaction for the analysis of the correct, $F(9, 378)= 3.15$, $\underline{MSE}=.132$, $\underline{p}<.01$, and alternate conclusions, $F(9, 378)= 3.66$, $\underline{MSE}=.120$, $\underline{p}<.01$. The suggestion of a diagnosis had the largest impact on the diagnostic conclusions for ECGs 8 and 10. The number of diagnoses available at test did not impact on the percent of cases in which they opted for the correct diagnosis, $F(1, 42)= 1.10$, $\underline{MSE}=.154$, $\underline{p}= .30$, or on the percent of cases in which they opted for the alternate diagnosis, $F(1, 42)= .005$, $\underline{MSE}=.191$, $\underline{p}=.94$.

Feature Detection

Depicted Features

Consistent with the previous studies of this thesis, one of the hypotheses was that the suggestion of a diagnosis would also have an impact on the features identified by the students. This was confirmed by the finding that the suggestion of a diagnosis had an impact on the percent of depicted features identified by the students (see Table 12). There was a main effect of diagnosis on the percent of correct, $F(1, 42)= 20.26$, $\underline{MSE}=.124$, $\underline{p}<.01$ and alternate features, $F(1, 42)= 18.27$, $\underline{MSE}=.257$, $\underline{p}<.01$, identified by the students, but not on the percent of irrelevant features identified, $F(1, 42)= .070$, $\underline{MSE}=.085$, $\underline{p}=.79$. Students identified more correct depicted features when suggested the correct diagnosis than when suggested the alternate diagnosis (84% vs. 69%). Conversely, students who were biased towards the alternate diagnosis identified more alternate features than did the students biased towards the correct diagnosis (76% vs. 55%).

The number of diagnoses available at the time of test did not influence the percent of correct, $F(1, 42)= .530$, $\underline{MSE}=.124$, $\underline{p}=.47$, or the percent of alternate features, $F(1,$

42)= 1.47, <u>MSE</u>=.257, p=.23, but did influence the percent of irrelevant features identified, F(1, 42)= 13.30, <u>MSE</u>=.085, p<.01. Compared to the students who had all 10 diagnoses available at the time of test, the students who only had to consider 3 diagnoses at the time of test identified fewer irrelevant features (6.3% vs. 16.3%).

A surprising finding was the lack of significant diagnosis by test condition interaction for the identification of the correct, F(1, 42)= .000, <u>MSE</u>=.124, p=.99, alternate, F(1, 42)= .080, <u>MSE</u>=.257, p=.78, and irrelevant, F(1, 42)= .126, <u>MSE</u>=.085, p=.72, features. As can be observed in Table 12, the influence of the diagnostic suggestion was the same, regardless of whether the students had 3 or 10 diagnoses available at the time of test.

There was some variability in the detectability of the features in the various ECGs, as evidenced by significant main effects of ECGs on the percent of correct, F(9, 378)= 10.34, <u>MSE</u>=.092, p<.01, alternate, F(8, 333)= 11.91, <u>MSE</u>=.128, p<.01, and irrelevant features, F(9, 378)= 6.98, <u>MSE</u>=.035, p<.01, identified by the students. The ECGs interacted with the test condition (3dx vs. 10dx) available at the time of test for the identification of the correct features, F(9, 378)= 8.67, <u>MSE</u>=.092, p<.01, alternate features, F(8, 333)= 4.29, <u>MSE</u>=.128, p<.01, and irrelevant features, F(9, 378)= 2.25, <u>MSE</u>=.035, p<.05, and with the diagnosis for the identification of the alternate features, F(8, 333)= 2.13, <u>MSE</u>=.128, p<.05,

Potential features

The aim of the present study was also to investigate whether the suggestion of a diagnosis or the number of diagnoses available at the time of test would have an impact on the identification of features that were consistent with the possible diagnoses but not present in the ECGs, features that consisted of misidentifications of normal variations.

The results of the ANOVAs reveal a main effect of diagnosis for the identification of the alternate potential features, $F(1, 42)= 15.56$, $MSE=.147$, $p<.01$, but not for the correct, $F(1, 42)= .671$, $MSE=.147$, $p=.42$, nor for the irrelevant potential features, $F(1, 42)= .033$, $MSE=.029$, $p=.86$. Overall, the students who were biased towards the alternate diagnosis identified more alternate potential features than did those biased towards the correct diagnosis (29% vs. 13%)

The number of diagnoses that the students had available to them at the time of test did not have a significant effect on the identification of either the correct, $F(1, 42)= .148$, $MSE=.147$, $p=.70$, alternate, $F(1, 42)= .073$, $MSE=.147$, $p=.79$, or irrelevant features, $F(1, 42)= .008$, $MSE=.030$, $p=.93$. The test condition did not interact significantly with the diagnosis on the identification of the correct, $F(1, 42)= 1.41$, $MSE=.147$, $p=.24$, and alternate features, $F(1, 42)= 1.26$, $MSE=.147$, $p=.27$, but showed a significant interaction on the identification of the irrelevant features, $F(1, 42)= 4.94$, $MSE=.029$, $p< .05$. The influence of the diagnosis on the identification of the potential correct and alternate features was the same, regardless of whether the students were considering 10 or 3 diagnoses at the time of test.

There were main effects of ECGs for the identification of the correct, $F(3, 126)= 24.52$, $MSE=.131$, $p<.01$, alternate, $F(7, 294) = 2.45$, $MSE=.122$, $p<.05$, and irrelevant features, $F(9, 378)= 3.46$, $MSE=.025$, $p<.01$. There was a significant ECGs by test condition interaction on the identification of alternate features, $F(7, 294)= 3.39$, $MSE=.122$, $p<.01$.

Discussion

A surprising finding of this study was that the number of diagnoses available to the students at the time of test did not have an impact on their identification of the correct and alternate features, nor did it interact with the suggestion of a diagnosis. The initial

hypothesis was that the students would be less susceptible to being biased by a diagnostic suggestion when only considering 3 possibilities, as the competing alternatives were more likely to be evaluated and to provide an alternative interpretation to the clinical data. A possible explanation for the lack of an effect of the number of diagnoses available at the time of test lies in the nature of the diagnoses that were grouped together for each booklet. In both booklets 1 and 2, the diagnoses were the ones that had the most overlapping features, the ones most confusable with each other. Thus, even though the number of diagnoses available to the students was reduced to three diagnoses, the two alternate diagnoses available to them were the ones that would have been the source of confusion even had they had all 10 diagnoses to consider. Support for this explanation comes from the finding that few irrelevant features were identified by the students evaluating the 10 alternatives. These irrelevant features would be features of a diagnosis other than the correct or the alternate one. If the students identified as many of these features as of the features of the correct or alternate diagnosis, this would indicate that they were considering other diagnoses. Overall, the subjects only identified 11.2% of the depicted irrelevant features when they had all 10 possibilities available to them. While this amount is larger than the amount of irrelevant features identified when the students were only evaluating 3 possibilities, it is less than the percent of correct and alternate features identified. This suggests that they were already focusing on the two most likely diagnoses rather than on the other alternatives. Perhaps there is something inherent in the materials of the present study that rendered it relatively easy for the students to focus on the most confusable diagnoses. It would be of interest to run the same study with different materials from an area of medicine in which each diagnostic presentation has more than one or two confusable diagnoses. The extension of the present findings to other areas of medicine would strengthen the generalizability of the finding that the number of

diagnostic alternative available upon diagnosis has no impact on the influence of a suggested diagnosis on feature interpretation.

Despite the failure to observe an impact of the number of diagnostic alternatives available at the time of test, this study does extend the generalizability of the finding that a suggested diagnosis has an effect on the features identified by clinicians. This effect on feature interpretation was observed with different materials and different participants than those used in the previous studies. Thus, the impact of a diagnosis on feature interpretation is not limited to the features visible from the appearance of a patient, nor is it observed solely in individuals who have received extensive training in medicine. Extreme novices are as susceptible to the suggestion of a diagnosis as are residents.

Chapter 6: General Discussion

Although difficulties in the identification of features from patients' appearance are under-emphasized by researchers interested in describing the decision-making process of clinicians and by those interested in describing the mental organization of medical knowledge, numerous studies demonstrate that features are ambiguous and that their identification is variable. Diagnostic hypotheses have a significant impact on the identification of features from radiographs, ECG strips, and the appearance of patients. An important contribution of the present studies is that they allow for the identification of the processes underlying the interaction of the diagnostic hypotheses and the identification of features. The suggestion of a diagnosis serves both to guide attention to features relevant to the diagnosis and to change the identification given to the physical characteristics of a patient's appearance. When the diagnostic hypothesis is the correct one, it leads to an increase in the accuracy of identifying the correct features. For example, students in the present studies were more likely to identify the superclavicular lymph node from the gentleman with stomach cancer when that diagnosis was evoked than when the diagnostic hypothesis was liver cancer. When the diagnostic hypothesis is incorrect, it leads clinicians to misidentify the evidence such that it supports the incorrect diagnosis. In the present studies, the students did so either by misidentifying normal variation or by mislabelling the correct feature. For example, misidentification of a characteristic due to normal variation occurred when participants identified the normally tanned skin of the gentleman with stomach cancer as jaundice as suggested by the alternate diagnosis of liver cancer. Misidentification of a correct feature occurred when

participants misidentified the swollen parotid glands of the boy with mumps as a moon-shaped face consistent with the alternate diagnosis of Cushing's disease. This susceptibility to the influence of a diagnosis when identifying features was not a consequence of limited clinical experience, as it was also observed with greater levels of expertise.

The fact that the suggested diagnosis had a strong impact on the identification of features is remarkable given that the cases selected were particularly good examples of the disease they represented. Each photograph was selected from prototypical textbook illustrations. The clinical implications of these results are important. Many of the features of the diagnoses are features that are strongly associated with the particular diagnosis (e.g. moon-shaped face = Cushing's disease and malar rash = SLE). The failure to correctly identify one of these features is likely to have a strong impact on the diagnostic accuracy of a clinician. For example, once the malar rash has been misidentified as a regular rash, the clinician is less likely to consider lupus as a diagnosis, and is more likely to consider other incorrect diagnostic hypotheses.

Although the identification of features was strongly influenced by the suggestion of a diagnosis, it was not influenced by the amount of independent evidence supporting the suggested diagnosis. In Experiments 2 to 4, various techniques (subsequent suggestion of a competing diagnosis, varying the case history accompanying the suggested diagnosis, as well as reducing the size of the pool of alternatives) were employed to manipulate the degree of focus placed on the suggested diagnosis. These manipulations had little, and in some cases, no impact on the identification of features. This finding implies that the major biasing effect on feature identification is the availability of a diagnosis, rather than the degree of focus that is placed on it. The finding that changing the credibility of a diagnosis had no impact on the identification of features

is counter-intuitive. As discussed in Chapter 3, it was hypothesized that a diagnosis with a high prior probability would result in a high prior probability of its features. From a Bayesian perspective, the final probability of a diagnosis is the result of its prior probability and of the strength of the evidence supporting it. If a diagnosis has a high prior probability, it will have a high final probability, even if there is little evidence to support it. Presumably, the same argument could be extended to the identification of features. Their likelihood of being identified would be a function of their conditional probability and of the amount of physical evidence suggesting them. If the features' prior probability was high due to the diagnosis having a high probability, little physical evidence would be necessary for these features to be identified by clinicians. However, the results of the present thesis counter such an argument.

In contrast, the diagnostic ratings of the participants were influenced both by the availability of the diagnosis and by the evidence provided with the diagnosis. It was not possible to establish the influence of both of these factors on diagnostic conclusions in Experiments 3 and 4, as the case history was confounded with the suggested diagnosis in the former study, and no case history was presented in the latter one. However, these two factors were unconfounded in Experiment 2, and both the case history and the suggestion of the diagnosis had substantial effects on diagnostic ratings. Despite the strong impact of the case history on diagnostic ratings, it had little impact on the identification of the features. In Experiment 2A, the case history accompanying the suggestion of the alternate diagnosis had an impact on the diagnostic conclusions, but had little impact on the identification of features. A similar pattern occurred in Experiment 2B, for the case histories accompanying the suggestion of the correct diagnosis. Such patterns of results indicate that feature identification is affected by the availability of the diagnosis, while the diagnostic ratings are influenced both by the availability of the diagnosis and by the

weight of the evidence in favor of the diagnosis. Furthermore, the suggestion of a diagnosis had a similar effect on both the identification of features and the diagnostic conclusions, suggesting that the impact of a suggested diagnosis might be mediated by its effects on feature identification.

Another major result of the present studies is the fact that initial feature identification is robust to reinterpretation. Once biased towards a diagnosis, and having identified features in light of this bias, the students did not reevaluate the identity provided to the physical characteristics of a patient, even when subsequently presented with a diagnosis suggesting an alternative identity for that characteristic. In other words, once they had seen and labeled a feature, they were unable to unsee it.

Finally, a third important contribution of the current studies was the finding that the influence of the suggested diagnosis was mediated by the information contained in the stimuli. While both novices and more experienced clinicians identified more features of the alternate diagnosis when it was suggested to them, they nevertheless identified more of the correct features than the alternate features. This suggests that the influence of the diagnosis on feature identification does not result from a purely top-down process, but rather from an interaction of top-down processes and the information present in the stimuli. This finding is consistent with Rumelhart and McClelland's (1981) interactive activation network of word perception and it counters an extreme version of Patel and Groen's (1986) argument that novices use backward reasoning while experts use forward reasoning. This interaction between the diagnosis being evaluated and the identification of features needs to be addressed by researchers studying medical decision-making and the organization of medical knowledge if the goal is to develop veridical accounts of the processes involved in medical diagnostic tasks.

By focusing solely on correcting errors of base-rate neglect and of improperly weighting the evidence, both the normative and descriptive views of decision theory underestimate the ambiguity of clinical signs and the interaction between the diagnosis and feature identification. While computer-aided diagnostic tools resulting from this approach are effective in reducing errors of base-rate neglect and inadequate weighting of the evidence, their effectiveness will be limited by the clinician's ability to accurately identify features from a patient's appearance. If the diagnoses generated early in the clinical case are incorrect, they will lead to the incorrect identification of the features that are given to the computer system. Without an accurate description of the patient's clinical signs, the list of diagnoses generated by the computer is less likely to contain the correct diagnosis.

In addition to the development of computer-aided diagnostic tools, proponents of decision theory espouse a second prescriptive approach of informing clinicians about the heuristics and biases that lead to diagnostic errors and of admonishing them to either consider alternative diagnoses or avoid any diagnostic hypothesis when faced with a patient (Elstein, 1999). However, as suggested by the results of Experiment 4, such a strategy will not always be effective in reducing errors. Reducing the pool of alternatives to make the most likely alternative more salient and more likely to be considered did not decrease the impact of suggesting a diagnosis on feature interpretation. Thus, rather than simply giving the general advice of generating diagnostic hypotheses or of considering alternatives, proponents of decision theory need to develop interventions aimed at increasing the likelihood that the correct diagnosis will be available to the clinician upon initial contact with a patient. Further support for this argument comes from Norman et al. (1999), who observed that students who had been asked to consider a diagnostic hypothesis prior to searching for features in ECGs showed higher accuracy in diagnostic

conclusion and identification of features, compared to students who had been instructed to search for the features prior to generating a diagnostic hypothesis. This result indicates that having a diagnosis in mind during feature interpretation is beneficial.

Similarly, any theory attempting to describe the organization of diagnostic knowledge of clinicians needs to account for the interaction between the diagnostic hypothesis and feature identification. Theories of categorization assume that the largest problem in classifying a diagnosis lies in matching the input to some representation of the category, whether it be a prototype, a prior instance or a schema. However, they typically do not address how the activation of a category could serve to change the interpretation of features. Prototypes and semantic-network theories treat features as independent information that must be combined and matched with stored knowledge regarding medical categories. Most of these theories make the implicit assumption that the diagnostic task is a process in which information is treated in a bottom-up fashion. By failing to address the interaction between knowledge of diagnostic categories and the identification of features, they cannot fully account for the organization and application of the clinicians' mental organization of medical knowledge.

Proponents of instance theories have addressed the interaction of diagnostic hypotheses and feature identification in their explanation of the organization of medical knowledge. When clinicians encounter a new case, it is compared with the most similar prior instance in memory. This instance, by having already been categorized, will suggest the diagnosis. In addition, the similarity between the prior instance and the new case will help clinicians resolve any ambiguity in the features by suggesting an interpretation. If the most similar prior instance corresponds to the correct diagnosis, it will lead to an accurate diagnostic decision and feature identification. If the instance corresponds to an incorrect diagnosis, it will have the opposite effect. Thus, identification of the features

suggested by the prior instance will depend on the example recalled. This prediction was confirmed by Allen et al. (1992) and Brooks et al. (1991) who found that both the diagnostic conclusion and identification of features from dermatological slides were influenced by their similarity to prior cases seen in practice.

The generalizability of the results in the present studies is limited by a number of factors. One of the potential concerns with this line of research lies in the use of photographs rather than real patients, leading this research open to the argument that clinicians rarely work with such constrained and limited information. Typically, clinicians gather significant information about the clinical history and are able to examine the patients from more than one angle or as they move around. However, the use of photographs in medical education is quite common. Clinicians maintain libraries of patient photographs for instruction and communication with colleagues and students. In addition, textbooks contain photographs of patients as visual examples of the verbal description of the features. The photographs used in the present studies were prototypical examples taken from medical textbooks. Presumably, their inclusion as teaching materials indicates that they were considered by the authors of the textbooks to be particularly good examples of the diagnosis they represented.

A second limitation to the generalizability of the current findings is that all the effects were observed with visual stimuli. According to anecdotal evidence provided by clinicians, the evidence most strongly weighted in reaching a diagnosis is gathered from the case history, leading to the suggestion that misidentifications of visual features would have relatively little impact on the final diagnostic decision. The clinical signs have a potentially large impact on the final decision because they are perceived very early in the encounter with the patient, often prior to the gathering of any formal clinical history.

Researchers have demonstrated that clinicians make active interpretations of the clinical data as soon as it is presented (Gale & Marsden, 1982) and generate diagnostic hypotheses very early during the clinical encounter (Neufeld et al., 1981). Thus, the interpretation of the visual signs could be very important. If an incorrect diagnosis were to be generated on the basis of an early misidentification of the features of the patient, it would then misguide the search for further information from the clinical history and laboratory tests. Experimentally, this hypothesis could be tested in a study similar to the studies in this thesis, with the modification of asking clinicians to indicate what additional information they would like to receive regarding the patient. If, as hypothesized, the initial identification of the features has an impact on the subsequent gathering of data, those clinicians who misidentify the features should seek new information that is consistent with the initial misidentification.

The argument that clinicians rely strongly on the information contained in the case history provides the impetus to develop similar studies to investigate whether a provisional diagnosis would influence the identification of the information reported verbally by the patients. Typically, patients do not present their complaints in an interpreted form and one of the clinician's skills is to identify the finding. For example, when complaining of a belly ache, patients will rarely report that they are experiencing an acute upper left quadrant abdominal pain. Rather, the clinician must provide this interpretation to the patient's complaint that it hurts "here" and that it started "a couple of hours ago". It is possible that a suggested diagnosis would have a similar influence on the identification of features from a patient's verbal reports as it has on the identification of visual features.

Educational and Clinical Implications

In addition to demonstrating a phenomenon that needs to be accounted for in the study of medical decision making and medical knowledge, the results in the present studies provide interesting challenges to medical educators. The first challenge lies in designing and developing studies to test interventions applicable during the clinical encounter to help students reevaluate the evidence when they initially considered an incorrect diagnosis. The second challenge lies in developing a curriculum that will lead to optimal learning of the medical knowledge and skills in correctly identifying features.

Based on the results of the present studies, it appears unlikely that interventions aimed at reducing the degree of focus placed on a suggested diagnosis will be effective in reducing erroneous feature identification resulting from considering an incorrect diagnosis. Attempts to reduce the credibility of the diagnosis by presenting it with an inconsistent case history or by making plausible alternatives more salient did not lead a reduction of the influence of the diagnosis on the identification of features. Similarly, even the direct suggestion of a strongly competitive diagnosis did not lead to the re-evaluation of the data. A possible explanation for the ineffectiveness of these manipulations in changing the influence of the diagnosis might be that they did not require active reinterpretation on the part of the participants. The suggested diagnosis, by being the most available alternative, was most likely selectively tested, leading to any other diagnosis being discounted because of the sufficiency of the evidence in favor of the initial diagnosis. It remains to be seen whether an intervention requiring students (or trainees) to actively test alternative diagnoses might be effective in reducing the influence of a suggested diagnosis on feature identification. Such interventions could be in the form of instructing students to argue against the suggested diagnosis, to argue for alternative diagnoses or to argue for different interpretations of the features identified. In studies of

social decision making, it has been shown that when participants were forced to consider the alternatives, either by arguing for and against each potential outcome, they were less susceptible to being biased towards one outcome (Arkes, 1981; Slovic & Fischoff, 1977).

Alternatively, instructional interventions aimed at increasing the likelihood that students will generate the correct diagnostic hypothesis early in the encounter with a patient might be effective in reducing diagnostic errors. Neufeld et al. (1981) and Norman et al. (1994) have demonstrated that novices and experts generate diagnostic hypotheses early in a case and then proceed to test these hypotheses. The largest difference in performance was the accuracy of the diagnostic hypotheses. There are a number of instructional methods that might lead to increased learning of medical knowledge, thus increasing the likelihood of generating the correct diagnosis upon initial contact with a patient.

Presently, medical education emphasizes learning the underlying disease processes of a diagnosis and verbal lists of features. Little emphasis is placed on recognizing the examples of a disease in varied forms and contexts in which they can occur. If medical education increases the emphasis placed on learning the variability of the presentation of visual features, their identification may be more likely to trigger the correct diagnosis when seeing a new patient. One way of achieving this would be to increase the amount of mixed practice available to students. Schmidt and Bjork (1992) reported a study in which variation at the time of practice led to increased performance at the time of test. They cite a study by Nitsch, in which students were asked to learn novel concept-words such as "crinch" that was defined as offending someone. The subjects either learned these words in a constant context (all in a restaurant setting) or in a variable context (numerous settings). When asked to recognize novel examples of the concept, the subjects who had learned the words under conditions of variable context performed better

than did the ones who learned the words in a constant context. Although Nitch's study was conducted in the field of recognition memory, its results suggest a potential method to increase the likelihood that students will recognize a new case as an example of the correct diagnosis. Schmidt and Bjork (1992) argue that the increase in performance was the result of systematically altering practice that encouraged additional or systematic information processing activities which lead to better recall of the relevant information at the time of test. They further argued that an important principle underlying this finding was the overlap between the processes necessary for performance at the test and the processes practiced during acquisition. If some of the conditions at the time of acquisition force the learner to engage in processes that are critical at the time of test, then they will be effective for learning.

This argument has implications for the education of feature identification for students. Rather than teach students to recognize examples of a disease within the context of a specific diagnostic category, their instruction should occur in a manner that will require them to use the processes necessary for diagnosing new patients. In the clinical setting, clinicians do not know the true diagnosis, and must therefore recognize the new case as being yet another variation of the disease. According to Schmidt and Bjork's (1992) argument, optimal identification of diseases in the clinical setting will occur if students practice the identification of examples of the diseases out of context of the instructional chapters during training. In support for this idea, Hatala (unpublished results) has shown an advantage for mixed practice when diagnosing new cases. She taught six ECG diagnoses to novice medical students, organized by diagnoses. One group then practiced identifying features and diagnosing new cases in a blocked fashion, such that they practiced with 6 examples of one diagnosis before moving on to 6 examples of the next diagnosis, and so on. The other group practiced with the same examples, but saw

them in a mixed fashion. Accuracy in diagnosing new cases was higher for the group of students who had received mixed practice. The explanation for this higher accuracy for the mixed-practice group was that this type of practice led them to attend to and learn the characteristics important in diagnosing cases in clinical practice.

Implementing mixed practice for diagnosing various clinical conditions into the curriculum might be insufficient to reduce errors in feature identification if the students are not exposed to the range of variation in the presentation of a diagnosis. In Experiments 1B and 2B, residents in family medicine showed a susceptibility to being influenced by the diagnosis when identifying features. Having completed a minimum of six months of training, they had had much practice at recognizing features from patients in other contexts than the suggested diagnosis. This result suggested that mere practice at recognizing features from patients is not sufficient to reduce the susceptibility of their feature identification to a suggested diagnosis. Rather, the organization of the knowledge or their degree of exposure to the variability in clinical presentations might play in important role.

Although prototypical instruction might be beneficial for the development of an organized representation of diagnostic categories, it might also lead to an absence of exposure to the variability of disease presentation. Norman, Brooks and their colleagues (1989) have shown that previous occurrences of diseases can increase the availability of those diseases, which could influence the initial generation of hypotheses. When encountering a new case, clinicians will often opt for a particular diagnosis simply because it bears resemblance to a prior case of that particular diagnosis. If, in medical education, care was taken to teach students cases that covered the range of variability in the presentation of diseases, they would be more likely to recall a prior instance of the correct diagnosis when encountering a new case. They would be more likely to have seen

a similar case before and its recall would help in the identification of the features and in determining its diagnostic category. In such an educational system, the aim would be to take advantage of the fact that the diagnosis considered influences feature identification and the fact that clinicians generate hypotheses early in the encounter with patients, rather than attempt to render feature identification independent of the diagnosis.

A third way of manipulating the structure of the information taught to students is to place the emphasis on the features that distinguish one disease from another, rather than on the likelihood of features given a particular disease. The latter approach, called the independent form, is typical of medical textbooks, in which students are trained to learn the features associated with a particular disease. The former approach, called the contrastive form, is one in which students learn how a particular disease differs from other diseases that are frequent alternatives for the same pattern of features. To test the efficacy of each type of learning, Klayman and Brown (1993) taught students to recognize cases of two fictitious diseases. One group of students learned the categories separately (independent form) while the other group learned the categories in parallel (contrastive form). The argument was that this latter form of instruction would allow them to easily identify the differences between the two categories. At test, the students who received the contrastive training showed more sensitivity to the features that discriminated between the two categories and demonstrated more diagnostic accuracy than the students who received the independent training. These results sugest that training that focuses on learning the features that distinguish diagnostic categories from one another, rather than simply focusing on the features typical of one diagnosis, lead to higher diagnostic accuracy and to a higher sensitivity to the features crucial to discriminating between the categories. Further evidence for the advantage of teaching by using contrasting cases was reported by Avrahami et al. (1997). When asked to teach

categories to naive participants, their subjects produced sequences consisting of several ideal positive cases of the category, followed by ideal negative cases and then borderline cases. The presence of negative cases early in the instruction sequence suggests that optimal learning of categories will occur when they are placed in contrast with each other. Such a structure will allow students to learn not only the features that are typical of a category, but also the features that are diagnostic of a category.

Another educational structure which might increase the likelihood of students generating the correct diagnostic hypothesis in a clinical encounter is to train them to generate abstracted representations early in the encounter. This is the argument put forward by Bordage (1999), based on his theory of semantic axes. In his theory, semantic qualifiers consist of abstractions of the features along oppositional axes (acute versus chronic presentation). These qualifiers can be used to generate early problem representations, which in turn can serve to guide the search for relevant diagnostic possibilities (e.g. diseases with acute presentations versus diseases with chronic presentations). Bordage and his colleagues (Chang et al., 1998) have shown that clinicians who generated the correct diagnostic hypotheses showed a greater use of semantic qualifiers than did clinicians who generated incorrect diagnostic hypotheses. Thus, teaching students to represent clinical problems along these semantic qualifiers is likely to increase their diagnostic accuracy.

Regardless of the proper method of instruction, researchers and medical educators need to take into account the interaction between the available diagnosis and the identification of features. As shown in the present studies, the evaluation of a diagnostic hypothesis has a strong impact on the accuracy of feature identification. This impact is strong enough to lead clinicians to identify features that are not present in the medical data if they are considering an incorrect diagnosis. Although the independent evidence in

favor of a diagnosis will have a large impact on the probability of concluding for the correct diagnosis, it has relatively little impact on the identification of features. Rather, the identification of features appears to be influenced by the mere availability of that diagnosis. The availability of the diagnosis has a similar impact on diagnostic ratings and on feature identification, suggesting that its effect on diagnostic conclusions might be mediated by its effects on feature identification. Many descriptions of the medical decision task, as well as many educational approaches, place most of the emphasis on getting clinicians to appropriately put together the clinical signs. In doing so, they ignore an important process underlying the medical decision task, namely the interaction between diagnostic conclusions and feature interpretation.

References

Allen S. W., Norman, G. R., & Brooks, L. R. (1992). Experimental studies of learning dermatological diagnosis: The impact of examples. Teaching and Learning in Medicine, 4(1), 35-44.

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. Journal of Consulting and Clinical Psychology, 49(3), 323-330.

Avrahami, J., Kareev, Y., Bogot, Y., Caspi, R., Dunaevsky, S., & Lerner, S. (1997). Learning by examples: Implications for the process of category acquisition. The Quarterly Journal of Experimental Psychology, 50A(3), 586-606.

Barnett, G. O., Cimino, J. J., Hupp, J. A., & Hoffer, E. P. (1987). Dxplain: An evolving diagnostic decision-support system. Journal of the American Medical Association, 258(1), 67-74.

Berbaum, K. S., Franken, E. A., Dorfman, D. D., Barloon, T., Ell, S. R., Lu, C. H., Smith, W., & Abu-Yousef, M. M. (1986). Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Investigative Radiology, 21, 532-539.

Berbaum, K. S., El-Khoury, G. Y., Franken, E. A., Kathol, M., Montgomery, W. J., & Hesson, W. (1988). Impact of clinical history on fracture detection with radiography. Radiology, 168(2), 507-511.

Bergus, G. R., Chapman, G. B., Gjerde, C., & Elstein, A. S. (1995). Clinical reasoning about symptoms despite preexisting disease: Sources of errors and order effects. Family Medicine, 27, 314-320.

Berner, E. S., Webster, G. D., Shugerman, A. A., Jackson, J. R., Algina, J., Baker, A. L., Ball, E. V., Cobbs, C. G., Dennis, V. W., Frenkel, E. P., Hudson, L. D., Mancall,

E. L., Rackley, C. E., & Tauton, O. D. (1994). Performance of four computer-based diagnostic systems. The New England Journal of Medicine, 330(25), 1792-1796.

Bordage, G. (1999). Why did I miss the diagnosis? Some cognitive explanations and educational implications. Academic Medicine, 74(10, Suppl), S138-S143.

Bordage, G., & Lemieux, M. (1991) Semantic structures and diagnostic thinking in experts and novices. Academic Medicine, 66(9, Suppl), S70-S72.

Bordage, G., & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: Categories and prototypes. Medical Education, 18, 406-416.

Boshuizen, H.P.A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. Cognitive Science, 16, 153-184.

Brooks, L. R. (1990). Concept formation and particularizing learning. In P. Hanson & S. Davis (Eds.), Vancouver Studies in Cognitive Science, 1, (pp.141-171). Vancouver, BC: University of British Columbia Press.

Brooks, L. R., LeBlanc, V, R., & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. Psychological Science, 11(2), 112-117.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. Journal of Experimental Psychology: General, 120(3), 278-287.

Chang, R. W., Bordage, G., & Connell, K. J. (1998). The importance of early problem representation during case presentations. Academic Medicine, 73(10, Suppl), S109-S111.

Cunnington, J. P. W., Turnbull, J. M., Regehr, G., Marriott, M., & Norman, G. R. (1997). The effects of presentation order in clinical decision making. Academic Medicine, 72(Suppl.), S40-S42.

Custers, E. J. F. M., Regehr, G., & Norman, G. R. (1996). Mental Representations of Medical Diagnostic Knowledge: A Review. Academic Medicine, 71(10, Suppl), S55-S61.

Detmer, D. E., Fryback, D. G., & Gassner, K. (1978). Heuristics and biases in medical decision-making. Journal of Medical Education, 53, 682-683.

Elstein, A. S. (1999). Heuristics and biases: Selected errors in clinical reasoning. Academic Medicine, 74(7), 791-794.

Fauci, A. S., Braunwald, E., Isselbacher, K. J., Wilson, J. D., Martin, J. B., Kasper, D. L., Hauser, S. L., & Longo, D. L. (Eds.). (1998). Harrison's Principles of Internal Medicine (14th ed.). New-York: McGraw-Hill

Friedman, C. P., Elstein, A. S., Wolf, F. M., Murphy, G. C., Franz, T. M., Heckerling, P. S., Fine, P. L., Miller, T. M., & Abraham, V. (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation: A multisite study of 2 systems. Journal of American Medical Association, 282(19), 1851-1856.

Fischoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. Psychological Review, 90(3), 239-259.

Gale, J., & Marsden, P. (1982). Clinical problem solving: The beginning of the process. Medical Education, 16, 22-26.

Hatala, R., Norman, G. R., & Brooks, L. R. (1998). The impact on a clinical scenario upon accuracy of electrocardiogram interpretation. Manuscript submitted for publication.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 16, 321-338.

Heller, R. F., Saltztein, H. D., & Caspe, W. B. (1992). Heuristics in medical and non-medical decision-making. The Quarterly Journal of Experimental Psychology, 44A(2), 211-235.

Klayman, J. (1995). Varieties of confirmation bias. The Psychology of Learning and motivation, 32, 385-418.

Klayman, J., & Brown, K., (1993). Debias the environment instead of the judge: An alternative approach to reducing error in diagnostic (and other) judgment. Cognition, 49, 97-122.

Komatsu, L. K. (1992). Recent views of conceptual structure. Psychological Bulletin, 112(3), 500-526.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1981). Models of competence in solving physics problems. Cognitive Science, 4, 317-345.

LeBlanc, V. R., Cunnington, J., Brooks, L. R., & Norman, G. R. (April, 1999). Reinterpreting the Classics: The Etiology of Errors in the Interpretation of Obvious Features. Structured poster presented at the annual meeting of the American Educational Research Association, Montreal; PQ.

Lemieux, M., & Bordage, G. (1992). Propositional versus structural semantic analyses of medical diagnostic thinking. Cognitive Science, 16, 185-204.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effect in letter perception: Part 1. An account of basic findings. Psychological Review, 88(5), 375-407.

McGaghie, W. C., Boerger, R. L., McCrimmon, D. R., & Ravitch, M. M. (1994). Agreement among experts about the structure of concepts in pulmonary physiology. Academic Medicine, 69(10, Suppl.), S78-S80.

Medin, D. L. (1989). Concepts and Conceptual Structure. American Psychologist, 44(12), 1469-1481.

Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. Cognitive Psychology, 12, 496-522.

Myles-Worsley, M., Johnston, W. A., & Simon, M. A. (1988). The influence of expertise on x-ray image processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(3), 553-557.

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. The Quarterly Journal of Experimental Psychology, 46A(4), 759-778.

Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: A cross-sectional and longitudinal analysis. Medical Education, 15, 315-322.

Norman, G. R., Brooks, L. R., Coblentz, C. L., & Babcook, C. J. (1992). The correlation of feature identification and category judgments in diagnostic radiology. Memory and Cognition, 20(4), 344-355.

Norman, G. R., Brooks, L. R., Colle, C., & Hatala, R. (1999). The benefit of diagnostic hypotheses in clinical reasoning: Experimental study of an instructional intervention for forward and backward reasoning. Cognition and Instruction, 17(4), 433-448.

Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J.
(1989). The development of expertise in dermatology. Archives of Dermatology, 125,
1063-1068.

Norman, G. R., Trott, A. D., Brooks, L. R., & Smith, E.K.M. (1994). Cognitive
differences in clinical reasoning related to postgraduate training. Teaching and Learning
in Medicine, 6(2), 114-120.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in
categorization, identification, and recognition. Journal of Experimental Psychology:
Learning, Memory, and Cognition, 15(2), 282-304.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in
medical reasoning. Cognitive Science, 10, 91-116.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. Journal of
Experimental Psychology, 77(3, pt 1), 353-365.

Regehr, G., & Norman, G. R. (1996). Issues in cognitive psychology:
Implications for professional education. Academic Medicine, 71(9), 988-1001.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of
stimulus material. Journal of Experimental Psychology, 81, 274-280.

Rosch, E. (1977). Human Categorization. In N. Warren (Ed.), Studies in Cross
Cultural Psychology, 1, (pp.3-49). London: Academic Press.

Rosch, E., & Mervis, C. R. (1975). Family Resemblance: Studies in the internal
structure of categories. Cognitive Psychology, 7, 573-605.

Rumelhart, D. E. (1985). Towards an interactive model of reading. Theoretical
Models and Processes of Reading(H. singer, Ed.)., International Reading Association.

Sanbonmatsu, D. M., Posovac, S. S., Kardes, F. R., & Mantel, S. P. (1998).
Selective hypothesis testing. Psychonomic Bulletin and Review, 5(2), 197-220.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. Memory and Cognition, 21(3), 338-351.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3(4), 207-217.

Shortliffe, E. H. (1987). Computer programs to support clinical decision making. Journal of the American Medical Association, 258(1), 61-66.

Slovic, P., & Fischoff, B., (1977). On the psychology of experimental surprises. Journal of Experimental Psychology: Human Perception and Performance, 3, 544-551.

Trope, Y & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational mechanisms. In E.T. Higgins & A. W. Kruglanski (Eds.), Social Psychology: Handbook of Basic Principles, (pp. 239-270). New York: Guildford Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. Science, 185, 1124-1131.

Wallsten, T. S. (1981). Physician and medical student bias in evaluation diagnostic information. Medical Decision Making, 1(2), 145-164.

Whittlesea, B. W., & Brooks, L. R. (1988). Critical influence of particular experiences in the perception of letter, words, and phrases. Memory & Cognition, 16(5), 387-399.

Table 1. Experiment 1A. Initial demonstration: The mean percent of cases in which the students opted for the correct and alternate diagnoses.

	Bias	
	Correct History and Diagnosis	Alternate History and Diagnosis
Mean percent of cases:		
Opted for correct dx.	80.0% (3.6%)	6.9% (2.6%)
Opted for alternate dx.	3.1% (1.9%)	66.9% (5.0%)

*numbers in parentheses represent the standard error

Table 2. Experiment 1A. Initial demonstration: The mean percent of features of the correct and alternate diagnoses identified by the students.

	Bias	
	Correct History and Diagnosis	Alternate History and Diagnosis
Mean percent of features:		
Consistent with correct dx.	48.8% (3.9%)	35.1% (4.3%)
Consistent with alternate dx.	5.7% (1.5%)	26.5% (2.8%)

*numbers in parentheses represent the standard error

Table 3. Experiment 1B. Initial demonstration: The mean percent of cases in which the residents opted for the correct and alternate diagnoses.

	Bias	
	Correct History and Diagnosis	Alternate History and Diagnosis
Mean percent of cases:		
Opted for correct dx.	80.6% (4.1%)	12.5% (3.5%)
Opted for alternate dx.	0% (0%)	69.4% (4.7%)

*numbers in parentheses represent the standard error

Table 4. Experiment 1B. Initial demonstration: The mean percent of features of the correct and alternate diagnoses identified by the residents

	Bias	
	Correct History and Diagnosis	Alternate History and Diagnosis
Mean percent of features:		
Consistent with correct dx.	52.7% (3.9%)	40.0% (4.4%)
Consistent with alternate dx.	8.4% (1.6%)	23.8% (2.4%)

*numbers in parentheses represent the standard error

Table 5. Experiment 2A. Credibility study: The mean percent of cases in which the students opted for the correct and alternate diagnoses.

	Condition			
	Correct Dx. Correct Hx.	Correct Dx. Alternate Hx.	Alternate Dx. Correct Hx.	Alternate Dx. Alternate Hx.
Mean percent of cases:				
Opted for correct dx.	68.8% (4.6)	41.3% (4.6)	23.7% (4.6)	1.9% (4.6)
Opted for alternate dx.	2.5% (4.8)	10.4% (5.0)	29.4% (4.8)	71.3% (4.8)

*numbers in parentheses represent the standard error

Table 6. Experiment 2A. Credibility study: The mean percent of features of the correct

and alternate diagnoses identified by the students.

	Condition			
	Correct Dx. Correct Hx.	Correct Dx. Alternate Hx.	Alternate Dx. Correct Hx.	Alternate Dx. Alternate Hx.
Mean percent of features:				
Consistent with correct dx.	53.7% (3.0)	42.5% (3.0)	41.7% (3.0)	41.2% (3.0)
Consistent with alternate dx.	12.7% (2.5)	18.1% (2.5)	25.2% (2.5)	27.5% (2.5)

*numbers in parentheses represent the standard error

Table 7. Experiment 2B. Credibility study: The mean percent of cases in which the residents opted for the correct and alternate diagnoses.

	Condition			
	Correct Dx. Correct Hx.	Correct Dx. Alternate Hx.	Alternate Dx. Correct Hx.	Alternate Dx. Alternate Hx.
Mean percent of cases:				
Opted for correct dx.	75.5% (6.6)	46.9% (6.6)	26.5% (6.6)	6.2% (6.6)
Opted for alternate dx.	9.3% (7.6)	18.8% (7.6)	35.7% (7.0)	71.0% (7.0)

*numbers in parentheses represent the standard error

Table 8. Experiment 2B. Credibility study: The mean percent of features of the correct and alternate diagnoses identified by the residents.

	Condition			
	Correct Dx. Correct Hx.	Correct Dx. Alternate Hx.	Alternate Dx. Correct Hx.	Alternate Dx. Alternate Hx.
Mean percent of features:				
Consistent with correct dx.	45.2% (3.6)	45.2% (3.6)	39.9% (3.6)	32.7% (3.6)
Consistent with alternate dx.	12.1% (3.0)	14.8% (3.0)	21.9% (3.0)	35.1% (3.0)

*numbers in parentheses represent the standard error

Table 9. Experiment 3. Impact of a competing diagnosis: The mean percent of cases in which the students opted for the correct and alternate diagnoses.

	Bias	
	Correct Hx. & Dx First Alternate Dx Second	Alternate Hx. & Dx First Correct Dx Second
Mean percent of cases:		
Opted for correct dx.	78.4% (3.7%)	35.2% (4.8%)
	80.0% (3.6%)	**6.9% (2.6%)**
Opted for alternate dx.	7.4% (2.2%)	60.2% (4.9%)
	3.1% (1.9%)	**66.9% (5.0%)**

*numbers in parentheses represent the standard error. The numbers in bold are percent of cases in which the students in Experiment 1A opted for the correct or alternate diagnoses in the equivalent conditions.

Table 10. Experiment 3. Impact of a competing diagnosis: The mean percent of features of the correct and alternate diagnoses identified by the students.

	Bias	
	Correct Hx. & Dx First Alternate Dx Second	Alternate Hx. & Dx First Correct Dx Second
Mean percent of features:		
Consistent with correct dx.	43.7% (3.3%)	40.6% (3.5%)
	48.8% (3.9%)	**35.1% (4.3%)**
Consistent with alternate dx.	13.9% (2.1%)	21.4% (2.8%)
	5.7% (1.5%)	**26.5% (2.8%)**

*numbers in parentheses represent the standard error. The numbers in bold are percent of correct and alternate features identified by the students in Experiment 1A in the equivalent conditions.

Table 11. Experiment 4. Manipulating the number of diagnoses available at test: The mean percent of cases in which the students opted for the correct and alternate diagnoses.

	Bias	
	Correct Dx.	Alternate Dx.
*3 Diagnoses available a time of test:		
Mean percent of cases:		
Opted for correct dx.	86.3% (3.6%)	49.2% (3.6%)
Opted for alternate dx.	7.5% (4.0%)	46.3% (4.0%)
*10 Diagnoses available a time of test:		
Mean percent of cases:		
Opted for correct dx.	81.4% (3.7%)	46.4% (3.7%)
Opted for alternate dx.	7.3% (4.2%)	45.9% (4.2%)

*numbers in parentheses represent the standard error

Table 12. Experiment 4. Manipulating the number of diagnoses available at test: The mean percent of features of the correct and alternate diagnoses identified by the students.

	Bias	
	Correct Dx.	Alternate Dx.
Mean percent of depicted features:		
3Dx at test:		
Consistent with correct dx.	82.7% (3.2%)	67.9% (3.2%)
Consistent with alternate dx.	58.8% (4.9%)	78.7% (4.9%)
Irrelevant features	6.1% (2.7%)	6.4% (2.7%)
10Dx at test:		
Consistent with correct dx.	85.1% (3.4%)	70.3% (3.4%)
Consistent with alternate dx.	51.3% (5.1%)	74.1% (5.1%)
Irrelevant features	17.1% (2.8%)	15.4% (2.8%)
Mean percent of potential features:		
3Dx at test		
Consistent with correct dx.	25.0% (5.5%)	27.1% (5.5%)
Consistent with alternate dx.	14.6% (3.9%)	25.9% (3.9%)
Irrelevant features	2.1% (0.5%)	3.2% (0.5%)

1ODx at test

Consistent with correct dx.	29.5% (5.5%)	18.2% (5.8%)
Consistent with alternate dx.	11.2% (3.9%)	31.4% (4.1%)
Irrelevant features	3.3% (0.5%)	2.1% (0.5%)

*numbers in parentheses represent the standard error

Appendix A: Correct and alternative diagnoses and case histories used in Experiments 1, 2 and 3

Case 3V

Correct: A man came to the emergency room with haematemesis and mild epigastric pain. The tentative is **stomach cancer.**

-anorexia/paleness
-weight loss (cachexia/wasting)
-left superclavicular lymph node

Alternative: A 64 year old man presents with a long history of cirrhosis. He presents with new onset of anorexia and weight loss. The tentative diagnosis is **liver cancer.**

-jaundice
-weight loss
-fever and sweating

Case 7V

Correct: A lady presented with gradual onset of malaise and weakness associated with an 8 kilogram weight loss. The tentative diagnosis is **polymyositis.**

-proximal muscle weakness
-heliotrope rash
-fever

Alternative: A 55 year old woman presents with gradual fatigue, lethargy and cold intolerance. The tentative diagnosis is **hypothyroidism.**

-hair is dry and falling out
-periorbital edema
-rough, doughy skin

Case 9V

Correct: A woman presented with primary amenorrhea. The tentative diagnosis is **Turner's Syndrome.**

-webbing of the neck
-abnormal facies (narrow maxilla, micromandible, epicanthal folds)
-prominent low-set or deformed ears

Alternative: This mentally handicapped 30 year old woman was discovered at the age of 14 to have hypothyroidism. The tentative diagnosis is **cretinism.**

-broad flat nose
-widely set eyes
-dry skin

Case 12V

Correct: A pregnant woman presented with fever and joint pain. The tentative diagnosis is **systemic lupus erythematosis.**

-malar (butterfly) rash

Alternative: A 24 year old woman recently had a first seizure and was started on Dilation. She presents to the emergency department with a sore mouth and pain on swallowing. The tentative diagnosis is **Steven's-Johnson syndrome.**

-mucosal lesions
-small blisters on purpuric macules
-fever

Case 13V

Correct: A woman presented with generalized muscle weakness and amenorrhea. The tentative diagnosis is **Cushing's disease.**

-easy bruisability
-deposition of adipose tissue (moon-shaped face, buffalo hump)
-plethoric face
-virilizing signs (acne, hirsutism)

Alternative: A 25 year old woman presents with recent rapid gain weight, swelling of the ankles and frothy urine. The tentative diagnosis is **nephrotic syndrome.**

-facial edema
-periorbital edema
-ill appearance

Case 16V

Correct: A boy presented with fever, malaise and pain on swallowing. The tentative diagnosis is **mumps.**

-malaise
-swelling of parotid glands

Alternative: This young man presents with rapid weight gain, new onset diabetes and hypertension. The tentative diagnosis is **Cushing's disease.**

-deposition of adipose tissue (moon facies, buffalo hump)
-plethoric face

Case 18V

      Correct: A woman presented with increased fatigue and weakness. The tentative diagnosis is **hypothyroidism.**

                -weight increase
                -hair is dry and falling out
                -myxedema (protruding tongue; periorbital edema; rough doughy
                          skin)

      Alternative: This 60 year old lifelong smoker presents with increasing somnolence and confusion. The tentative diagnosis is **respiratory failure.**

                -cyanosis
                -diltated pupils
                -tired, drowsy-looking
                -distressed, anxious looking
                -pale
                -sweating
   -pursed lips

Case 27V

      Correct: A child was brought to the emergency room because he was passing dark colored urine. The tentative diagnosis is **acute glomerulonephritis.**

                -periorbital edema
                -fever
                -mild anemia

      Alternative: A boy presented with fever, malaise and pain on swallowing. The tentative diagnosis is **mumps.**

      .

                -malaise
                -swelling of parotid glands

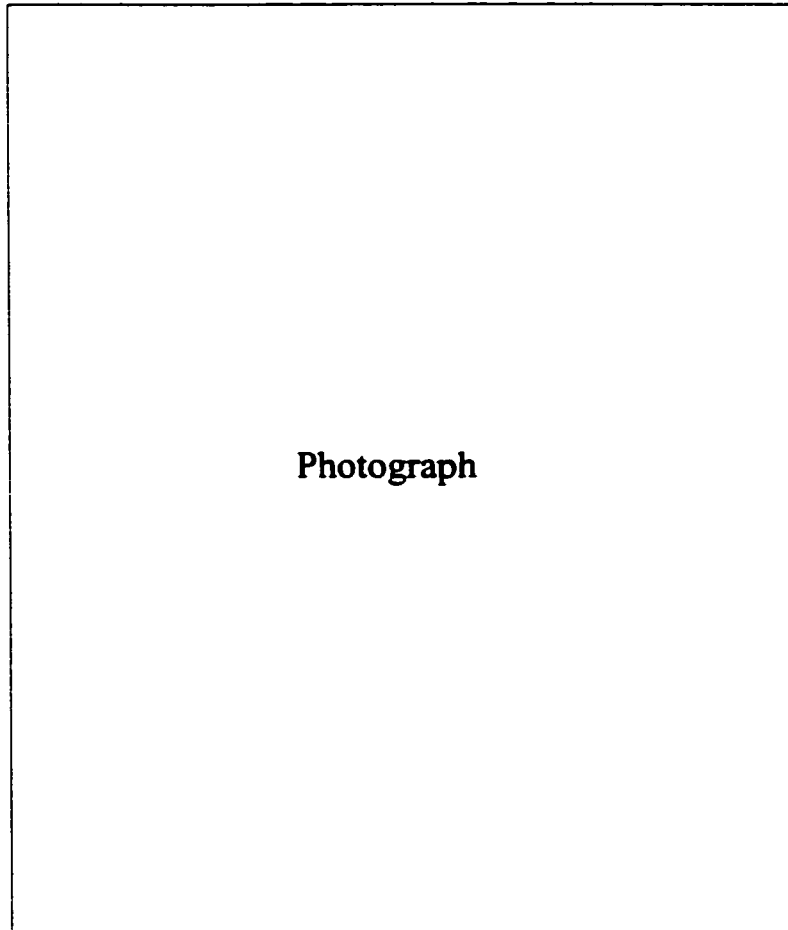**Appendix B: Examples of head-and-shoulder photographs used in Experiments 1-3.**
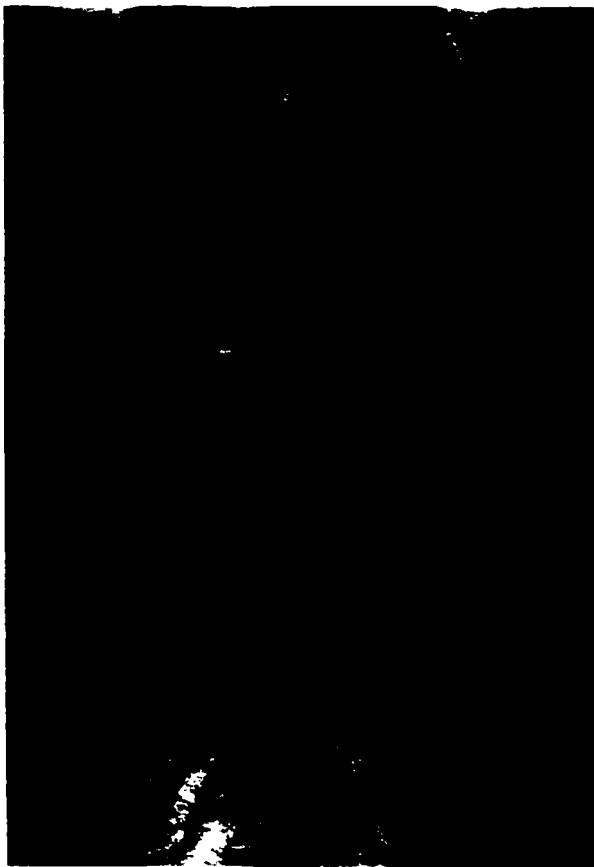
Scenario 3: Stomach Cancer

Scenario 16: Mumps

Appendix C: Example of biasing information and response sheet from Experiment 1.

Case 3V

Photograph

A man came to the emergency room with haematemesis and mild epigastric pain. The tentative is stomach cancer.

A) Please look at the photograph, and identify **any and all** clinically important features.

_____   _____   _____

_____   _____   _____

_____   _____   _____

_____   _____   _____

B) Considering all of the information, rate the likelihood of the tentative diagnosis. Are there any other diagnoses that you are seriously considering? What is their likelihood? What is the likelihood that it might be something else?

Stomach Cancer: _____%

_____: _____%

_____: _____%

Something else _____%

Total   100%

Appendix D: Example of answer sheet used in Experiment 3

A) In addition to liver cancer, the differential diagnosis includes stomach cancer. Please look at the photograph, and identify **any and all** clinically important features.

_____ _____ _____

_____ _____ _____

_____ _____ _____

_____ _____ _____

B) Considering all of the information, rate the likelihood of both the tentative and differential diagnoses. Are there any other diagnoses that you are seriously considering? What is their likelihood? What is the likelihood that it might be something else?

Liver Cancer: _____%

Stomach Cancer   _____%

_____: _____%

_____: _____%

Something else   _____%

                    Total   100%

Appendix E: Rules of ECG diagnoses used in Experiment 4 (10 and 3 diagnosis test conditions)

Features of Various Disorders

Normal
-predominant S wave in V1 and V2
-predominant R wave in V5 and V6
-R wave gets larger from V2 to V4
-T wave is positive in leads with and R wave, but negative in aVR

Right Ventricular Hypertrophy
-Predominant R wave in V1 (R > S wave or only R wave)
-R wave in V1 is greater than 7mm
-R wave may get progressively smaller from V2 to V4

Left Ventricular Hypertrophy with Strain
-Large S waves in V1 and V2   the sum of the largest R wave and the
-Large R waves in V5 and V6   largest S wave is > 35mm
-ST depression in V4 to V6 and/or
-T waves inverted in V4 to V6

Left Bundle Branch Block
-Widening of QRS complex (>3 small blocks wide)
-RSR' (rabbit ears) in V5 and V6

Right Bundle Branch Block
-Widening of QRS complex (>3 small blocks wide)
-RSR' (rabbit ears) in V1 and V2

Acute Anterior Myocardial Infarction
-ST elevation in V1 to V4
-Q waves present in V1 to V4 (not necessary for diagnosis)

Acute Inferior Myocardial Infarction
-ST elevation in II, III, aVF
-Q waves present in II, III, aVF (not necessary for diagnosis)

Ischemia
-ST depression and/or T wave inversion in any of the leads

Pericarditis
-ST elevation across many leads

Hyperkalemia
-Peaked T waves across many leads, particularly from V2 to V6
-As hyperkalemia progresses, widening of QRS complex can occur.

Features of Various Disorders
(3 diagnoses test condition)

Left Ventricular Hypertrophy with Strain
-Large S waves in V1 and V2   the sum of the largest R wave and the
-Large R waves in V5 and V6   largest S wave is > 35mm
-ST depression in V4 to V6 and/or
-T waves inverted in V4 to V6

Acute Anterior Myocardial Infarction
-ST elevation in V1 to V4
-Q waves present in V1 to V4 (not necessary for diagnosis)

Ischemia
-ST depression and/or T wave inversion in any of the leads

This ECG was shown to a group of Psychology 1A03 students and one of the diagnoses raised was Hyperkalemia

A. Please look at the ECG and list any and all clinically important features that are present

_____   _____   _____

_____   _____   _____

B: Considering all of the information, rate the likelihood of the suggested diagnosis. Are there any other diagnoses that you are considering? What is their likelihood? What is the likelihood that it might be something else?

Hyperkalemia   _____%

_____ :   _____%

_____ :   _____%

Something Else: _____%

                                         Total   100%