

Thinking about diagnostic thinking: a 30-year perspective

Arthur S. Elstein

Received: 14 July 2009 / Accepted: 14 July 2009 / Published online: 11 August 2009
© Springer Science+Business Media B.V. 2009

Abstract This paper has five objectives: (a) to review the scientific background of, and major findings reported in, Medical Problem Solving, now widely recognized as a classic in the field; (b) to compare these results with some of the findings in a recent best-selling collection of case studies; (c) to summarize criticisms of the hypothesis-testing model and to show how these led to greater emphasis on the role of clinical experience and prior knowledge in diagnostic reasoning; (d) to review some common errors in diagnostic reasoning; (e) to examine strategies to reduce the rate of diagnostic errors, including evidence-based medicine and systematic reviews to augment personal knowledge, guidelines and clinical algorithms, computer-based diagnostic decision support systems and second opinions to facilitate deliberation, and better feedback.

Keywords Clinical judgment · Problem solving · Clinical reasoning · Cognitive processes · Hypothetico-deductive method · Expertise · Intuitive versus analytic reasoning

Accounts of clinical reasoning fall into one of two categories. The first comprises narratives or reports of interesting or puzzling cases, clinical mysteries that are somehow explicated by the clever reasoning of an expert. The narrative includes both the solution of the mystery and the problem solver's account of how the puzzle was unraveled. These case reports resemble detective stories, with the expert diagnostician cast in the role of detective.¹ A recent example of this genre is “[How Doctors Think](#)”, a collection of stories of clinical reasoning that has become a best-seller (Groopman 2007).

The second broad category of studies falls within the realm of cognitive psychology. Investigators have analyzed clinical reasoning in a variety of domains—internal medicine, surgery, pathology, etc.—using cases or materials that might be mysteries but often are

¹ The fictional detective Sherlock Holmes, perhaps the model of this genre, was inspired by one of Sir Arthur Conan Doyle's teachers in medical school, Dr. Joseph Bell (http://www.siracd.com/work_bell.shtml).

A. S. Elstein (✉)

Department of Medical Education, University of Illinois at Chicago, Chicago, IL, USA
e-mail: aelstein@sbglobal.net

intended to be representative of the domain. Rather than offering the solution and reasoning of one clinician, the researchers collect data from a number of clinicians working the case to see if general patterns of performance, success and error can be identified. Some studies use clinicians of varying levels of experience, thereby hoping to study the development of expert reasoning or to discern patterns of error. Others concentrate on experienced physicians and aim to produce an account of what makes expert performance possible. These investigations are based on samples of both subjects and cases, although the samples are often relatively small.

Two recent reviews (Ericsson 2007; Norman 2005) of the second category of studies date the beginning of this line of work to the publication of “**Medical Problem Solving**” (Elstein et al. 1978). It is exceedingly gratifying to be so recognized. When my colleagues and I began this project at Michigan State University 40 years ago, we never dreamed that our book would turn out to be a classic in medical education, cited as the foundation of the field of research on clinical judgment.²

For this reason, this paper begins with a review of the state of the field when the Medical Inquiry Project (as we called it) was conceived and implemented and then summarizes the major finding of that study. Next some of Groopman’s case studies will be examined because they vividly illustrate several important conclusions of the research and the strengths and weaknesses of the clinical narrative approach. The next two sections review some of the criticisms of the hypothetico-deductive model and some of the major errors in diagnostic reasoning that have been formulated by two somewhat divergent approaches within cognitive psychology—the problem solving and judgment/decision making paradigms. The essay concludes with some suggestions to reduce the error rate in clinical diagnosis.

Medical Problem Solving: the research program

The Medical Inquiry Project was aimed at uncovering and characterizing the underlying fundamental problem solving process of clinical diagnosis. This process was assumed to be “general” and to be the basis of all diagnostic reasoning. Because the process was assumed to be general and universal, extensive sampling of cases was not necessary, and each physician who participated was exposed to only three cases in the form of “high-fidelity simulations”, a form known now as “standardized patients” (Swanson and Stillman 1990), plus some additional paper cases. Some of the paper simulations were borrowed from work on clinical simulations known as “Patient management problems” (PMPs; McGuire and Solomon 1971).

In the light of our findings and subsequent research, the assumption of a single, universal problem-solving process common to all cases and all physicians may seem unreasonable, but from the perspective of the then-current view of problem solving, it was quite plausible. Take language as an example: how many spoken sentences do you need to hear to judge whether or not a speaker is a native or non-native speaker of English? Not very many. Chomsky’s work on a universal grammar that was assumed to underpin all

² To give credit, a case can be made that the field of research on clinical diagnosis really began earlier with studies of variation in clinical judgment that are nowadays regrettably overlooked (Bakwin 1945; Yerushalmy 1953; Lusted 1968). The reasons for neglect are unclear and in any event are beyond the scope of this paper, but it is a plain fact that they are cited neither by Ericsson nor Norman, nor by the vast majority of psychologists, physicians and educators who have done research on clinical reasoning in the past 30 years.

languages was one influence on our strategy (Chomsky 1957). Newell and Simon (1972), Newell et al. (1958) had been working toward a “general problem solver” (GPS). Their research on human (non-medical) problem solving similarly employed small sets of problems and intensive analysis of the thinking aloud of relatively few subjects. Likewise, deGroot’s study of master chess players used just a few middle-game positions as the task (De Groot 1965).

Early in our work, Lee Shulman and I visited the Institute of Personality Assessment and Research at the University of California Berkeley. Ken Hammond was in the group to which we presented our plans and some pilot data. As the Brunswikian he is, Hammond warned us to sample tasks more extensively. He argued that tasks differ more than people and that we were not sampling the environment adequately. We did not understand the force of his comments until much later, but his views have influenced much subsequent research: both cases and physicians have to be adequately sampled.

The assumption of a general problem-solving process also meant that any physician could be expected to exemplify the process. General internists and family physicians were the subjects, not specialists in the cases selected. Peer ratings of excellence were obtained to categorize the physicians as either experts or non-experts (the terms used in the book are “criterial” and “non-criterial”). An earlier review of this project (Elstein et al. 1990) has discussed the problems arising from using peer ratings to identify expertise. Questions about whether the clinical reasoning of specialists and generalists would be similar did not arise until later.

The major findings of our studies are quite familiar and can be briefly summarized. First, given the small sample of case materials used, expert and non-expert physicians could not be distinguished. The objective performance of expert diagnosticians was not significantly better than that of non-experts. All the subjects generated diagnostic hypotheses, all collected data to test hypotheses. The “experts” were not demonstrably more accurate or efficient. The groups might differ, but evidence was lacking.

Second, diagnostic problems were approached and solved by a hypothetico-deductive method. A small set of findings was used to generate a set of diagnostic hypotheses and these hypotheses guided subsequent data collection. This did indeed appear to be a general method, but since this method was employed whether or not the final diagnosis was accurate, it was difficult to claim much for it.

Third, between three and five diagnostic hypotheses were considered simultaneously, although any physician in our sample could easily enumerate more possibilities. This finding, a magic number of 4 ± 1 , linked our results to work on the size of short-term memory (Miller 1956) and to Newell and Simon’s notion of limited or bounded rationality (Newell et al. 1958; Newell and Simon 1972). Thus, our analysis was far more grounded in cognitive theory than prior research on clinical reasoning or the usual clinical narrative.

Finding #4 was that expertise was content- or case-specific. Physician performance was variable, and the outcome on one case was a poor predictor of performance on another. In medical education, this has been a very influential finding, especially in the assessment of clinical competence. In case-based examinations, test designers now use more cases to get satisfactory generalizability (Swanson and Stillman 1990; Bordage and Page 1987). The PMP method, which asked many questions about a small number of cases, was gradually replaced by approaches that sample cases more broadly and ask fewer questions about each case.

Taken as a whole, these findings implied that expertise was more dependent on specific clinical experience than we had assumed. The concluding chapter of “[Medical Problem](#)

[Solving](#)” asked whether limited medical licensure was a desirable direction. Instead, the medical profession has proliferated specialties, another way of limiting practice.

How Doctors Think: analysis of case reports

There are a number of points of contact between these results and the clinical narratives in “[How Doctors Think](#)”. Like the Michigan State team 40 years ago, Groopman believes that beneath the apparent differences in subject-matter knowledge of different specialties, competent performance has important similarities across domains: Hypotheses are generated early, and expertise is acquired by repeated practice and feedback, especially about mistakes. He stresses the role of prior experience, a point repeated in the research literature of the past 30 years. In addition, he appeals frequently to principles from the psychological literature on clinical reasoning, especially the topic of heuristics and biases, to provide theoretical grounding for the stories. (At the same time, he argues that decision trees and Bayesian reasoning, elements of the same theory, are unlikely to be clinically useful, while others see these techniques as helpful remedies for the cognitive problems disclosed by psychological research.) Finally, because the book is so well written, it quite likely will have more general readers than all of the other papers and books cited in this review combined. So it makes sense to review some of his case studies to see what they do and do not teach.

The introductory chapter tells the story of a patient called Anne Dodge. She has been incorrectly diagnosed with anorexia, bulimia and/or irritable bowel syndrome and treated unsuccessfully for several years by different experts. Her failure to respond to a treatment regimen that she says she is following faithfully is taken as further evidence that her problem is in some way or other “in her head”. At last, she is referred to an experienced gastroenterologist who sets aside her thick chart and all previous laboratory studies and asks her to simply tell her story from the beginning. He quickly suspects celiac disease (malabsorption syndrome), and after some additional laboratory studies, makes that diagnosis and sets her on a proper treatment regimen. How did he do it? Why did others miss the diagnosis?

The hero of this story says that he made the diagnosis because he listened to the patient (and presumably her previous physicians did not). He explicitly recalls Osler who, in the era before the clinical laboratory had developed, urged physicians to listen to their patients, they are telling you the diagnosis. Groopman emphasizes the importance of asking open-ended questions to allow the patient to tell her own story and then listening carefully to that account. Unfortunately, we have no way of knowing from this narrative if Anne Dodge’s previous physicians used open-ended questions to allow her to tell her own story or if they used highly structured interviews. Maybe they did not listen carefully, or perhaps they listened carefully but misinterpreted what she said and still made a diagnostic error. In “[Medical Problem Solving](#)”, we reported that thorough data collection and accurate data interpretation were uncorrelated. That might be the case here too.

It turns out that the successful consultant had extensive prior experience with malabsorption syndrome because he had done research at NIH on that problem. Indeed, he had once been fooled by a patient who was exhaustively worked up for that problem and was accidentally discovered to be using laxatives surreptitiously. She really was a psychological problem. It is unclear why this experience predisposed him to give more weight rather than less to Anne Dodge’s account of her illness, but it does explain why he considers both psychological and biological components in a patient’s experience. In trying

to explain why her previous physicians erred, Groopman speaks of the power of an early frame (or hypothesis). Granted that the early frame provides a powerful lens through which subsequent data are interpreted, this does not explain why those physicians did not consider malabsorption syndrome in their preliminary frames. Did they lack extensive prior experience? Were they all convinced the issue was psychological? We just don't know. Thus, the case can be understood to highlight the importance of the clinician's prior experience and deep knowledge of the disease in question, but it casts little light on the riddle of hypothesis generation.

The next case takes place in a very dramatic setting: a young physician's very first night on call as an intern at the Massachusetts General Hospital. He is speaking with a patient, a 66-year-old African-American man with hypertension that has been difficult to control and who had been admitted 2 days earlier with chest pain. As he is about to say goodbye, the patient "shot upright in bed. His eyes widened. His jaw fell slack. His chest began to heave violently." And he was unable to speak. The intern was paralyzed by fear. He forgot everything he knew and simply did not know what to do. At that moment, he was rescued by a visiting cardiologist who had trained some years before at MGH and just happened to be visiting the hospital. He passed by the patient's room at precisely the right moment. The narrative now becomes a story of heroic rescue. The experienced physician recognized the emergency (a ruptured aortic valve), knew what to do, and responded quickly. Groopman contrasts this apparently effortless demonstration of know-how, developed by extensive clinical experience, with the intern's sense of helplessness, exacerbated by the fact that he had been an outstanding medical student. He carried a pack of index cards in his pocket (nowadays it would be a PDA) but in the moment of crisis he forgot everything he knew (or so it seemed to him), while his rescuer knew exactly where to listen in the chest, how to interpret what he heard, and what had to be done immediately.

This story makes two points relevant to our understanding of diagnostic reasoning. It vividly demonstrates the difference between propositional knowledge, the kind acquired in the early years of a medical education, and the practical know-how acquired by post-graduate medical education and further experience. Second, it emphasizes, again, the role of prior clinical experience, now in the service of rapid pattern recognition. It is not clear if the behavior of the expert cardiologist should even be called "thinking" (Hamm 1988). It is surely a different kind of thinking than was called into play in the case of Anne Dodge. Some investigators identify this rapid, poorly verbalized, highly overlearned process as "intuition"; others prefer "pattern recognition".

The last selection involves a patient with debilitating pain and swelling in his dominant right hand. The patient is Groopman himself. His personal experience seeking a diagnosis and treatment plan for this problem is a cautionary tale of diagnostic and treatment variation. Over the course of 3 years, he had consulted six hand surgeons and received four different opinions about what was wrong and what to do about it. The diagnoses and recommendations of five hand surgeons are summarized, but no interviews with them are reported, so nothing is said about their thought processes. Finally he chose surgery with the fifth in the series, because he recommended the same procedure as the fourth and had more experience with it. This story is told to yet a sixth hand surgeon, one who had just finished a term as president of the American Society for Surgery of the Hand and was about to assume presidency of the American Orthopedic Association. Few patients would get to bring their medical mystery to so prominent an expert. This consultant fortunately agrees with the diagnosis and plan of the fourth and fifth ("fortunately," because what if he had thought one of the others was more correct?). The question both he and Groopman ponder is, why did it take 3 years to find a surgeon who thought about the problem correctly?

This case history of diagnostic variability leading to different treatment plans is likely to upset lay readers who might believe that there is little variation between expert physicians. This concern is exacerbated by some additional facts: Groopman and his wife are both physicians and she accompanied him to each of these consultations to be sure the right questions were asked. Presumably they asked better questions and received more detailed explanations of each consultant's recommendations than the average patient would get. If this is the level of expertise that patients and their advocates need to thread their way through the health-care system to receive competent advice, patients are in even bigger trouble than we thought.

"How Doctors Think" is particularly interesting because it both celebrates master clinicians and yet recognizes the value of systematic, algorithmic approaches that could be used by any clinician. It uses principles of cognitive psychology to explain both successes and diagnostic errors. Groopman has read widely about the acquisition of expertise, the role of pattern recognition, the importance of practice and feedback. The literature on heuristics and biases (Kahneman et al. 1982), a mainstream theme in the psychology of judgment and decision making, seems to capture most of his interest. He refers repeatedly to availability, representativeness, attribution errors and framing effects. These concepts flow from a line of research on decision making that developed largely after the publication of "[Medical Problem Solving](#)", and it has become a dominant theme in current research on the psychology of decision making generally, and of clinical decision making in particular (Chapman and Elstein 2000; Elstein 1999).

Critique of Medical Problem Solving

As both Ericsson's (2007) and Norman's (2005) reviews note, the findings of this book shifted the emphasis away from a search for general processes to concern with the structure of long-term memory and how it is built through repeated practice and experiences. Criticisms of the hypothetico-deductive model were not long in coming. Many investigators (e.g., Bordage 1994; Bordage and Lemieux 1991; Patel and Groen 1986, 1991; Patel et al. 1986; Schmidt et al. 1990; Gruppen and Frohna 2002) have explored how memory structure and/or knowledge structure affect diagnostic reasoning. All place more emphasis on knowledge structure and rapid retrieval than on formal reasoning processes. They showed convincingly that formal hypothesis generation and testing is but one form of diagnostic thinking, certainly not a complete account of the process.

Kassirer and Kopelman (1991) have suggested that novices are more likely to use a hypothesis-testing strategy until they develop the highly compiled, elaborated knowledge of experienced clinicians. In a defense of the hypothetico-deductive method, Elstein (1994) proposed that it would be used when pattern recognition or recalling a previous similar instance fail to produce a satisfactory solution. That argument failed to deal with the question of how a physician determines whether or not a solution is satisfactory. Response to treatment is one way to answer this question, but obviously it did not work in the case of Anne Dodge.

Looking back at the studies in "[Medical Problem Solving](#)", we may ask if the emphasis on hypothetico-deductive reasoning over pattern recognition came about because the overall design of the project led to data that fit this model. And why? It is very possible that (a) the problems were more familiar to specialists in neurology, gastroenterology, and infectious disease than general internists, and (b) the research method itself encouraged physicians to be systematic: they knew they were being observed and videotaped, and

would be asked to justify retrospectively their procedure and conclusions. We perhaps underestimated the degree to which the experimental setting biased the physicians to respond in a particular way.

We know now that experienced physicians can and do use all kinds of methods. In current cognitive theory, rapid, non-verbal, intuitive cognition is characterized as System 1, while the slower, more formal process of hypothesis formation and testing is a function of System 2 (Schwartz and Elstein 2008).

But how do clinicians decide which approach is best suited for the case at hand? And what is the error rate in these decisions? Given that there is more than one way to solve a diagnostic problem, the difficult question for clinical practice is this: when does the physician need to engage in a slow, careful logical process of hypothesis generation and testing, and when will short-cut methods, like pattern recognition or recalling the solution to a previous case work just as well or better? Or, as Gropman and other thoughtful physicians worry, are time pressure and economics determining that quick, simple methods will dominate even when more formal approaches are called for?

Errors in diagnostic reasoning

Several recurring errors that lead to diagnostic mistakes have been identified and reviewed more extensively by many investigators (e.g., Kassirer and Kopelman 1991; Schwartz and Elstein 2008): Limited or bounded rationality implies that not all of the hypotheses in a physician's long-term memory can be simultaneously evaluated. The search is typically for a satisfactory solution, within the practical constraints of the clinical environment, not necessarily the optimal solution. The formation of early hypotheses is essential to limit the problem space, but there is always the hazard of premature closure. If one chooses to rely on pattern recognition, what is to guarantee that the right pattern or illness script will be selected? And what if the patterns or scripts stored in the individual knowledge base are in some way mistaken or faulty? Rapid pattern recognition, when it is right, it is a triumph of efficiency and experience. When wrong, it can be viewed as a special case of premature closure. The problem solver may interpret neutral cues as supporting a favored hypothesis, in order to keep the problem relatively simple (Klayman 1995) or fail to use the data to test competing hypotheses (Wolf et al. 1985). If the problem solver's knowledge is deficient or dispersed and not well-elaborated, his behavior may look clueless, displaying an inadequate theory of the case (Bordage 1999).

Behavioral decision research has identified a variety of psychological processes that affect both probability estimation (diagnosis) and choice (of diagnostic or treatment alternatives). The list includes availability, representativeness, anchoring and adjustment, framing effects, omission bias, overconfidence, hindsight bias, and other principles (Kahneman et al. 1982; Chapman and Elstein 2000; Elstein 1999; Schwartz and Elstein 2008; Dawson and Arkes 1987; Dawson et al. 1988). These heuristics and biases are often useful shortcuts, necessary for getting the work done in real time. Yet they may lead to errors in diagnosis or management. The troubling question remaining is: Given that our intuition is not perfect and that rational analytic thought is too time consuming, when should we trust our clinical intuition and when is a more systematic rational approach needed? How should we decide that question? Here, in my opinion, the next generation of researchers on clinical cognition can and should devote some effort.

Improving clinical practice

Having reviewed, all too briefly, some of the main causes of diagnostic errors, let us move on to considering how to reduce them.

Increased knowledge

If a significant component of successful diagnosis is prior knowledge, then remediable knowledge deficiencies should be implicated in errors. Indeed, Evidence-Based Medicine is based on two linked propositions: that physicians' personal knowledge cannot always be sufficient and that they have difficulty retrieving clinically relevant information quickly and evaluating the published literature accurately once they find it. EBM proposes to address these problems by showing physicians how to formulate an answerable clinical question to focus the literature search and a set of templates for critical appraisal of what is retrieved (Sackett et al. 1997). This strategy turned out to be less useful than had been hoped, and has been augmented by increasingly providing busy practitioners with systematic reviews of the literature, by such institutions as the Cochrane Collaboration. As a general principle, more knowledge is always better than less, but sooner or later, every physician will be confronted with a situation where knowledge is incomplete. Can these situations be recognized so that the busy practitioner will know when to seek consultation or search the literature?

It would be good if physicians were as well acquainted with the relevant principles of cognitive psychology as they are with comparable principles in pathophysiology. But this knowledge must be organized in ways that are clinically useful, otherwise it may not help in practice any more than the intern's detailed biomedical knowledge helped with his first case. Appeals to rationality and the necessity of acquiring new knowledge are necessary but insufficient. Another serious problem with relying on the clinical literature to augment personal knowledge is the extent to which the trustworthiness of research results in the published literature has been eroded by ongoing stories of manipulation, bordering on deceit (Sox and Rennie 2008). The medical profession's concern with this issue appears to have grown in the past decade.

Guidelines and clinical algorithms

The possibilities of errors and deficiencies in clinical practice have been widely recognized by specialty associations and government agencies. Clinical guidelines and algorithms have been extensively developed in almost every area of medicine, aimed at providing physicians with practical recommendations to enhance medical care. An updated review of this extensive literature is beyond the scope of this paper. The interested reader is referred to (Elstein et al. 2002). Perhaps the greatest virtue of this approach is that well-designed algorithms and guidelines try to remedy deficiencies in human judgment by incorporating principles of statistics, decision theory, and epidemiology in a clinically useful format. These principles have been around for over 20 years in textbooks (Kassirer and Kopelman 1991; Albert et al. 1988; Sackett et al. 1991; Sox et al. 1988; Hunink et al. 2001; Schwartz and Bergus 2008) but their impact on clinical practice still seems modest at best.

Reflection and second opinions

The method of hypothesis generation and testing is a form of reflection. It offers the opportunity for a physician to think about alternatives. Given the realities of clinical practice, it cannot be applied to all cases nor is it necessary; we need a better sense of when it should be employed.

Developers of systems for computer-assisted diagnosis hoped that they would provide convenient, inexpensive and accurate second opinions. In my judgment, it has not worked out that way. The difficulties and problems have been well documented (Berner et al. 1994; Friedman et al. 1999). Still, a more effective system may be just around the corner. It is possible that the findings in the electronic medical record of a multi-problem patient could be automatically entered into a diagnostic decision support system that would be sufficiently intelligent to distinguish the unknown problem from the list of diagnoses already identified. To my knowledge, such a system is not yet available, but given the pace of development of computer applications, it would foolhardy to forecast the future.

The use of consultation and second opinions is another way to encourage reflection and deliberation. Groopman tells of a radiology practice in which a fraction of all radiological images are routinely subjected to a second reading. This practice should be especially effective with all kinds of images. Decision analysis (Hunink et al. 2001) is another way of subjecting intuitive judgments and decisions to critical review. If more off-the-shelf decision trees were available, clinicians would be able to compare their planning with the results implied by a more formal analysis. Of course there are tradeoffs: additional costs to the health care system, perhaps some delay in starting treatment. Do we know if the reduced error rate is sufficient to justify the increased costs and delays? Or is this another direction for future research?

Thoughtful physicians (Groopman 2007; Kassirer and Kopelman 1991) have noted that time pressure is likely to further degrade performance: Physicians are under pressure to see more patients per hour and every time Medicare payments are reduced, that pressure increases. Clinical practitioners, professional associations and medical educators are obliged to call this problem to the attention of policymakers.

Debiasing by improved feedback

Debiasing strategies, including metacognition (which seems similar to reflection) should be tried (Croskerry 2003). But it is unlikely they will be as effective as proponents hope, because the biases are not simple knowledge deficiencies. Overconfidence and premature closure seem to be deeply engrained in the way we think (Elstein 1999). However, providing prompt feedback about predictions has proven to be a reasonably effective debiasing procedure. Weather forecasters are less biased in their predictions than other types of experts because the system is engineered to provide rapid feedback on accuracy. Ericsson (2007) correctly observes that expert performance is acquired by practice and feedback. Medical students and residents get lots of supervised practice. But how good is the feedback once in clinical practice? How do we learn from errors if we don't know that an error has occurred? Or if we are not sure if a bad outcome is due to error or just bad luck? Just compare the feedback available to health professionals with that provided continually to concert musicians and professional baseball players, two domains where immediate feedback is regularly provided and where mistakes are carefully reviewed. Improving feedback to clinical practitioners may be the most effective debiasing procedure available.

Indeed, improving how feedback is provided and used in the clinical setting has been identified as a priority task for reducing diagnostic errors (Schiff et al. 2005).

Conclusion

Diagnostic errors can never be entirely eliminated (Graber et al. 2002). Human reasoning is not perfect, and so mistakes in interpretation and inference will be made. Sooner or later, situations will be encountered where one's knowledge will be incomplete. Feedback in the clinical setting is spotty. Second, even if our knowledge were complete and our inference processes were perfect, clinical evidence is not. Patients may not tell entirely accurate histories, physical findings and laboratory tests do not perfectly distinguish between diseased and disease-free populations. The sensitivity and specificity of every diagnostic test are typically <1.0. We can and should try to reduce the error rate, but some errors are no-fault products of the deficiencies identified. The benefit of more widely disseminated knowledge of the psychology of diagnostic reasoning may be to facilitate humility and attitude change.

Many years ago, I heard Ken Hammond attribute a maxim to La Rochefoucauld: "Everyone complains about his memory and no one complains about his judgment." Perhaps the research reviewed here will demonstrate why we should be concerned.

Acknowledgments An earlier version of this paper was presented as a keynote address at a conference, "Diagnostic Error in Medicine," held in Phoenix, AZ, May 31-June 1, 2008. I thank the organizing committee—Eta Berner, Pat Croskerry, Mark Graber, and Gordon Schiff—for the invitation and for encouraging personal reflections on the subject. My co-authors and students have taught me a great deal. I owe much to my colleagues in the field, some cited in this paper and some not. Errors of facts and interpretation are my sole responsibility.

References

- Albert, D. A., Munson, R., & Resnik, M. D. (1988). *Reasoning in medicine*. Baltimore: Johns Hopkins University Press.
- Bakwin, H. (1945). Pseudodoxia pediatrica. *New England Journal of Medicine*, 232, 691–697.
- Berner, E. S., Webster, G. D., Shugerman, A. A., Jackson, J. R., Algina, J., Baker, A. L., et al. (1994). Performance of four computer-based diagnostic systems. *New England Journal of Medicine*, 330, 1792–1796.
- Bordage, G. (1994). Elaborated knowledge: A key to successful diagnostic thinking. *Academic Medicine*, 69, 883–885.
- Bordage, G. (1999). Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Academic Medicine*, 74, S138–S143.
- Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Academic Medicine*, 66(9), S70–S72.
- Bordage, G., & Page, G. (1987). An alternative approach to PMPs: The "key features" concept. In I. R. Hart & R. M. Harden (Eds.), *Further developments in assessing clinical competence* (pp. 59–75). Montreal: Can-Heal Publications.
- Chapman, G. B., & Elstein, A. S. (2000). Cognitive processes and biases in medical decision making. In G. B. Chapman & F. Sonnenberg (Eds.), *Decision making in health care: Theory, psychology, and applications* (pp. 183–210). Cambridge: Cambridge University Press.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, 78, 775–780.
- Dawson, N., & Arkes, H. R. (1987). Systematic errors in medical decision making: Judgment limitations. *Journal of General Internal Medicine*, 2, 183–187.

- Dawson, N. V., Arkes, H. R., Siciliano, C., Blinkhorn, R., Lakshmanan, M., & Petrelli, M. (1988). Hindsight bias: An impediment to accurate probability estimation in clinicopathological conferences. *Medical Decision Making*, 8, 259–264.
- De Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.
- Elstein, A. S. (1994). What goes around comes around: The return of the hypothetico-deductive strategy. *Teaching and Learning in Medicine*, 6, 121–123.
- Elstein, A. S. (1999). Heuristics and biases: Selected errors in clinical reasoning. *Academic Medicine*, 74, 791–794.
- Elstein, A. S., Schwartz, A., & Nendaz, M. (2002). Medical decision making. In G. Norman, C. van der Vleuten, & D. Dolmans (Eds.), *International handbook of medical education*. Boston: Kluwer.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, Mass.: Harvard University Press.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical problem solving: A ten-year retrospective. *Evaluation and the Health Professions*, 13, 5–36.
- Ericsson, K. A. (2007). An expert-performance perspective of research on medical expertise: The study of clinical performance. *Medical Education*, 41, 1124–1130.
- Friedman, C. P., Elstein, A. S., Wolf, F. M., Murphy, G., Franz, T., Miller, J., et al. (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation: A multi-site study of 2 systems. *JAMA*, 282, 1851–1856.
- Graber, M., Gordon, R., & Franklin, N. (2002). Reducing diagnostic errors in medicine: What's the goal? *Academic Medicine*, 77, 981–992.
- Groopman, J. (2007). *How doctors think*. New York: Houghton Mifflin.
- Gruppen, L. D., & Frohna, A. Z. (2002). Clinical reasoning. In G. Norman, C. van der Vleuten, & D. Dolmans (Eds.), *International handbook of medical education*. Boston: Kluwer.
- Hamm, R. M. (1988). Clinical intuition and clinical analysis: Expertise and the cognitive continuum. In J. Dowie & A. Elstein (Eds.), *Professional judgment: A reader in clinical decision making*. New York: Cambridge University Press.
- Hunink, M., Glasziou, P., Siegel, J., et al. (2001). *Decision making in health, medicine: Integrating evidence and values*. New York: Cambridge University Press.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kassirer, J. P., & Kopelman, R. I. (1991). *Learning clinical reasoning*. Baltimore: Williams & Wilkins.
- Klayman, J. (1995). Varieties of confirmation bias. In J. Busemeyer, R. Hastie & D. L. Medin (Eds.), *Decision making from a cognitive perspective: The psychology of learning and motivation* (32, 385–418).
- Lusted, L. B. (1968). *Introduction to medical decision making*. Springfield, IL: Thomas.
- McGuire, C. H., & Solomon, L. (1971). *Clinical simulations*. New York: Appleton-Century-Crofts.
- Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological Review*, 63, 81–97.
- Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, 65, 151–166.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Norman, G. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education*, 39, 418–427.
- Patel, V. L., & Groen, G. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science*, 10, 91–116.
- Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93–125). New York: Cambridge University Press.
- Patel, V. L., Groen, G., & Frederiksen, C. H. (1986). Differences between medical students and doctors in memory for clinical cases. *Medical Education*, 20, 3–9.
- Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston: Little Brown.
- Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1997). *Evidence-based medicine: How to practice and teach EBM*. New York: Churchill Livingstone.
- Schiff, G. D., Kim, S., Abrams, R., Cosby, K., Lambert, B., Elstein, A. S., et al. (2005). Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project. *Advances in Patient Safety*, 2, 255–278. Available at www.ahrq.gov/downloads/pub/advances/vol2/Schiff.pdf.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, 65, 611–621.

- Schwartz, A., & Bergus, G. (2008). *Medical decision making: A physician's guide*. New York: Cambridge University Press.
- Schwartz, A., & Elstein, A. S. (2008). Clinical reasoning in medicine. In J. Higgs, M. Jones, S. Loftus, & N. C. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 223–234). Boston: Elsevier.
- Sox, H. C., Jr., Blatt, M. A., Higgins, M. C., & Marton, K. I. (1988). *Medical decision making*. Boston: Butterworths.
- Sox, H. C., & Rennie, D. (2008). Seeding trials: Just say “No”. *Annals of Internal Medicine*, *149*, 279–280.
- Swanson, D. B., & Stillman, P. L. (1990). Use of standardized patients for teaching and assessing clinical skills. *Evaluation & the Health Professions*, *13*, 79–103.
- Wolf, F. M., Gruppen, L. D., & Billi, J. E. (1985). Differential diagnosis and the competing hypotheses heuristic: A practical approach to judgment under uncertainty and Bayesian probability. *JAMA*, *253*, 2858–2862.
- Yerushalmy, J. (1953). The reliability of chest roentgenography and its clinical implications. *Diseases of the Chest*, *24*(2), 133–147.