The influence of task demands and experience on diagnostic accuracy:

Investigating the assumptions of a default interventionist dual systems model

Sandra D. Monteiro, M.Sc.

McMaster University, monteisd@mcmaster.ca

Models of Medical Diagnosis

The influence of task demands and experience on diagnostic accuracy: Investigating the assumptions of a default interventionist dual process model

By

Sandra Monteiro, M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2013) McMaster University

(Psychology) Hamilton, Ontario

TITLE: The influence of task demands and experience on diagnostic accuracy: Investigating the assumptions of a default interventionist dual systems model

AUTHOR: Sandra Deodit Monteiro, M.Sc. (McMaster University)

SUPERVISOR: Professor Geoff Norman

NUMBER OF PAGES: xvii, 153

Abstract

There are various dual process models of human cognition. While many models of cognitive control propose processes that are selected exclusively or in combination, a default-interventionist model of reasoning assumes that processing occurs in serial stages. System 1 processes are believed to recruit unconscious memory retrieval processes by default and precede System 2 processes (Evans & Stanovich, 2013; Kahneman, 2011). System 1 processes are also considered to be overly sensitive to the automatic influences of the environment and thereby also to various cognitive biases and errors; hence System 1 is inferior. On the other hand System 2, which represent conscious logic and normative reasoning processes, is not considered susceptible to such automatic influences and thereby capable of overriding errors made through System 1 reasoning; hence System 2 is superior. This default-interventionist model has become highly influential in theories about best practices in medical education (Croskerry, 2009; 2003; Klein, 2005; Redelmeier, 2005), and has encouraged a view that increased conscious processing and reflective thought will improve performance. Such a view is in stark contrast to models of human memory in psychology that suggest contextual or automatic influences of the environment are not only critical for learning, but also critical for adaptive processing and the development of expertise (Yonelinas, 2002; Larsen & Roediger, 2012). In this thesis I investigate and critique several assumptions of the default-interventionist model by testing the relationship between processing time, reflective thought, experience and accuracy. The results of two large studies do not support basic assumptions presented in the literature and instead demonstrate that experience and knowledge are better predictors of performance.

Acknowledgements

I always knew this section would be a difficult piece to write, because there are so many people at McMaster who provided support, advice, laughter and friendship through the years and encouraged me to believe in myself. I was able to believe this could happen because other people believed I could make it happen. With all the support and encouragement, there was very little chance that I would fail. Of course behind the scenes there was even more support and encouragement from family and friends outside of McMaster. So, I find it difficult to express my gratitude properly to everyone – words do seem inadequate. This does not mean that I will not try. I can only hope that I can repay everyone in some way beyond this meagre gesture of thanks.

I dedicate this completed work to everyone who helped along the way. I suppose I should explain in more detail who they are and how they helped and so I begin at the beginning, with my husband Jerry who never waivered in his support of my return to full-time school. Thank you for believing in me even when I did not. I am also grateful for my children who balanced out my hectic academic life. I love you all. This passage to my family should far outshine any other in this thesis, but that is all I wish to write here about the most important people in my life – they will hear the rest in person.

Having made the decision to enter a graduate program I was given my first opportunity to practice and learn new research skills working for Mel Rutherford.

Mel, I felt very fortunate to be involved in your first major research project at McMaster. Thanks for making me a part of your research team.

Of course the most important vote of confidence I received at McMaster was from Bruce Milliken who agreed to be my first graduate supervisor. As I have said before to many people Bruce, you have an infectious joy for research, one that is recognized by all of your students. It is truly a pleasure to watch you get animated about a controversial issue on cognitive control and negative priming – as odd as that may sound to the inexperienced reader. You have an enviable talent for weaving a story. More importantly, you have extremely high standards for research design and are very patient with sharing your knowledge and experience. I would like to think I have absorbed that very high standard for excellence in research, but sadly, try as I might, I could not match your commitment to the finer details of attention processing. Of course, being an understanding and excellent supervisor, you continued to foster my interest in research and encouraged me to pursue a more rewarding field. I am certain, Bruce, that you were central to the early success of a research partnership with my new graduate supervisor and ultimately a new career path. Thank you for giving me an excellent start to graduate research and helping ensure my success in the program.

To everyone that I shared lab space with in the Milliken Lab – Ellen MacLellan, Adam Spadaro, Dave Thomson, Maria D'Angelo, and Chris Fiacconi, thanks for all your helpful advice through the years. Thanks to everyone in the

Shedden lab (Miriam Benarroch, Nicole LeBarr and John Grundy) for making me feel welcome at all the lab meetings and for some fruitful collaborations.

Over the past 6 years, there have also been some outstanding faculty and graduate students in the department of Psychology, Neuroscience and Behaviour that deserve recognition for their generosity of time and support.

Scott Watter, you let me ramble on about some flighty research ideas with the hopes that one day I would find the right path – it worked. Thank you Scott, for your encouragement and occasional chats over beer.

Judy Shedden, you were very generous with your time and resources and always made me feel like a member of your own lab. Thank you for allowing me to tag along.

Karin Humphreys, you were always kind and encouraging as an advocate for my success as a graduate student. Thank you for taking the time to offer guidance.

As graduate students, many of us struggle to stand out from the crowd of equally talented people that we work with everyday. But the fellow graduate students that I am fortunate enough to call friends are quite unique and stand out for their strength, courage and kindness.

Ellen MacLellan: great friend, a mother, grandmother, all around talented and passionate researcher and just brilliant at everything. I met Ellen when she

was a research assistant to Bruce Milliken and she was always encouraging, fun and helpful. Ellen you helped me see that it is not so crazy to go back to school and it was really great taking our first steps as graduate students together. Thank you for all the fun times and smart advice.

Miriam Benarroch: loyal and awesome friend, new bride, creative spirit and very talented statistician and programmer. Miriam you have a patience and drive that is outmatched. I have seen you accomplish many amazing feats over the past few years. I could not have completed my Master's thesis without you and you have been an excellent and supportive friend ever since. Miriam, our friendship has been one of the most valuable things I have gained at Mac.

Molly Pottruff: friend, roller derby expert, genuine in every way, outstanding researcher, excellent writer and editor but sadly addicted to the use of commas (joke between friends). Molly, I truly respect your skills as a researcher and have enjoyed our many conversations over the years. Thank you for being a good friend and colleague and soon to be co-author I am sure.

Now to address everyone at my new research home at the Program for Education, Research and Design (PERD) – a family of researchers formerly governed by Geoff Norman. First I want to address the many people who keep the department working and help make research ideas a reality: Elizabeth Howey, Katie Zazulak, Sharyn Kreuger, David Keane, Anita Riddel and Mary Lou Schmuck. You have all made me feel welcome and valued as a new addition to

the department. Next, I want to express a thank you to all the faculty: Kelly Dore, Lawrence Grierson, Meghan McConnell and Harold Reiter. You have all given me the opportunity to gain valuable experience and I have learned so much from our conversations and collaborations.

Of course that was not all the faculty – I have yet to thank Geoff Norman. I cannot say that I am at a loss for words – if anything, I am overwhelmed by them. However, I will keep this passage short as well. Geoff, you have always managed to make me feel like an equal, even as you explained what I did wrong. You never delivered any bad news without making sure that you had a healthy dose of good news to follow. I have enjoyed all our conversations as I always felt that you valued my opinion – and that is the greatest gift to a student – helping me realize my strengths. I thanked Bruce earlier for helping me change directions in research– but I thank you, Geoff, for changing my outlook. Of course there's also the finer points of helping me get this thesis done in about 2 years.

I thoroughly enjoyed the journey.

Statement of Academic Contribution

In this thesis I present a critique of a popular dual systems model of human reasoning that has influenced models of medical diagnosis and strategies for error reduction.

The first chapter is an introduction and is a general overview of dual process models of reasoning and memory. This chapter was written entirely by me.

The second chapter is a literature review of diagnostic reasoning and dual process models in psychology. In particular I focus on models of memory and reasoning that have influenced research in medical education. This chapter was conceptualized by myself and my supervisor, Dr. Geoff Norman. For the purpose of preparing a manuscript the review paper was primarily written by myself with supplementary information and edits provided by Dr. Norman. The emphasis is on the important implications of each model for education strategies and learning outcomes. This review chapter has been prepared and will be published in *Teaching and Learning in Medicine*, but also serves to set the stage for the research discussed in chapters 3 thru 5.

Each data chapter has been written in a manuscript format and is intended for submission to an academic journal. Therefore, the 3 data chapters are meant to stand alone as well as contribute to the overall understanding of a dual systems approach to medical diagnosis. Specifically, in the third chapter, I address 3 basic assumptions attributed to a default-interventionist dual systems model of reasoning: the assumption that a more rapid response, typically linked to reduced System 2 processing, is more error prone than a slower response; the assumption that interruptions limit available cognitive resources and reduce accuracy compared to task performance with no interruptions and the assumption that reliance on System 2 processing should benefit more experienced physicians to a greater extent than less experienced physicians.

Chapter 3 is an original experiment conducted in collaboration with several co-investigators including Dr. Meredith Young, Dr. Tim Wood, Dr. Danielle Blouin (MD) and Dr. Glen Bandiera (MD). The data were collected at several test sites in Canadian Universities as well as Canadian and U.S. Health care centres. The basic experiment design was originally developed by Dr. Geoff Norman and Dr. Jonathan Sherbino (MD) and is described in two preceding studies already in print (Sherbino et al. 2012; Norman et al. 2013). However, the added experimental factors of distractions and experience developed out of discussions between my Dr. Norman and myself. The computer programming of this study was completed by Elizabeth Howey. Dr. David Keane assisted with making the experiment available online for emergency physicians in Canada and the US. I was solely responsible for organizing, scoring and analyzing the data. The written portion is a manuscript prepared for a special edition on medical education in JAMA and is primarily my writing, with comments from all co-authors. (Manuscript

JAMA13-4184) The final version submitted to JAMA and appearing in this thesis was edited only by myself and Geoff Norman.

In chapter 4, I test the assumption that increased reliance on System 2 processing, through additional reflection and the opportunity to revise a previous decision, should improve accuracy. The results of this study are re-analyzed and re-interpreted for chapter 5 to contribute to an understanding of when participants revised diagnoses and the relationship between response time and diagnostic accuracy. Chapter 4 and 5 are prepared as separate manuscripts, although they address the results of a single experiment.

Chapter 4 is a collaborative project between Dr. Geoff Norman, Dr. Jonathan Sherbino (MD), Dr. Ameen Patel (MD), Dr. Ian Mazzetti (MD), Dr. Amanda Gardhouse (MD), Elizabeth Howey and myself. The design was modelled after previous studies by Geoff Norman and Jonathan Sherbino and included suggestions from Elizabeth Howey. Several co-authors (AG, IM and AP) assisted with recruitment of participants. Elizabeth Howey was responsible for the programming. I was solely responsible for data collection, organization, scoring and analysis. The written portion is entirely my work with edits by Geoff Norman.

For chapter 5, I re-analyzed a subset of data from chapter 4 to address an important concern regarding self-assessment of knowledge and a physician's ability to identify errors. The concept for the analysis was suggested by me and developed further through discussion with Dr. Norman and Dr. Sherbino (MD).

The conceptualization and re-interpretation is my own, as well as the writing and analyses. Geoff Norman assisted with editing the final draft.

The final chapter is a summary of all the findings in the thesis and addresses the main concerns introduced in the introduction. This chapter was written entirely by me and concludes the thesis.

Table of Contents

List of Figures

CHAPTER 3

CHAPTER 4

CHAPTER 5

Preface

This thesis has been prepared in a ‘sandwich thesis’ format, therefore there is a general introduction with 3 distinct data chapters supporting the arguments presented in the introduction. This first chapter is a general introduction to dual process models from psychology and a historical perspective of the impact of dual process models in medical education research. The second chapter expands on the main points in greater detail and provides a current review of the intersection between medical education research and theories from psychology. Chapters 3 thru 5 describe original experiments. Chapter 6 is a general discussion and conclusion to this thesis.

Chapter 1: Introduction

Sandra Monteiro

The mystery of medical diagnosis

> *…the consultant went into the room with an initial hypothesis; after a quick look at the patient, a new hypothesis popped into his mind. This kind of reasoning implies an amazing sequence of psychological events: perceiving the features of the situation, quickly accessing relevant hypotheses, checking for signs and symptoms that confirm and rule out those competing hypotheses, and using related knowledge to guide appropriate investigations and treatment. These events happen at such a high speed that students have trouble understanding the reasoning process and perceive only the outcome.* [1]

Supported by early research in cognitive psychology, a model of the Hypothetico-Deductive method was proposed to describe expertise in medical diagnosis. [2] Similar to the process of the scientific method and theories of problem solving, this model proposed that expert physicians were able to rapidly determine diagnostic hypotheses and then rely on a well developed deductive process to test and adjust each hypothesis. [3] Unfortunately further research [4] showed that novices also used the same approach, and it likely reflected the basic limitations of human information processing rather than expertise. Critically, the focus remained on the processes involved, rather than differences in the knowledge representations between experts and novices.

A second early process-oriented model took a cue from artificial intelligence and proposed that experts used forward reasoning, from data to diagnosis, whereas novices used backward reasoning, from hypothesis to data. [5,6,7] Although the theories captured expertise to some degree, the definitions of forward and backward appeared elusive, and the role of experience in either process was unclear; the distinction between expert and novice emerged as less clear than originally proposed. [8] Importantly, neither backward/forward reasoning models nor hypothesis generation lent themselves to the development of educational strategies for medical education or diagnostic error reduction. In other words, early research on reasoning and problem solving processes did not identify concrete skills that could be taught in medical school and residency.[2] Consequently, several authors developed theories about the knowledge structures that supported initial diagnostic hypotheses. Analytic knowledge structures, included illness scripts [9], semantic axes [10], and propositional networks[6], while non-analytic knowledge structures were primarily described by prior experience and exemplars. [11, 12, 13] These theories do lend themselves to several strategies for teaching and practicing medicine.

Historically, this progression from processes to knowledge structures is interesting as recent medical education research has seen a revival of interest in process based theories of medical reasoning. There is a classic literature that describes medical diagnosis as a categorization task, proposing two forms of parallel categorization processes: analytic and non-analytic. Analytic processes are described as slower and retrieve detailed assessments of features and knowledge of diagnostic rules [14], while non-analytic processes are described as faster and support judgments based on similarity and pattern recognition [12,13]. The primary purpose of early research in this area was to describe these processes as they apply to expert and novice performance and categorization of disease symptoms. [11, 12, 13, 14, 15, 16, 17] A related and more recent perspective is that reasoning may proceed by either a first response system of rapid, memory retrieval processes or a slower, second response system of logical, reasoning processes. Although certain characteristics of this perspective are similar to the theories on analytic and non-analytic processes, this perspective is unique to a popular "dual process" model of reasoning. It is this class of theory that is the subject of my thesis.

Dual process models

Dual process frameworks are the corner stone of many theories in psychology and support a wide variety of research programs. The standard framework assumes independence between two forms of cognitive processes that operate either autonomously (i.e. unconsciously) or require conscious attention and working memory. Dual process accounts that distinguish between conscious and unconscious processes are general enough to help categorize human behaviour based on any number of attributes and can support research in almost any domain including child development and education. Importantly, each dual process theory is based on different assumptions which may determine the applicability of that theory to the real world. In other words, not all dual process theories are the same and even though many are quite good at predicting behaviour, some have not been tested fully. This thesis was developed to address a concern that, although the assumptions of a very popular dual process model of reasoning have not been fully tested in an ecologically valid manner by medical education researchers, it has had a strong influence on medical education programs. [18, 19, 20, 21, 22, 23] This theory has been described as a "default-interventionist" account of cognition. [24] In this thesis I argue against this particular dual process model of reasoning by investigating and critiquing several of the model's assumptions and predictions for medical diagnosis. With this introductory chapter, I begin by distinguishing between this model and another major class of dual process models .

Parallel-Competitive and Default-Interventionist models

Various models that depict human cognition as a duality rely on their own unique terminology and assumptions. As a consequence, there can be a great deal of confusion between models because of different assumptions around parallel and serial processing and the role of memory. For the purposes of this thesis, I rely on the existing literature by distinguishing between models that assume parallel-competitive processing and those that assume serial or default-interventionist processing.[24] Parallel-competitive models are common to memory research and include models of concept formation, categorization and recognition. [24, 25, 26, 27, 28] For example categorization models offer exemplar based processes, which contribute to the identification of a category label based on rapid recruitment of a similar prior exemplar and prototype based processes, which contribute to the identification of a category label based on an assessment of category rules. [25, 26] Theoretically, both forms of knowledge could exist and both forms of processing could occur in parallel. Occasionally, processes may recruit competing evidence for different category labels, in which case a more directed search and analysis of rules is required to determine the correct or most appropriate answer.

Another model that falls within a parallel-competitive framework is a dual process model of recognition memory, which proposes that a previously seen item or person can be recognized through a conscious recollection process that recruits specific details about the prior event and a familiarity based process that is more vague. [28] Again there is a distinction between a conscious, accessible process and one that operates unconsciously. Although both processes are thought to contribute to recognition memory in varying proportions, they are believed to operate in parallel. [27, 28] Competition would occur in the case of disagreement between the information retrieved by each process. Common to memory models is the assumption that a retrieval process can be more or less appropriate for a given task, but there is no assumption that one process is inherently superior.

In contrast, default-interventionist models offer two systems of thinking processes that operate in a serial manner. System 1 is considered rapid and autonomous while System 2 is considered slower and deliberate. [29, 30, 31] Interestingly, System 1 characteristics alone are comparable to parallel-competitive models. [24] That is, System 1 processes are described as parallel, with unlimited processing capacity, and capable of rapid retrieval from memory. [24]

System 2 is then proposed as an additional set of logical processes for human reasoning. [24, 30]

Default-Interventionist Dual Process Reasoning

The development of a default interventionist dual systems model is often linked to Amos Tversky and Daniel Kahneman. Tversky and Kahneman [31, 32] began exploring the idea of dual processes within the field of economics and the assessment of risk and reward. To summarize, Tversky and Kahneman argue that people make decisions by relying either on error prone heuristics (i.e. hasty short cuts) or by relying on normative rules of logic (i.e. accounts for probabilities). [31, 32] The use of time and effort saving shortcuts is associated with a system of processes generally referred to as System 1. Reliance on careful consideration on the other hand is associated with another, more advanced system of processes called System 2.

These time saving shortcuts are referred to as prototype heuristics which are simply abstract rules stored in memory that lead to rapid solutions.[33] Prototype heuristics can be helpful shortcuts, but are also considered biased by the recency and frequency of events.[34, 35] According to this aspect of the dual process account, very frequent events will strengthen the existing prototype in memory, creating a faulty heuristic that is only representative of recent events, thereby biasing subsequent decisions. For emphasis, System 1 processes are linked to errors caused by the recruitment of these faulty or inappropriate heuristics. [31] To complete this dual process account, Kahneman argues that reliance on System 2 processes is initiated in general by a failure of System 1. Although both systems might lead to the same solution, in difficult situations System 1 will either fail to produce an answer or fail to produce a valid answer. One problem with this model is that it does not allow a way to predict difficult situations. So it is unclear how a person might prepare in advance to rely on System 2. The theory also proposes, although the mechanism responsible is not clear, that certain features of the problem will require the methodical approach of System 2 thinking. [31] Proponents of this model believe that with training, System 2 processes can independently assess the results of System 1 processing, identify errors and correct them using a logical sequence of established decision making rules. [31, 34, 35] Therefore, System 2 reasoning is touted as the pinnacle of human accomplishment.

Typically evidence in favour of this model of reasoning comes from logic problems in which the correct normative solution cannot be derived from experience. The key to this theory, then is the assumption that the normative logical solutions are the ideal universal goal; that good logical decisions can only be made after proper consideration of all available information. Following this assumption, any claim that humans are capable of following normative rules of logic, would by default, assume that humans are also capable of accessing and processing all available and relevant information for any situation. However there is a great deal of evidence that humans have a limited cognitive load and working memory.[36] A creative concept of bounded rationality explains how humans can reach past the boundary of a limited cognitive load and recruit System 2 reasoning. [31] Although humans are generally limited by the cognitive architecture of the brain, individual differences in intellect, working memory and generalized training in logic are believed to increase reliance on System 2 and overcome these limitations.

As the default-interventionist dual systems model has gained popularity and even influenced certain curriculum changes in medical education[17, 37], one goal of this thesis is to highlight the implications for medical education of these and other assumptions related to the model.

Medical education strategies for diagnostic error reduction

Several strategies for reducing errors have been proposed including slowing down to increase reliance on System 2, an intense method of reflective practice and training in metacognition. Slowing down is proposed to increase reliance on System 2 processing, thereby reducing the influence of System 1 and allowing the opportunity for increased rational thought. This assumption was recently tested in two large studies with medical residents. [38, 39] In one study participants were required to diagnose medical cases under an extreme time pressure [38] (i.e. while imagining there were many patients waiting in the emergency department) while in the other study, a comparable cohort of participants were allowed ample time and were encouraged to be systematic and attend to all cues. [39] The group given instructions to manage their time, did have faster response times while the other group had slower overall response times. If the assumption that slowing down will increase reliance on System 2 and improve accuracy is correct, then the group with slower response times should have had better performance. A comparative study of the two groups revealed no difference in performance, suggesting that additional response time or time spent in analytical processing a medical case does not necessarily improve performance. Yet, the strategy to slow down and reflect has grown in popularity in medical education. [21]

Another strategy for overcoming the deficiencies of System 1 is to train medical professionals in metacognitive skills aimed at identifying cognitive biases and faulty heuristics. [19, 20] Croskerry has proposed several cognitive biases that can result in diagnostic errors, in the hopes that bringing them to light will help prevent physicians from being misled. [19, 20] In addition to the benefit of raised awareness about cognitive biases, cognitive forcing strategies are believed to help overcome overreliance on System 1.[19, 20] Recent studies have shown that cognitive forcing strategies are not successful at improving diagnostic accuracy, [40, 41] although many still support them.[19, 20, 21]

The concept that further conscious reflection can improve judgement and problem solving can be traced back to the American philosopher John Dewey, who encouraged students to reason through every problem by seeking out multiple hypotheses and solutions. [42] Dewey argued that students and experts alike would benefit from developing generalized, critical thinking skills that would serve to simultaneously enhance learning as well as improve performance. [42] Dewey distinguished between reflective thought and random ideas or beliefs that could not be supported by concrete evidence. [42] However, Dewey's philosophy is not supported by evidence. Although Dewey argued against relying on experience, most problems do not require reflective thought if they are similar to ones experienced before. Therefore, with experience, memory for prior problems and solutions may produce *belief-based* reasoning whereby a solution appears obvious, yet cannot be justified right away. Consistent with Dewey's philosophy, experience dependent or *belief-based* reasoning is considered by many to be less than reliable.[34, 43, 44]

Building on this concept that reflection will improve performance, Sylvia Mamede and colleagues developed a structured form of reflection to assist in medical diagnosis.[45, 46] In several studies this structured form of reflective diagnosis was shown to improve performance when compared to a more rapid diagnostic process. [47, 48, 49, 50] However, in chapters 4 and 5, I note several methodological concerns with the studies by Mamede and colleagues and present my own original data from follow-up experiments, showing that a more realistic form of reflection is not always beneficial and can at times even be harmful. The primary goal of this thesis is to highlight these and other discrepancies between theory, research, and practice of diagnostic reasoning.

The Present Thesis

In the present thesis, I examine in detail the assumptions of a default interventionist dual process model and the implications for strategies in education. I begin with a review of the current literature on medical reasoning and propose alternate theories that are better suited for developing successful training programs. To provide support for my criticisms of a default interventionist dual process model, I test several assumptions of this model in chapters 3, 4 and 5.

In chapter 2 of the current thesis, I argue that a consistent theme in medical education is to focus on optimizing the reasoning processes involved at the time of diagnosis rather than the influence of memory. In contrast, dual process memory models focus on both the encoding and retrieval of knowledge. The type of processing used for learning is entirely dependent on the information being learned and the overall goals. [51, 52] Similarly, the type of process used for retrieving knowledge will depend on the task demands. In order to dissociate between different memory processes, experimental manipulations often induce response bias and affect accuracy. [11, 12, 13] and sadly, the evidence of response bias is often also evidence that memory cannot always be trusted for medical diagnosis.[16] Rather than discourage reliance on memory, I suggest that medical education would benefit more from understanding how memory succeeds and how to ensure that memory has the knowledge to support diagnostic expertise.

The study described in chapter 3 tests the assumption that decreased conscious processing, due to interruptions, increased time pressure or increased expertise will lead to decreased diagnostic accuracy. Junior medical residents and practicing emergency physicians were given instructions to proceed quickly or slowly and were required to complete a diagnostic task while managing interruptions on half the cases. If interruptions reduce available conscious resources, then performance should differ between interrupted and un-interrupted cases. Similarly, if speeded processing reduces System 2 resources, then accuracy should also be lower for cases solved more quickly compared to cases solved more slowly. Finally, we would expect that more experienced physicians will be more accurate, however, the literature also suggests that more experienced physicians rely on pattern recognition more often.[13,53] Therefore, it is possible that emergency physicians asked to go quickly will be more affected by interruptions, reducing their diagnostic accuracy compared to emergency physicians asked to go slowly. Alternatively, less experienced physicians are thought to rely on slower rule based processes in a diagnostic task.[13,53] Therefore it is possible that less experienced physicians (i.e. residents) asked to go quickly will be more affected by interruptions, reducing their diagnostic accuracy compared to residents asked to go more slowly. The results did not support any of these predictions and I discuss this experiment and these results in more detail in chapter 3.

The study reported in chapters 4 and 5 tests the assumption that reflective practice, also intended to increase conscious processing, can improve diagnostic performance and identify errors. There are several studies reporting a positive impact of reflective practice over rapid processing.[47,48,49,50] In all these studies however, a highly structured method was prescribed to residents who were asked to diagnose several medical cases and in most of these studies the cases had been previously manipulated to increase the likelihood of error. Chapters 4 and 5 in the present thesis highlight methodological concerns with previous studies and investigate a more ecologically valid method of reflection with a set of representative and straightforward cases.

Preface

The second chapter is a review paper prepared for a special 25th anniversary edition of *Teaching and Learning in Medicine*. This review presents a comparison of separate research programs in medical education. On one hand are theories influenced by research on reasoning and thinking and on the other hand are theories on knowledge structures and processes in memory. The division between reason and memory research in medical education is influenced by a similar division in psychology. In this review I contrast the two approaches and their implications for strategies to reduce errors in medicine.

Chapter 2: Review paper to appear in *Teaching and Learning in Medicine*

Diagnostic Reasoning and Remembering: How Physicians think and learn

Sandra Monteiro, M.Sc. and Geoffrey R. Norman, Ph.D.

Sandra Monteiro is a Ph.D. candidate in the Department of Psychology, McMaster University

Dr. Norman is a Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University

Corresponding Author

S.D. Monteiro, M.Sc.
McMaster University, Department of Psychology
1280 Main Street West,
Hamilton, Ontario, L8S 4L1, Canada
e-mail: monteisd@mcmaster.ca
phone : 905 525 9140 ext. 23114
fax : 905 572 7099

Abstract

In early studies of clinical reasoning, diagnostic expertise was viewed as the acquisition of general problem solving skills. However a now classic paper by Elstein [1] served as a catalyst to redefining medical expertise by knowledge and experience.

More recently, dual processing models of diagnostic reasoning, have become popular. These theories postulate a fast, unconscious (System 1) component and a slow, logical, analytical (System 2) component. In contrast to work on analytical and non-analytical knowledge as a basis for reasoning, these theories focus on the thinking process, not the nature of the knowledge retrieved. Ironically, this appears to be a revival of an outdated concept. Rather than defining diagnostic performance by problem solving skills, it is now being defined by processing strategy.

The version of dual processing that has received most attention in the literature in medical diagnosis might be labelled a "default / interventionist" model [2,3,4]. This model suggests that a default system of cognitive processes (System 1) is responsible for cognitive biases that lead to diagnostic errors and that System 2 intervenes to correct these errors. Consequently the best strategy for reducing errors is to make students aware of the biases and to encourage them to rely more on System 2.

In this review we briefly discuss the history of research in clinical

reasoning, then focus more specifically on the evidence to support dual processing models. We then examine critically the assumptions of the default interventionist dual process model and the evidence to justify its specific claims.. We then discuss the evidence for several strategies for reducing diagnostic errors and conclude by identifying knowledge gaps about clinical reasoning and provide suggestions for future research.

280 words

Introduction

A classic model of clinical reasoning defined diagnostic expertise by the hypothetico-deductive method, a specialized thinking process or ability to quickly generate a diagnostic hypothesis early in the clinical encounter. However this early model was abandoned in the 1980's, as detailed analyses revealed that both experts and novices generated early hypotheses largely as a strategy to reduce cognitive load, a consequence of the limited size of working memory. Moreover there was very little evidence that the process was an acquired skill [5,6]. Experts did not generate more hypotheses earlier; they just generated better hypotheses. In short, expertise resided in content knowledge, not process.

Patel & Groen then advanced a theory derived from the artificial intelligence literature that experts used "forward-reasoning" from data to diagnoses, and novices were more likely to use "backward-reasoning" from hypotheses to data [7]. However, others have shown that the link may be an artifact of the experimental situation more than a marker of expertise [8]. Moreover, forward and backward reasoning models also do not lend themselves to the development of strategies for medical education or diagnostic error reduction; it is not likely that admonishing students to improve their forward reasoning skills will transform them into experts.

Influenced by research in psychology and sociology, the field shifted and later theories postulated various forms of semantically rich medical knowledge –

semantic axes [9] , probability matrices [10], illness scripts [11], propositional networks [12]. In contrast to the theories that described the process of hypothesis generation, these theories identified possible knowledge representations as sources of hypotheses.

More recently, there has been renewed interest in the process (as opposed to the content) of diagnostic reasoning, especially in the context of diagnostic errors. The dominant model of reasoning is a dual process framework [13], in which reasoning may proceed by a fast, unconscious, retrieval process (System 1) or a more analytical, slow, deliberate and conscious logical process (System 2). Increased reliance on System 2 processing is assumed to improve diagnostic reasoning; the quality of knowledge representations is rarely questioned, as the focus has been less on modeling diagnostic expertise, then on identifying cognitive biases associated with diagnostic errors in reasoning [13,14] .

In the present paper, we contrast this perspective with the view, derived from a different tradition in psychology, that medical diagnosis is a categorization and memory task [15,16] dependent on analytical and experiential knowledge. From this 'memory' perspective, diagnoses are not reasoned so much as they are recognized. In other words, diagnosis is primarily recognition of similar and familiar, previously seen signs and symptoms. The processes involved are complex, but considered secondary and only serve to describe how memory functions in accessing prior knowledge. Learning as improving the quality of knowledge representations is fundamental to this perspective, as it facilitates the recognition of symptoms and disease categories.

By outlining how these two perspectives have developed, we hope to provide a better understanding of what they have to offer. We begin by reviewing the influence of reasoning theories in medical education and the evidence for strategies to improve diagnostic reasoning. We then review the history of research based on memory theories and the evidence for strategies to improve diagnostic memory and learning. Finally, we conclude with a proposal to combine the two programs to improve learning of reasoning in medical education.

Theories of Reasoning Strategies in Psychology

Traditionally, research in problem-solving and decision-making was concerned with general thinking processes. Researchers were less interested in the contribution of memory and experience and more interested in how a solution was derived logically from the data [17]. In order to reduce effects of individuals' prior experiences, these thinking processes were frequently studied using artificially created probability problems (e.g. the famous Linda problem of Tversky & Kahneman [18]). Paradoxically, while the goal was to understand how humans solve problems in the absence of specialized knowledge, most people found these contrived problems difficult to solve without specialized and extensive training in logic and probability [19, 20]. These findings were consistently interpreted as demonstrating that people did not rely on normative logic, but instead recruited simple rules or "heuristics" which are prone to various "cognitive biases" [2]. Hence human reasoning is often described as irrational or subrational because of reliance on heuristics and biases. These result in suboptimal or incorrect responses, particularly for the artificial problems specifically designed to result in a failure of heuristics. Still, this "heuristics and biases" research program, has a strong following in medical education [3, 13, 14, 21, 22, 23].

Dual process framework

The research on heuristics and biases is closely aligned with a dual systems model of reasoning briefly defined earlier. The general description of this model, consisting of two independent systems, has become common knowledge with the popularity of Kahneman's best selling book *Thinking Fast and Slow* [2]. However, this model's theoretical framework, called a "Default-interventionist" model, is quite problematic for practical applications.

Default-interventionist models of reasoning propose separate systems that operate in stages (one after the other) or exclusively (one or the other). [2, 24, 25, 26, 27] In the default-interventionist model, the default mode of reasoning is System 1, which contributes to reasoning by rapidly recruiting heuristics. System 2 on the other hand intervenes when difficulty or bias arises, yet is underutilized because of cognitive limits in attention [2, 24]. As heuristics are associated with biases and are generally viewed as suboptimal strategies, System 1 is similarly viewed as suboptimal and error prone [2, 13, 23, 28, 29]. There is an inherent assumption within this framework that the processes are independent and capable of "*cognitive decoupling*"; the capacity for System 2 to independently assess, manipulate or even inhibit information retrieved by System 1 [20]. This concept of cognitive decoupling is critical to the assumption that control of one or the other system is under conscious influence, leading several authors to conclude that reliance on heuristics can be discouraged through training to consciously recruit System 2 and normative rules, quite independent of content knowledge and experience [13, 22, 23, 28, 29, 30].

Interventions to Reduce Error

In medical education attempts to devise interventions to reduce errors have focused on three broad strategies that are assumed to increase reliance on System 2 processing – slowing down, reflection, and cognitive forcing.

Slowing Down

Kahneman proposes that System 1, while efficient, is inherently flawed, as it contributes to *basic* pattern recognition and does not engage logical reasoning processes [2]. Since System 1 has been characterized by a faster speed of processing compared to System 2, perhaps the simplest intervention is to simply admonish subjects to go slow, be systematic, be thorough, take their time, etc. Such an intervention follows directly from the dual process framework, and should result in increased reliance on analytical, System 2, processes. It also seems common sense that more time spent reasoning through a problem should result in better (more accurate) solutions.

Surprisingly, this general assumption that faster responses are more error prone than slower responses is not supported by research in medical education. One observational study showed that accuracy was associated with *faster*, not slower response times 31. Further experimental studies testing an intervention directed at slowing down did show increased response time, but no effect on accuracy [32, 33]. Finally a study, in which distractions were introduced to hinder System 2 processing during diagnosis, showed a small increase in response time but no effect on accuracy [34]. In light of the discussion of the previous section, all of these interventions amount to manipulation of the amount of cognitive resources allocated to System 2 thinking, yet all have shown no benefit. The idea, that errors can be eliminated by slowing down to increase reliance on System 2 thinking, finds no support from these studies.

Reflective Practice

Another common sense conviction that conscious reflection can improve judgement also fits within a default interventionist framework. Conscious reflection is generally considered slower than intuition and is associated with recruitment of System 2 processes. The concept of reflective thought was popularized in medicine by *The Reflective Practitioner* [35] which has two forms. Increased awareness and introspection while treating a patient comprise reflecting *in* action, while retrospective analyses of decisions comprise reflecting *on* action. [35]

Several studies have examined the influence of reflection on the accuracy of hypothesis generation. [36, 37, 38, 39, 40] Mamede and colleagues manipulated the ambiguity of several cases in a diagnostic task to induce unstructured reflective thought, but found that it did not improve accuracy. [37] A structured reflective method was then tested where the participants listed possible diagnoses, identified critical features, then completed a matrix relating features to diagnoses, and eventually submitted a conclusion. This method, designed to support a deliberate, System 2 approach to diagnosis has had mixed results. [38, 39,40] In one study it reduced errors only for diagnoses where the potential for error was pointed out. [38] In two other studies, although the reflective method did not provide an overall advantage, it did improve accuracy for a subset of cases that the authors identified as difficult. [39, 40] In all these studies however, any advantage provided by the reflective method was quite small and idiosyncratic to the study, and was based on a structured and intensive procedure.

The authors justify the structure and intensity of their procedure by comparing their method to that of deliberate practice proposed by Ericsson and Charness [41] the premise being, if reflection can benefit early practice during the learning process, it may also benefit professional practice. [36] However, the characteristics of deliberate practice have had different effects on novice compared to expert performance. [42] The benefits of a structured program based on the principles of deliberate practice may be limited to the early stages of learning a skill and may prove too cumbersome to maintain once expertise has been achieved. There is also evidence to suggest that a reflexive process is recruited for areas of expertise, while reflection is reserved only for unfamiliar content. [43, 44] It is conceivable that forcing the use of reflective processes prevents a physician from recruiting prior experience, their primary asset, or from feeling like an expert. Moreover, there is evidence that more intuitive diagnoses can be highly accurate [45, 46] calling into question the need for such measures.

Cognitive Forcing Strategies

Consistent with a heuristics and biases research program, Croskerry instructs both novice and expert physicians to proceed slowly and be aware of over 30 sources of cognitive bias. [3, 14] *Cognitive forcing strategies* are a set of warnings that encourage metacognition (a heightened analytic inspection of one's own thought processes) for the purposes of preventing cognitive biases and errors. [14] The promise of thinking strategies that reduce errors is very appealing, but inevitably cognitive forcing strategies have not been shown to be beneficial. [47]

Cognitive forcing strategies devalue the role of experiential and formal knowledge by discouraging reliance on hypotheses to guide the identification of relevant symptoms; the view is that all errors result from several distinct cognitive biases, which, if eliminated, would reduce errors. Defining these discrete elements of cognition as responsible for error is a one sided argument.

As one example, "confirmation bias", the active pursuit of data to support an initial hypothesis, is viewed as a common source of error. However, this process may be an asset. For instance, a working hypothesis affects the relevance of features in a written medical case. [48, 49] As well, Brooks and colleagues [50] demonstrated that feature lists contained more accurate items when participants had a working hypothesis and Norman and colleagues [46] showed that overall diagnostic accuracy was improved when participants followed a form of 'backward reasoning' by starting with a hypothesis. These studies suggest that the active pursuit of supportive data is a useful process, often facilitated by a correct initial hypothesis, and labelled as 'confirmation bias' only in the event that the end result is an error.

Conceptual Problems with biases and the Default-Interventionist model

In summary, the evidence suggests that strategies to increase reflection and awareness of reasoning biases are neither necessary nor sufficient for reducing errors in medicine. One reason may be inadequacy of the model.

There are several fundamental assumptions associated with this model, which are open to critical examination. First, the notion that the two systems are "decoupled" and under conscious control may be incorrect. System 1 processing amounts to retrieval of knowledge from memory, and likely proceeds on a time scale of the order of hundreds of milliseconds. Conscious attempts to "speed up" or "slow down" may well alter the amount of cognitive resources devoted to

System 2, analytical reasoning, but are unlikely to influence the rapid retrieval processes of System 1. Second, it is fallacious to associate cognitive bias solely with System 1. Biases like "confirmation bias", actively seeking information to rule in a diagnosis, "anchoring and adjustment" – changing probabilities of outcomes by adjusting from a baseline, "premature closure"- arriving at a conclusion without accounting for critical information, all arise during the process of gathering additional data and explicitly weighting alternatives, which is a conscious, System 2, activity.

These fundamentally weak assumptions have led to disagreement among different dual process theorists, and even between different arguments from the same authors. [24, 51, 52, 53, 54] While Evans and Stanovich have argued that both types of processing could lead to errors and that it would be incorrect to assume that one process was superior, elsewhere [25, 26, 27] they argue that System 2 is associated with rationality and intelligence, which have been interpreted to mean that System 2 is superior.

Although many authors distance themselves from these assumptions, it is difficult to disentangle the architecture of a default interventionist dual systems model from the assumption that System 1 is inferior and System 2 is superior. This confusion in the literature has led many to consider abandoning dual process models altogether. [51, 54, 55, 56, 57]

Not all dual process frameworks are problematic however. Parallel-competitive models propose separate processes that influence each other and operate simultaneously. [27] Parallel operating frameworks are used in models of memory such as categorization [58, 59, 60] and recognition [61], in which each process is linked to a different form of knowledge; non-analytic experiential knowledge and analytic rule based knowledge. The process dissociation method was developed to calculate the relative contribution of these parallel processes, as within this framework, neither process is considered superior and no task is considered a pure measure of a single mode of processing. [62] These dual process models, focussing on access of memory, may be better suited for application to medical education and strategies to reduce error.

The Role of Memory in Reasoning

In contrast to research on reasoning, research on memory focuses on how information is encoded and recalled to solve problems and maintain goals. Performance on each new problem or task is assessed in relation to prior experience, which may be a characteristic of individual experience, as in research with chess experts [63] or may be experimenter-controlled where participants learn materials specific to the experiment (e.g. lists of words). [17] More important, research on human memory identifies several complex factors affecting recall errors (e.g. false memory, failure to recognize, etc.) and rarely emphasizes reasoning-related errors. For example, the phrasing of a question can influence whether an object is recognized as old or new [64] and contextual information can cue the recall of different information, affecting accuracy. [65, 66] Therefore, memory researchers propose learning strategies to ensure that complex factors such as context, have a reduced impact on recall accuracy. [66, 67]

We will return to memory-based strategies to improve learning, but for now we will focus on discussing two parallel processing models of memory that are relevant to improving our understanding of medical diagnosis: categorization and recognition.

Categorization

The ability to identify objects in our environment is explained by models of categorization and concept formation. There are two classes of categorization models: prototype and exemplar. [58] The main distinction is that, in a prototype model, the role of memory is to retrieve an abstracted, average representation of individual experiences; in the exemplar model, the role of memory is to retrieve relevant closely matched individual experiences and their features.

The current literature on categorization continues to debate the relevance of prototype vs. exemplar models. [68, 69] For the purpose of this review, we do not take a side in this debate, as we believe that both forms of knowledge are important for different aspects of learning. For example, novices are possibly limited to reliance on prototypical or analytic knowledge, having few exemplars or experiences to draw upon. As well, both prototypes and exemplars can be shown to influence categorization, largely as a function of experimental design. [70]

Regardless of the model, accuracy in categorization is undoubtedly related to the quality and quantity of experiences.

We have already mentioned that medical diagnosis has been described as a categorization task [15, 16, 71] and there is sufficient evidence supporting the theory that similar prior experiences influence diagnostic accuracy. However, the identification of previously seen disease categories by reliance on similar prior exemplars is sometimes disparagingly referred to as "basic" pattern recognition (Croskerry, 2009). We propose that recognition is itself a complex process that contributes to categorization and accurate medical diagnosis.

Recognition

A model of recognition memory proposes that a previously seen item or person can be recognized through a recollection process that recruits specific details about the prior event and a familiarity based process that is more vague. [62, 72, 73] For example, recognizing that a pattern of symptoms has been seen before can be accomplished by recalling the details of an identical description in a textbook, by a vague sense that a similar pattern has been seen before on another patient, or by some combination. [61, 67] Importantly, both processes are thought to contribute in parallel to recognition memory, although in varying proportions, and there is no known method for measuring the pure influence of any one process as both processes are equally susceptible to environmental or contextual influences; therefore equally susceptible to error. [62, 74, 75] Learning new information also relies on previously stored experiences, as familiarity can activate previously stored knowledge, while recollection can facilitate associations between old and new information in memory and help define how new information is perceived. [61, 76]

There is clearly some overlap between a memory-based model of reasoning involving experiential and analytical knowledge forms and dual processing models, where System 1 amounts to retrieval of prior experiences and System 2 is related to application of rules relating features to categories. However the emphasis in memory models is not on the process of retrieval but on the nature of the knowledge retrieved.

Memory Based Strategies to Enhance Reasoning

Based on these principles of memory, there are several well established strategies that can improve learning. Here we discuss two strategies that have already been investigated and demonstrated to be successful in a medical education context: test enhanced learning, and mixed practice.

Test enhanced learning

The value of testing for assessment is well known and increasingly so is the value of testing for learning. Testing has been shown to improve retention for material compared to not being tested or studying alone. [76] The testing effect or "test enhanced learning" effect has been demonstrated in several studies by

Roediger and colleagues, both in general and medical education. [76] The effect can be understood within the recognition model, whereby retrieval of information from long-term memory (i.e. recollection) can facilitate the learning of new information. More generally, the act of repeatedly retrieving information strengthens associations in memory by providing contextual variability and increasing meaning. Furthermore, tests that require the production of answers (e.g. short answer, fill-in-the-blank, essay) have led to better retention compared to multiple choice tests, [76] which is consistent with findings in psychology that recollection is improved for items that were generated during learning. [62]

Mixed practice

To the extent that effective reasoning, based on a memory model, is largely derived from an extensive experiential and analytical knowledge base, the emphasis for strategies to improve reasoning skills changes from practising a process to acquiring examples. Although the importance of practice has been emphasized in the literature on deliberate practice, memory models go further than simply examining the amount of practice. In particular, the categorization task of diagnosis requires learning those features that discriminate one category from another. Yet much of our instructional material is based on lists of supporting features, (e.g. the signs and symptoms of myasthenia gravis are…) which are of little value in discriminating myasthenia gravis form multiple sclerosis.

One solution is acquisition of examples from “mixed practice”, where confusable examples are learned together and features to distinguish one from another are examined. In one study, mixed practice was contrasted to “blocked” practice (one category at a time) in the teaching and practice of ECG interpretation skills. Students who learned by mixed practice showed a 17% increase in test scores compared to students who practiced in the traditional blocked method. [77]

A change of perspective is due

Research into sources of diagnostic errors in medicine has had a disproportionately strong influence in recent medical education research. Retrospective analyses of diagnostic errors have estimated that 74% of fatal diagnostic errors arose from reasoning biases like premature closure and availability bias. [78, 79] Although these retrospective analyses may themselves be susceptible to hindsight bias, these error rates attract a great deal of attention. [80] A significant portion of medical education literature has been dedicated to identifying the sources of cognitive errors, rather than identifying the best strategies for learning the prerequisite knowledge to avoid errors.

> *Most of our thinking in clinical practice is of the inductive type, and we should understand its nature and limitations. Inductive thinking is the logic of experience. … Those who have learned well from experience are said to have clinical acumen, and while there is no substitute for experience, we might shorten the road by teaching some of the basic flaws and biases known to be present in everyday thinking.* [81]

Croskerry describes the nature of expertise and correctly acknowledges the importance of experience in reducing errors. However he draws the conclusion that attention should be diverted to teaching 'flaws and biases'. Errors are viewed only as unnecessary and preventable problems that can be eliminated by improving reasoning skills. [3, 13, 14, 81] To this end, diagnostic errors are typically associated with cognitive biases in reasoning rather than gaps in knowledge. [13, 78, 79]

However, this attribution to cognitive biases is only one possibility. A recent retrospective study of chart reviews similar to the Graber and colleagues study, [79] Zwaan and colleagues, [82] using a different taxonomy, did not identify thinking errors derived from cognitive biases. Instead, they found that most errors (58%) were "mistakes" defined as "an intended act, but the physician does not *know* it is incorrect" [emphasis added]. Errors were related to lack of knowledge, not bias.

As well, the old adage that we should learn from our mistakes is quite applicable to medical education. It is through the process of applying knowledge, making mistakes and learning from them that novices become experts. Errors are a necessary element of early learning; focusing learning on the weaknesses in knowledge identified by errors ensures that future opportunities for error are reduced during professional practice. This point has been made forcefully by Eva. [83] The benefits of learning from errors can be gained through the strategic use of testing and mixed practice.

We argue that it is insufficient to target the cognitive biases that are purported to act as sources of diagnostic error while ignoring the value of prior experience and knowledge. If physicians spend more time practicing getting it right, and acquiring experiential knowledge during training, they should make fewer errors later on. Strategies based on explication of cognitive biases have, to date, been shown to be ineffective. [47, 84, 85] A new approach is needed; we suggest incorporating strategies grounded in memory research.

To accomplish this, we cannot rely on a natural distribution of medical problems to expose physicians to the variety of patients, symptoms and disease necessary to form strong knowledge representations. The nature of rare or atypical diseases makes them unreliable learning experiences for medical students and residents. Medical education programs must create the situations necessary for learning by taking advantage of various simulation based learning techniques and applying the principles of deliberate practice early in medical training. [86]

References

1 Elstein, A. S., Shulman, L. S., Sprafka, S. A., & Allal, L. (1978). Medical problem solving: an analysis of clinical reasoning (Vol. 2). Cambridge, MA: Harvard University Press.

2 Kahneman D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

3 Croskerry P. (2003a). Cognitive forcing strategies in clinical decision making. *Annals of emergency medicine*, 41(1): 110-120.

4 Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual review of psychology*, 59, 255-278.

5 Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical Problem Solving A Ten-Year Retrospective. *Evaluation & the health professions*, 13(1), 5-36.

6 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(5), 315-322.

7 Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive science*, 10(1), 91-116.

8 Eva, K. W., Brooks, L. R., & Norman, G. R. (2002). Forward reasoning as a hallmark of expertise in medicine: Logical, psychological, and

phenomenological inconsistencies. *Advances in psychological research*, 8, 41-69.

9 Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Academic Medicine*.

10 Papa, F. J., Shores, J. H., & Meyer, S. (1990). Effects of pattern matching, pattern discrimination, and experience in the development of diagnostic expertise. *Academic Medicine*, 65(9), S21-2.

11 Barrows, H. S., & Feltovich, P. J. (1987). The clinical reasoning process. *Medical education*, 21(2), 86-91.

12 Patel VL, Groen GJ, Scott HM. (1988) Biomedical knowledge in explanations of clinical problems by medical students. *Medical Education*, 22, 398-406.

13 Croskerry P. (2009). Clinical cognition and diagnostic error: applications of a dual process theory of reasoning. *Advances in health sciences education*, 14(1): 27-35.

14 Croskerry P. (2003b). Cognitive forcing strategies in clinical decision making. *Annals of emergency medicine*, 41(1): 110-120.

15 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.

16 Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: the role of experience. *Medical education*, 41(12), 1140-1145.

17 Galotti, K. M. (2010). Cognitive psychology: In and out of the laboratory. Cengage Learning.

18 Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

19 Kahneman, D. (2003). Experiences of collaborative research. *American Psychologist*, 58(9), 723.

20 Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology*, 94(4), 672.

21 de Vries, M., Witteman, C. L., Holland, R. W., & Dijksterhuis, A. (2010). The unconscious thought effect in clinical decision making: an example in diagnosis. *Medical decision making*, 30(5), 578-581.

22 Klein, J. G. (2005). Five pitfalls in decisions about diagnosis and prescribing.BMJ: *British Medical Journal*, 330(7494), 781.

23 Redelmeier, D. A. (2005). The cognitive psychology of missed diagnoses. *Annals of internal medicine*, 142(2), 115-120.

24 Evans, J. S. B., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition Advancing the Debate. *Perspectives on psychological science*, 8(3), 223-241.

25 Evans, J. S. B. (2010). Intuition and reasoning: a dual-process perspective. *Psychological inquiry*, 21(4), 313-326.

26 Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, 7(10), 454-459.

27 Evans, J. S. B., & Elqayam, S. (2007). Dual-processing explains base-rate neglect, but which dual-process theory and how?. *Behavioral and Brain Sciences*, 30(03), 261-262.

28 De Neys, W. (2010). Heuristic bias, conflict, and rationality in decision-making. In Towards a theory of thinking (pp. 23-33). Springer Berlin Heidelberg.

29 Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological bulletin*,116(1), 117.

30 De Neys, W. (2006). Dual Processing in Reasoning Two Systems but One Reasoner. *Psychological science*, 17(5), 428-433.

31 Sherbino, J., Dore, K.L., Wood, T.J., Young, M.E., Gaissmeier, W., Krueger, S., Norman, G.R. (2012). The relationship between response time and diagnostic accuracy. *Academic medicine*, 87: 785-91.

32 Norman, G. R., Sherbino, J., Dore, K., Wood, T.J., Young, M., Gaissmeier W, Monteiro S, Kreuger S. The etiology of diagnostic errors: A controlled trial of system 1 vs. system 2 reasoning. *Academic Medicine*, (in press) 2013.

33 Ilgen, J.S., Bowen, J.L., McIntyre, L.A., Banh, K.V., Barnes, D., Coates, W.C., Druck, J., Fix, M.L., Rimple, D., Yarris, L.M., & Eva, K.W. (2013). The impact of instruction to use first impressions or directed search on candidate diagnostic performance and the utility of vignette-based assessment. *Academic Medicine*. 2013.

34 Monteiro SM, Sherbino JD, Ilgen JS, Dore KL, Wood TJ, Young ME, Bandiera G, Blouin D, Gaissmeier W, Norman GR. Disrupting diagnostic reasoning: The effect of interruptions on the diagnostic performance of residents and emergency physicians. In prep.

35 Schon D. (1983). The reflective practitioner: how professionals think in action. New York, Basic Books.

36 Mamede, S., & Schmidt, H. G. (2004). The structure of reflective practice in medicine. *Medical Education*, 38(12), 1302-1308.

37 Mamede, S., Schmidt, H.G., Rikers, R.M.J.P, Penaforte, J.C. and Coelho-Filho, J.M. (2007). Breaking Down Automaticity: Case Ambiguity and the Shift to Reflective Approaches in Clinical Reasoning. *Medical education*, 41: 1185 - 1192.

38 Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M., van Saase, J. L., van Guldener, C., & Schmidt, H. G. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA: The Journal of the American Medical Association*, *304*(11), 1198-1203.

39 Mamede, S., Schmidt, H.G., Rikers, R.M.J.P., Penaforte, J.C., Coelho-Filho, J.M. 2008a. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. Influence of Perceived Difficulty of Cases on Physicians' Diagnostic Reasoning. *Academic medicine*, 83(12): 1210-1216.

40 Mamede, S., Schmidt, H. and Penaforte, J.C. 2008b. Effects of reflective practice on the accuracy of medical diagnosis. *Medical education*, 42: 468:475.

41 Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American psychologist*, 49(8), 725.

42 Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of experimental psychology applied*, 8(1), 6-16.

43 Eva, K. W., & Regehr, G. (2007). Knowing when to look it up: A new conception of self-assessment ability. *Academic medicine*, 82(10), S81-S84.

44 Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2004). Evidence-based and intuition-based self-knowledge: An fMRI study. *Journal of personality and social psychology*, 87(4), 421-435.

45 Woolley, A., & Kostopoulou, O. (2013). Clinical intuition in family medicine: more than first impressions. *The annals of family medicine*, 11(1), 60-66.

46 Norman, G. R., Colle, C., Brooks, L. R., & Hatala, R. (2000). A disadvantage for searching medical stimuli without a diagnostic hypothesis. *Cognition and instruction*, 17, 433-48.

47 Sherbino J, Kulasegaram K, Howey E, Norman G. (2013). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. *Academic emergency medicine*.

48 LeBlanc, V. R., Norman, G. R., & Brooks, L. R. (2001). Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Academic medicine*, 76(10), S18-S20.

49 LeBlanc, V. R., Brooks, L. R., & Norman, G. R. (2002). Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Academic medicine*, 77(10), S67-S69.

50 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.

51 Kruglanski, A. W. (2013). Only one? The default interventionist perspective as a unimodel—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, 8(3), 242-247.

52 Thompson, V. A. (2013). Why It Matters The Implications of Autonomous Processes for Dual Process Theories—Commentary on Evans & Stanovich (2013). *Perspectives on psychological science*, 8(3), 253-256.

53 Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning*, 19(1), 1-26.

54 Keren, G. (2013). A Tale of Two Systems A Scientific Advance or a Theoretical Stone Soup? Commentary on Evans & Stanovich (2013). *Perspectives on psychological science*, 8(3), 257-262.

55 Custers, E. J. (2013). Medical education and cognitive continuum theory: An alternative perspective on medical problem solving and clinical reasoning. *Academic Medicine*, *88*(8), 1074-1080.

56 Osman, M. (2013). A case study: dual-process theories of higher cognition—Commentary on Evans & Stanovich (2013). *Perspectives on psychological science*, 8(3), 248-252.

57 Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic bulletin & review*, 11(6), 988-1010.

58 Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of experimental psychology: Learning, memory, and cognition*, 24(6), 1411.

59 Smith, E. E., & Medin, D. L. (1981). Categories and concepts (pp. 51-83). Cambridge, MA: Harvard University Press.

60 Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory & cognition*, 22(4), 377-386.

61 Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46(3), 441-517.

62 Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, 30(5), 513-541.

63 Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive psychology*, 31(1), 1-40.

64 Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, 13(5), 585-589.

65 Godden, D. R., & Baddeley, A. D. (1975). Context‐dependent memory in two natural environments: On land and underwater. *British Journal of psychology*,66(3), 325-331.

66 Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, 8(2), 203-220.

67 Neath, I. (1998). Human memory: An introduction to research, data, and theory. Thomson Brooks/Cole Publishing Co.

68 Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological science*, 9(4), 247-255.

69 Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological bulletin*, 138(1), 102.

70 Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: the effects of category size, category structure, and stimulus complexity.

Journal of experimental psychology: Learning, memory, and cognition, 27(3), 775.

71 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(6), 406-416.

72 Yonelinas, A. P., & Jacoby, L. L. (1996). Response bias and the process-dissociation procedure. *Journal of experimental psychology: general*, 125(4), 422-434.

73 Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian journal of experimental psychology*, 48(4), 516.

74 Hockley, W. E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of experimental psychology: Learning, memory, and cognition*, 34(6), 1412.

75 Aggleton, J. P., & Brown, M. W. (2006). Interleaving brain systems for episodic and recognition memory. *Trends in cognitive sciences*, 10(10), 455-463.

76 Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2008). Test‐enhanced learning in medical education. *Medical education*, 42(10), 959-966.

77 Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in health sciences education*, 8(1), 17-26.

78 Chisholm CD, Weaver CS, Whenmouth L & Giles B. (2011). A task analysis of emergency physician activities in academic and community settings. *Annals of emergency medicine*. 58(2): 117-122.

79 Graber ML, Gordon R, Franklin N. (2002). Reducing diagnostic errors in medicine: what's the goal? *Academic medicine*. 77(10): 981-992.

80 Eva KW, Norman GR. (2005) Heuristics and biases--a biased perspective on clinical reasoning. *Medical education*, 39, 870-2.

81 Croskerry, P. (2000). The Cognitive Imperative Thinking about How We Think. *Academic emergency medicine*, 7(11), 1223-1231.

82 Zwaan, L., Thijs, A., Wagner, C., van der Wal, G., & Timmermans, D. R. (2012). Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Academic medicine*, 87(2), 149-156

83 Eva, K. W. (2009). Diagnostic error in medical education: where wrongs can make rights. *Advances in health sciences education*, 14(1), 71-81.

84 Sherbino, J., Dore, K.L., Siu, E., Norman, G.R. (2011). The effectiveness of cognitive forcing strategies to decrease diagnostic error: An exploratory study. *Teaching and learning in medicine*, 23 (1): 78-84.

85 Ark T., Brooks LR & Eva K. (2006) Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Academic medicine,* 81, 405-409.

86 Norman, G. (2012). Medical education: past, present and future. *Perspectives in medical education*, 1 (1), 6-14

Preface

The first data chapter is a manuscript submitted to the Journal of the *American Medical Association (JAMA)* for consideration to be included in a special edition on medical education. The study described in this chapter is a follow-up study to Norman et al., (2013) and Sherbino et al., (2012) and addressed several assumptions of a default-interventionist model of medical reasoning. Specifically, this study addressed the assumption that interruptions would reduce diagnostic accuracy (Chisholm et al., 2011) and the assumption that all physicians, including experienced physicians, would benefit from taking more time during diagnosis (Croskerry, 2009). This study contributes to an understanding of the role of response time, experience and interruptions in modulations of diagnostic accuracy.

Chapter 3:

Disrupting diagnostic reasoning: The effect of interruptions on the diagnostic performance of residents and emergency physicians.

Sandra D. Monteiro, M.Sc., Jonathan David Sherbino, M.D., Jonathan Seth Ilgen, M.D., Kelly L. Dore, Ph.D., Timothy J. Wood, Ph.D., Meredith E. Young, Ph.D., Glen Bandiera, M.D., Danielle Blouin, M.D., Wolfgang Gaissmaier, PhD., Geoffrey R. Norman, PhD., Elizabeth Howey, BSc.

Corresponding Author

S.D. Monteiro, M.Sc.

McMaster University, Department of Psychology

1280 Main Street West,

Hamilton, Ontario, L8S 4L1, Canada

e-mail: monteisd@mcmaster.ca

phone : 905 525 9140 ext. 23114

fax : 905 572 7099

Sandra Monteiro is a Ph.D. candidate in the Department of Psychology, McMaster University

Dr. Sherbino is an Associate Professor, Department of Medicine, McMaster University,

Dr. Ilgen is an Assistant Professor, Division of Emergency Medicine, School of Medicine, University of Washington

Dr. Dore is an Assistant Professor, Department of Medicine, McMaster University,

Dr. Wood is an Associate Professor, Department of Medicine, University of Ottawa

Dr. Young is an Assistant Professor in the Department of Medicine, McGill University

Dr. Bandiera is the Director of Postgraduate Programs, Department of Medicine and PGME, University of Toronto

Dr. Blouin is the Director of Faculty Development for Health Sciences, Department of Emergency Medicine, Queen's University

Dr. Gaissmaier, Chief Research Scientist, Harding Center for Risk Literacy, Max Planck Institute for Human Development

Dr. Norman is a Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University

This study was funded by a grant from the Medical Council of Canada

Key Words: Clinical reasoning, Diagnostic error, Interruptions, Reflection, Diagnosis

words (3025)

Importance This study contributes to an understanding of the impact of experience, time pressure and interruptions on medical reasoning.

Context Diagnostic errors in an emergency department may occur more often as a result of increased time pressure and interruptions. Previous studies have reported that increased time pressure did not result in increased diagnostic error. The effect of interruptions has not been tested. As well, it is unclear whether experience will modulate the combined effects of time pressure and interruptions.

Objective To investigate whether increased time pressure, interruptions, and experience level affect diagnostic accuracy and response time.

Design, Setting and Participants Residents (*N* = 152) were recruited from several Medical Council of Canada (MCC) Qualifying Examination, Part II test sites (McMaster, Toronto, McGill, Ottawa and Queens). Emergency physicians (*N* = 46) were recruited from health centers in two cities (Hamilton and Seattle). Participants were randomly assigned to two instructional conditions (fast (time pressure) or slow (no time pressure)). Case interruptions were manipulated as a within subject factor (present or absent).

Main Outcome Measures Diagnostic accuracy and response time (RT).

Results Accuracy was not affected by interruptions or time pressure. Only experience level was related to diagnostic accuracy; emergency physicians were more accurate (70%) than residents (43%), $F = 235.0$, $p < 0.0001$. Response time (RT) was shorter in the fast condition (55s) than in the slow condition (73s), $F = 21.8$, $p < 0.0001$).

Conclusion Concerns that diagnostic errors are more frequent with increased time pressure and frequent interruptions were not substantiated.

251 words

Introduction

Diagnostic errors are a significant problem in emergency medicine. While estimates vary, several studies have indicated that the prevalence of diagnostic errors in an emergency department (ED) range from 0.6% to 12% (1). Notably, a characteristic feature of any emergency department is the presence of multiple interruptions. Observational studies estimated that emergency physicians experience 10-20 interruptions per hour, which may increase pressure to rely on shortcuts to complete tasks faster (2,3,4,5).

Certainly, there is some evidence that interruptions have a negative impact on performance in other tasks. One observational study involving nurses at two Australian hospitals, reported a positive relationship between errors in drug administration and rate of interruptions (6). Another study using materials related to product management and logistics showed an effect of disruptions on numerical calculation tasks (7), although while interruptions reduced performance on complex tasks, they improved performance on relatively simpler tasks. Application of these findings to medical diagnosis is not clear.

It is also unclear which strategies can reduce the negative impact of interruptions. One broad strategy for reducing diagnostic errors is to essentially increase self-awareness (8). Chisholm (9) argues that Cognitive Forcing Strategies (CFS) – generalized techniques for increasing self-awareness - should be employed to deal with disruptions:

> *There is evidence supporting the negative effect of interruptions on task performance and subject perception of stress and we agree with the speculation by Westbrook et al (7) that interruptions likely cause emergency medicine providers to compensate through task short cuts or failure to reengage in the task. This suggests that teaching cognitive forcing strategies to reorient after an interruption, and especially after a break in task, may be beneficial to ED providers. (p. 121)*

However, two recent studies have failed to show any benefit of CFS in reducing diagnostic errors (10,11), suggesting that generalized strategies are quite inappropriate. Another strategy for reducing error is to essentially slow down the diagnostic reasoning process. Kahneman (12) writes:

> *"How can we improve judgments and decisions…? The short answer is that little can be achieved without a considerable investment of effort. …System 1 is not readily educable…."*
>
> *The way to block errors that originate in System 1 is simple in principle: recognize that you are in a conceptual minefield,* ***slow down****, and ask for reinforcement from System 2". [emphasis ours]* (p. 417)

Kahneman (12) refers to the dual systems model of reasoning which relates errors to the use of inappropriate processing (13,14). This model suggests that System 1 engages fast, non-analytical processes that require few cognitive resources whereas System 2 engages slower, deliberate analytical processes that require a great deal of cognitive effort. It has been theorized that overreliance on System 1 processes may result in cognitive biases that lead to diagnostic errors (14). Conversely, the slow, analytical thinking strategies of System 2 are thought to have a central role in correcting the errors resulting from the biases of System 1.

Three recent studies tested this theory and showed that instructions did have an influence on response time in a diagnostic task, yet had no impact on diagnostic error rates, as accuracy was equivalent for cases diagnosed fast or slow (15,16,17). However, these studies took place in a quiet, interruption free environment, so do not provide a direct test of whether strategies designed to encourage slow, deliberate reasoning will directly reduce the impact of interruptions on clinical reasoning. Moreover, it may be that less experienced physicians, who rely more on analytical processing than more experienced physicians, (18), may be more vulnerable to the effect of interruptions.

To address these issues, the present study investigated the effect of time pressure, experience and interruptions on accuracy in a diagnostic task. Similar to previous studies (15,16), participants were asked to diagnose a series of general medicine cases, and were randomly assigned to receive instructions that encouraged either a faster or slower approach. The critical additions of the present study design were 1) the inclusion of more experienced emergency medicine physicians and 2) the requirement that participants manage interruptions during the study. We addressed the following research questions:

1) What is the effect of increasing time pressure through interruptions and instructions to proceed quickly on response time and diagnostic accuracy?

2) What is the effect of experience on response time and diagnostic accuracy?

METHOD

Design

Emergency physicians and junior residents were recruited to complete a diagnostic task on a computer. Participants were randomly assigned to receive instructions to diagnose faster and within a shorter time limit (30 minutes) or more carefully within a longer time limit (45 minutes). The specific instructions will be outlined in detail below, however, the between-subjects condition created by the different instructions will be referred to as “fast” or “slow”. Interruptions were presented on half the cases in one of 2 counterbalanced orders. Therefore this study was a randomized mixed model design with two between-subjects factors of instruction (fast or slow) and experience level (emergency physician or resident) and one within-subject factor, interruptions (present or absent).

Procedure

Participants

Recruitment and Setting

Emergency Physicians

Forty-six emergency physicians were recruited from health centers in Hamilton, Ontario (*n* = 21) and Seattle, Washington (*n* =25). Emergency physicians received an e-mail inviting them to participate by authors JS (Hamilton) and JI (Seattle). The e-mail provided a web link through which the computer-based experiment could be downloaded directly to a personal computer. These physicians participated outside of work hours and in a quiet location of their choice. Upon following the study link, participants were asked to provide electronic consent, and then diagnose a series of clinical vignettes. Upon completion of the study, participants were offered an honorarium of $50.

Residents

Residents with a minimum of PGY2 training (*n* = 152) were recruited across Ontario and Quebec (Canada) following completion of the Medical Council of Canada Qualifying Examination Part II (MCCQE Part II). Medical residents were recruited at five sites: McGill University (*n* = 55), Queen's University (*n* = 21), University of Toronto (*n* = 30), University of Ottawa (*n* = 23) and McMaster University (*n* = 23). Participants provided written consent prior to diagnosing a series of clinical vignettes. Upon completion of the study, participants were offered an honorarium of $30.

Instructions:

Diagnostic Task

The experiment took place in a computer laboratory setting. Participants were randomly assigned to receive one of two sets of instructions (described below), and asked to diagnose 20 general medicine cases as accurately as possible.

Participants in the slow condition were instructed to be careful and thorough. They were shown a progress bar in the upper right hand corner of the screen (to show how many cases had been completed) and received these instructions:

> *For the next 45 minutes you will be presented with information about 20 general medical cases. We are asking you to provide a single best diagnosis for each one of them. Make sure you consider all the data before you arrive at your diagnosis.*

Participants in the fast condition were encouraged to be fast through two interventions. First, a digital timer was present on the top-right corner of the computer screen while a medical case was visible (contrasted to the 'progress bar' described above). Second, participants received these instructions:

> *Imagine that you are in a busy emergency department. As usual there is a large backlog of patients. You are about to see a series of 20 general medicine cases, and you have a limited time (that is about 30 minutes) to see them all. It's possible you won't be able to complete all the cases, but work as quickly as you can without sacrificing accuracy.*

All participants received the following instructions about entering a diagnosis:

> *All information for a single case will appear on one screen. You will need to click on a button to go to the diagnosis screen. You may spend as much time as you wish reading the case information, but once you advance to the diagnosis screen, you cannot go back. Your final diagnosis should be described in 5 words or less.*

Interruptions

All participants (emergency medicine physicians and residents) were informed they would have to attend to occasional interruptions. Two kinds of interruptions occurred: auditory and visual. The auditory interruption occurred first; after a delay the visual interruption occurred. The visual interruption was a multiple choice style question that replaced the medical case on the screen until the correct answer was selected. Interruptions occurred in exactly half the cases in a counterbalanced design. Participants were given the following instructions on how to manage the interruptions:

> *In approximately half the cases, you will be "interrupted". You will hear a simulated page asking you to call a line extension. You will need to remember the line extension and recall it at a later point. We ask that you please remember the line extension relying only on your memory. Please do not write down the extension using pen and paper. In addition, the screen will switch to a multiple-choice question, which you must answer correctly before going back to the case. After you have diagnosed the case, you will be asked to enter the line extension from the page. Please do not rely on any memory aids.*

Each "interrupted" case contained both interruptions. Previous studies (3,4) have indicated that an emergency physician can anticipate 10-20 interruptions per hour, which aligns well with our design (assuming the physician can see about 10 patients per hour, this amounts to 1-2 per case).

Materials

Diagnostic Task

The general medical cases selected were a subset of those used in a previous study (13) and were presented on a computer screen. One half of the screen presented written descriptions of the primary complaint, a relevant patient history and relevant test results. The other half of the screen displayed an image presenting findings of the physical exam and/or investigation (e.g. CT scan, x-ray). The measurement units were appropriate to the country of practice of the physician (all residents tested were presented with Canadian units of measurement).

Audio Interruptions

At predetermined time intervals during the vignettes, participants were interrupted with audio prompts that asked them to remember phone extensions. These extensions (i.e. numbers) were a string of four randomly selected digits from 1 to 5 and were different for each interrupted case. Residents who conducted the experiment in the computer lab were equipped with ear-bud style headphones to facilitate hearing the pager number. Subjects were prompted to enter these numbers in a separate screen after they provided the diagnosis for the interrupted case.

Visual Interruptions

Visual interruptions consisted of multiple-choice questions about general medical knowledge, designed by JS. An example is:

> The gold standard test to diagnose aneurismal subarachnoid haemorrhage is:
>
> a. Non-contrast CT of the brain
>
> b. CSF examination
>
> c. Skull radiographs

Measurement of Performance

Diagnostic Task

Using a previously-developed rubric (17), free text responses to the diagnostic cases were scored on a three-point scale, where 0 was incorrect, 1 was partially correct and 2 was completely correct.

Interruptions

To encourage participants to attend to the auditory interruptions, they received immediate feedback on the accuracy of their pager number entries. The percent of correct pager reports was calculated. When presented with the multiple choice question, participants were not permitted to return to the interrupted case until they selected the correct answer. The response time for each multiple-choice question was recorded.

Statistical Analysis

Case response times were computed for each participant, excluding the time taken to manage the visual (multiple choice) interruptions. Audio (pager) interruptions lasted four seconds, however, the case was still visible and it was still possible to review the case material. Therefore, in the primary analyses, the time required to listen to audio interruptions was not subtracted from the time taken to complete a case.

An overall accuracy score was computed for each participant based on the sum of the completed case scores (maximum possible = 40), divided by the number of cases completed. In this way average scores were adjusted for participants who did not complete all 20 cases. The overall accuracy score and case response times were submitted to separate univariate ANOVAs with two between-subject factors, a) Level of experience – resident or emergency physician and b) Study instructions – fast or slow.

An average accuracy score and average response time was calculated for interrupted and uninterrupted cases and submitted to separate repeated measures ANOVAs with one within subject factor of interruptions (present or absent) and two between subject factors of experience and instructions on response time and accuracy. The alpha level was 0.05 for all analyses.

Average case accuracy scores for emergency physicians were submitted to a preliminary one-way ANOVA to test the effect of country of practice (Canada vs. U.S.). No differences were found, ($p > 0.5$) so subsequent analyses were collapsed across this factor.

RESULTS

Effect of Experience Level on Accuracy and Response Time

Forty-five of 46 (98%) physicians, and 136 of 152 residents (89%) completed all 20 cases. The average accuracy scores are displayed in Figure. 1. Overall diagnostic accuracy for emergency physicians (70%) was significantly higher than for residents (43%) ($F = 234.0$, $df = 1$, $p < 0.0001$). Average response times for physicians and residents for each instruction condition are presented in Figure 2. On average, residents had longer processing times (69 seconds) to diagnose cases than experienced physicians (58 seconds), ($F = 9.0$, $df = 1$, $p < 0.005$).

Effect of Instructions on Accuracy, Response Time and Cases completed

There was no overall effect of instructions on accuracy and no interaction with experience level. For emergency physicians, accuracy under the fast condition was 70%, and 71% under the slow condition; for residents, accuracy was 43% under both conditions ($F < 0.05$, $df = 1$, $p = .84$), which is consistent with previous studies using similar materials and resident participants (14,15). Participants in the fast condition had significantly shorter response times (55 seconds) than participants in the slow condition (73 seconds) ($F = 22.2$, $df = 1$, $p < 0.0001$). As a result of taking longer to diagnose cases in the slow condition, fewer (93%) participants in the slow condition completed all cases than participants in the fast condition (97%) despite having an additional fifteen minutes available. A power calculation, based on an alpha of 0.05 and a beta of 0.20 indicated that the study would be able to detect a difference in accuracy of 4%.

Effect of Interruptions on Accuracy and Response Time.

There was no effect of interruptions on accuracy in either cohort (Residents: Interruptions, 43%, No interruptions 44%; Emergency physicians: Interruptions 70%, No interruptions 70%), $F = 0.01$; $df = 1$, $p = 0.9$. When cases were interrupted, participants took longer to report a diagnosis (67 seconds) compared to having no interruptions (60 seconds) (F=33.9, df = 1, p < 0.01). Adjusting the average response time by subtracting the time required to listen to the audio message, still revealed a significant 3-second slowing for cases with interruptions, (F = 6.14, df = 1, p = 0.01).

The overall accuracy for reporting the pager number was 84% for emergency physicians and 74% for residents, t = 2.6, df = 45, p < 0.01. Emergency physicians were faster (6 sec.) at identifying the correct answer to multiple-choice questions than residents (9 sec.), (t = -4.1, df = 45, p < 0.001).

DISCUSSION

It has been suggested that excessive interruptions during case management can have serious consequences, as physicians may be unable to automatically re-engage previous tasks (3,4,5). Within the context of diagnostic reasoning however, we found no evidence that interruptions adversely affected accuracy for either more experienced emergency physicians or less experienced residents. By contrast, the results of the present study suggest that emergency physicians and residents are quite efficient at dealing with interruptions. While the results appear counterintuitive, they are consistent with research in psychology related to task switching, which compares performance for repeating the same task to switching between two or more tasks (19). Typically, switching tasks results in longer response times, but if tasks are not confusable, such as diagnostic and administrative tasks, there is no impact on accuracy (19).

These results stand in contrast to predictions of a dual processing model that presumes errors can be mitigated by slowing down and engaging more analytical resources (12,13), and adds to the growing evidence that rapid processing does not affect diagnostic accuracy (15,16).

Although the medical cases included in this study were reasonably representative of emergency medicine, the study has some limitations. An obvious concern is that the findings are based on diagnosis from written cases, which may not be as rich as the “real world” of the emergency department. However, some of the results observed in this study support the validity of the approach. The emergency physicians outperformed junior residents, with higher diagnostic accuracy and faster response times. As well, participants responded to instructions to go fast or slow. Moreover, recent studies have shown that a) learning from written cases leads to clinical performance equivalent to learning from video cases or standardized patients (20) and b) assessment based on low fidelity simulations (a laptop presentation of an ECG) are equivalent to evaluation with real patients (21).

Finally, it has been suggested that with increasing experience physicians learn to rely on heuristics or pattern recognition rather than analytic reasoning (22). According to some authors, (12,13), such reliance on cognitive shortcuts or heuristics is associated with increased error rates. The present study found no loss of diagnostic accuracy for experts. Quite the converse, emergency physicians were faster, more accurate overall, and more efficient at dealing with interruptions.

CONCLUSION

Dual processing models of reasoning, as described by Kahneman (13) and others (12) suggest that slowing down and consciously engaging analytical (System 2) helps to reduce the influence of cognitive biases and thereby improve diagnostic performance. The results of the present study run counter to this hypothesis by demonstrating that such influences—whether induced by testing instructions, level of expertise or interruptions—do not adversely affect diagnostic performance. Consistent with past studies that aimed to induce reflection (20-22), longer response times were not associated with better performance. Future research must aim to understand the role of factors that have been demonstrated to influence diagnostic accuracy, such as experience.

ACKNOWLEDGEMENTS

Each author contributed substantially to the development of this study. Dr. Sherbino, Dr. Ilgen, Dr. Young, Dr. Wood, Dr. Bandiera, Dr. Blouin and Sandra Monteiro assisted in recruiting participants and collecting all data at the various sites. Sandra Monteiro managed, collated and analyzed all data in this study and takes responsibility for the integrity and accuracy of the analyses.

Elizabeth Howey BSc. contributed to programming and designing the experiment. David Keane PhD, a Research Associate at McMaster University, assisted with the development of the online format of the experiment and managed the storage and security of online data from emergency physicians.

Disclaimers

None of the authors report any conflict of interest.

Ethical Approval

The study received approval from medical ethics boards at McMaster University, McGill University, University of Toronto, Queen’s University and University of Ottawa.

Funding / Support

The study was supported by grants from the Medical Council of Canada. The funding organization was not directly involved in the preparation of this manuscript or in any part of the design or conduct of the study. The funding organization was not involved in data collection, management, analysis or interpretation of the data and did not review this manuscript prior to submission.

REFERENCES

1) Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *AJM.* 2008; 121 (5 Suppl) S2.
2) Chisholm CD, Collison EK, Nelson DR, Cordell WH. Emergency department workplace interruptions: Are emergency physicians "interrupt-driven" and "multitasking"? *Acad Emerg Med*. 2000; 7(11): 1239 – 1243.
3) Chisholm CD, Dornfeld AM, Nelson DR, Cordell WH. Work interrupted: A comparison of workplace interruptions in emergency departments and primary care offices. *Ann Emerg Med*. 2001; 38(2): 146-151.
4) Graber ML, Gordon R, Franklin N. Reducing diagnostic errors in medicine: what's the goal? *Acad Med*. 2002; 77(10): 981-992.
5) Westbrook JI, Woods A, Rob MI, Dunsmuir WTM, Day RO. Association of interruptions with an increased risk and severity of medication administration errors. *Archives of Internal Medicine*. 2010; 170(8): 683.
6) Westbrook JI, Coiera E, Dunsmuir, WTM, Brown BM, Kelk N, Paolini R, Tran C. The impact of interruptions on clinical task completion. Qual Saf Health Care. 2010; 19(4): 284 – 289.
7) Speir C, Valacich JS, Vessey I. The influence of task interruptions on individual decision-making: an information overload perspective. *Decis Sci*. 1999; 30(2): 337-360.
8) Croskerry P. Cognitive forcing strategies in clinical decision making. *Ann Emerg Med*. 2003; 41(1): 110-120.

9) Chisholm CD, Weaver CS, Whenmouth L, Giles B. A task analysis of emergency physician activities in academic and community settings. *Ann Emerg Med.* 2011; 58(2): 117-122.

10) Sherbino J, Dore KL, Siu E, Norman GR. The effectiveness of cognitive forcing strategies to decrease diagnostic error: An exploratory study. *TLM*. 2011; 23 (1): 78-84.

11) Sherbino J, Kulasegaram K, Howey E, Norman G. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. *Acad Emerg Med* (in press) 2013.,

12) Kahneman D. Thinking, fast and slow. Farrar, Straus and Giroux. 2011.

13) Evans J. St. B. Dual processing accounts of reasoning, judgment and social cognition. *Ann Rev Psyc*. 2008; 59: 255-278.

14) Croskerry P. Clinical cognition and diagnostic error: applications of a dual process theory of reasoning. *Adv HSci Ed*. 2009; 14(1): 27-35.

15) Norman GR, Sherbino J, Dore K, Wood TJ, Young M, Gaissmeier W, Monteiro S, Kreuger S. The etiology of diagnostic errors: A controlled trial of system 1 vs. system 2 reasoning. *Acad Med*. (in press) 2013.

16) Sherbino J, Dore KL, Wood TJ, Young ME, Gaissmeier W, Krueger S, Norman GR. The relationship between response time and diagnostic accuracy. *Acad Med.* 2012; 87: 785-91.

17) Ilgen JS, Bowen JL, McIntyre LA, Banh KV, Barnes D, Coates WC, Druck J, Fix ML, Rimple D, Yarris LM, Eva KW. The impact of instruction to use

first impressions or directed search on candidate diagnostic performance and the utility of vignette-based assessment. *Acad Med*. (in press) 2013.

18) Eva KW, Link CL, Lutfey KE, McKinlay JB. Swapping horses midstream: factors related to physicians' changing their minds about a diagnosis. *Acad Med. 2010* 85(7): 1112-1117.

19) Monsell S. Task Switching. *Trends in Cognitive Sciences*. 2003; 7(3): 134:140.

20) La Rochelle JS, Durning SJ, Pangaro LN, Artino AR, van der Vleuten CP, Schuwirth L. Authenticity of instruction and student performance: A prospective randomised trial. *Med Educ.* 2011; 45(8): 807-817.

21) Hatala R, Issenberg SB, Kassen B, Cole G, Bacchus CM, Scalese RJ. Assessing cardiac physical examination skills using simulation technology and real patients: a comparison study. *Med Educ.* 2008; 42(6): 628-636.

22) Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ*. 2004; 39: 98-106.

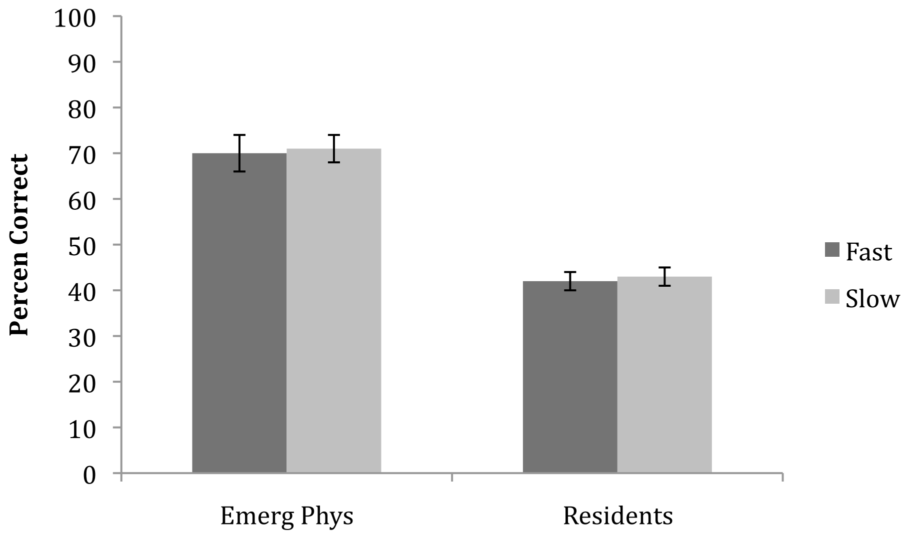


Figure 1. Average case accuracy presented for emergency physicians and residents.

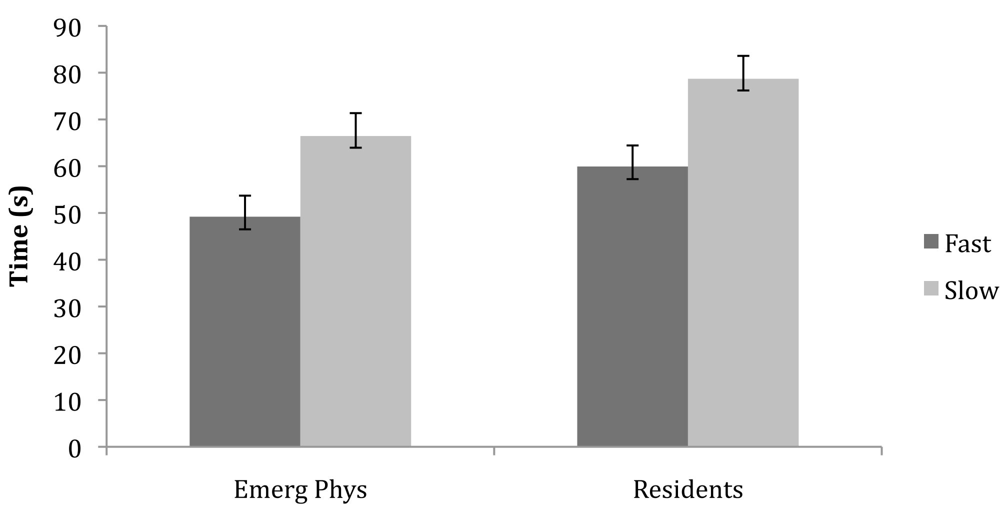


Figure 2. Average case response times for emergency physicians and residents.

Preface

The second data chapter addresses the assumption that reflection and structured hypothesis testing increase reliance on System 2 processing, and consequently improve diagnostic accuracy. Sylvia Mamede and colleagues have previously argued in favour of incorporating reflection during diagnosis. In particular Mamede and colleagues developed a structured rubric that guided physicians through a written medical case (Mamede et al., 2007, 2008a, 2008b, 2010). In several studies, Mamede and colleagues concluded that their structured form of reflection led to improved diagnostic accuracy compared to an unstructured and non-analytic, free-form method. However, in all these studies, several methodological concerns arise, calling their interpretation of the results into question. In particular, the structured rubric has little ecological validity and cannot be generalized to various medical settings since it is a time and labour-intensive method. The study presented in chapter 4 compares the diagnostic performance of physicians relying on a unstructured reflective approach to a unstructured non-reflective approach. This study contributes to an understanding of the role of experience and reflection in modulating diagnostic accuracy.

Chapter 4: Reflecting on Diagnostic Errors: Does it Help to Take a Second Look?

Running Title: Reflecting on Diagnostic Errors

Sandra Monteiro MSc., Geoffrey Norman PhD, Jonathan Sherbino MD, Ameen Patel MD, Ian Mazzetti MD, Amanda Gardhouse MD, Elizabeth Howey BSc

Sandra Monteiro is a Ph.D. candidate in the Department of Psychology, McMaster University

Dr. Norman is a Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University

Dr. Sherbino is an Associate Professor, Department of Medicine, McMaster University

Dr. Patel is a Professor, Department of Medicine, McMaster University

Dr. Mazzetti is and internal medicine resident, Department of Medicine, McMaster University

Dr. Gardhouse is an internal medicine resident, Department of Medicine, McMaster University

Elizabeth Howey is a research assistant for the Program of Education, Research and Design, McMaster University

Corresponding Author

S.D. Monteiro, M.Sc.
McMaster University, Department of Psychology
1280 Main Street West,
Hamilton, Ontario, L8S 4L1, Canada
e-mail: monteisd@mcmaster.ca
phone : 905 525 9140 ext. 23114
fax : 905 572 7099

Key Words: Clinical reasoning, Diagnostic error, Reflection, Diagnosis

words (3918)

Abstract

Background Rapid “pattern recognition” has been postulated to be associated with several cognitive biases, and reflective practice has been proposed as a strategy to reduce diagnostic errors. The results of several studies indicate that additional reflection may improve accuracy in some situations. However those results are not generalizable as the studies employed a methodology with low ecological validity and generally focused on junior residents.

Objective To investigate how experience level and naturalistic reflection, with or without case summaries, will affect performance.

Study Design Residents from post-graduate years 1, 2 and 3 were randomly assigned to receive instructions to take more or less time during initial diagnosis. Participants asked to use less time were then allowed to reflect on each case and decide whether to retain or revise their previous diagnosis. Half of these participants were allowed access to the full case summary, while the other half saw limited information. This was a mixed design with 2 primary between subject factors (post-graduate year and instructions) and a secondary between subject factor of case accessibility.

Main Outcome Measures Diagnostic accuracy and response time was collected for each case.

Results Participants did respond faster when instructed to take less time, (97sec. vs. 111sec.) $F = 4.1$, $p < 0.05$, however this did not affect overall accuracy (60%

vs. 64%, $p > 0.1$). The best predictor of accuracy was post-graduate year, as senior residents were more accurate (65%) than junior residents (55%), $p < 0.05$. The opportunity to reflect further showed a small, but marginally significant, benefit (60.5% vs. 61.5%, $F = 3.9$, $p = 0.06$).

Conclusion The results suggest that reflection alone does not improve performance. Experience plays a more important role in determining diagnostic success.

Introduction

A popular dual systems model of reasoning proposes separate systems of processes labelled System 1 and System 2 that operate within a distinct *default-interventionist dual-process* framework. [1,2] This framework features prominently in research on reasoning and bias and is described in detail by Evans and Stanovich.[1] Within this framework, System 1 is theorized to operate reflexively or by *default*, recruiting autonomous memory processes that are described as vulnerable to cognitive biases. [1,3] System 2 processes are thought to *intervene* at a later stage with deliberate rational thought, especially in the event of difficulty or an error. [1,3] Critically, System 2 processing is considered superior primarily because it facilitates *cognitive decoupling*; a form of higher consciousness that works *with* information from memory rather than being controlled *by* information from memory.[1,4] Understanding this framework is critical to understanding how this default interventionist dual systems model has influenced medical education.

Several authors reference this dual process model of reasoning to support strategies that they believe will reduce cognitive errors in medical diagnosis. [5,6,7,8,9,10,11,12,13,14,15,16,17,18] Cognitive errors have been postulated to be related to faulty reasoning or thinking and since System 1 is hypothesized to recruit fast and reflexive, easily biased reasoning processes, then System 1 is considered the most likely source of cognitive errors. [3,8] Consequently, the proposed solution to reduce these types of errors is to develop strategies that take advantage of the concept of cognitive decoupling: slow down, reflect on information retrieved from memory and rely on System 2. [3,8,10] Strategies that have been proposed include cognitive forcing strategies [8,9] and reflective practice. [10,11,12,13,14]

Cognitive forcing strategies encompass a set of metacognitive skills aimed at helping physicians recognize the limits of their memory and knowledge and think analytically about their own thoughts.[8,9] When tested experimentally, this strategy did not provide any advantage.[19,20]

Another strategy relies on a slightly different form of metacognition and incorporates the components of deliberate or reflective practice. [10,11,12,13,14,21] Building on the components of reflective practice, Mamede and colleagues designed a structured reflective method to guide physicians through uncertainty in diagnosis by helping them identify symptoms described in the medical case, list differential hypotheses and apply probability and confidence ratings in order to select the best diagnosis. [10,11,12,13,14] Proponents of this strategy also link it to classic educational philosophies of authors like John Dewey [22] and Donald Schon. [23] who have separately encouraged a form of reflective practice during learning; however, reflective practice has not previously been tested experimentally. To provide evidence in support of this strategy, Mamede and colleagues investigated the impact of a reflective method on diagnostic accuracy.

Using a diagnostic task with written medical cases, Mamede et al. compared the accuracy of cases diagnosed using the reflective method and cases diagnosed through pattern recognition.[12] Overall accuracy for cases diagnosed using the reflective method was not significantly different than cases diagnosed through rapid pattern recognition, however the authors did note an advantage of reflection for cases previously identified as difficult. [12] However, it remains to be seen whether the improvement resulted from the specific intervention related to the reflection exercise or simply from additional time spent on the task. When a similar study was conducted with medical students, residents and faculty, no benefit of reflection was found as only experience level determined performance.[24]

In another study, Mamede et al., tested the impact of the reflective method on diagnostic accuracy where there was an experimentally induced availability bias.[10] In phase one of the experiment, first and second year residents were asked to confirm an experimenter-provided diagnosis for several internal medicine cases. In phase 2, participants were presented with new cases and asked for a first impression diagnosis. Critically, half of the new cases mimicked a case with a different (incorrect) diagnosis from phase one. If participants were influenced by these similar prior cases during test, accuracy would be reduced compared to participants who did not rely on similarity. Indeed this is what they found, as residents made more errors on biased cases (i.e. availability bias). In the last phase of the experiment, participants were then asked to review those biased cases using the reflective method. The authors reported improved accuracy from added reflection, consistent with the view that errors resulted from a cognitive bias (availability) that could be reduced by additional reflection. [8,16,17]

However, in this study, only the biased cases were targeted for reflection. It is possible that residents responded to experimenter demands and presumed they should change their diagnosis. As well, the final review phase allowed participants full access to all the information presented in the case description so participants could more easily compare how well symptoms supported their previous diagnosis or differential diagnoses. Perhaps if they had limited access to information during the review process, as would be the case in clinical practice, reflection would have proved less beneficial.

Thus far, the studies that investigated the impact of reflective practice relied on a prescribed method of reflection. [11,12,24] Participants were given detailed instructions about evaluating and revising diagnoses which proved selectively beneficial for difficult cases in some studies.[11,12] Additionally, previous studies focused on the performance of junior residents [11,12,24,25,26] Prescribing a method of reflection for more experienced residents may have no effect at all or may be deleterious to performance. Indeed, one study that included participants of varied levels of experience found no benefit of reflection.[24] Therefore, it is worthwhile investigating how residents with different levels of training respond to a more naturalistic opportunity to reflect and whether a more naturalistic form of reflection will improve diagnostic accuracy.

In the present study, we were interested in answering the following questions:

Research Questions

1) Will instructions to approach diagnosis more slowly or more quickly affect diagnostic accuracy?
2) Are more senior residents more accurate overall
3) Are senior residents more able to use reflection more effectively than junior residents?
4) Will access to case details during reflection improve performance compared to having limited access?

If, as Kahneman [3] suggests, slowing down will improve performance, residents instructed to use less time, should rely more on System 1 in their responses and make more diagnostic errors than residents instructed to take more time. Although previous research suggests that this is not the case, [24,25,26] those studies did not go on to examine the effect of physicians' own decisions to reflect further and revise a diagnosis. Second previous studies did not examine the interaction between experience and reflection to determine if more experienced physicians are just generally more accurate or if they also employ reflection more effectively. If reflection alone provides an advantage, than reflection should provide a measurable benefit on all cases and for junior and senior residents alike. Finally, we examine whether the presence of a case summary facilitates reflection and results in higher accuracy compared to reflection without case details.

Method

Design

The study was a randomized mixed design comparing between group effects of instructions, resident level and access to case details and within-group effects of reflecting again on a prior case diagnosis.

Setting

Internal medicine residents from the teaching hospitals associated with McMaster University in Hamilton were invited to participate. The test sites were the Juravinski Cancer Centre, St. Joseph's Hospital and McMaster Children's Hospital. The study was conducted by SM using laptops set up in conference rooms within each of the test sites.

Participants

Recruitment

Residents were informed of the study by e-mail and invited to participate during the hour before morning rounds or during the lunch hour. In most situations, residents were compensated for their time with breakfast or lunch. All residents from post-graduate year 1 to 3 were eligible to participate. We recruited a total of 65 residents; 27 in post-graduate year (PGY)1, 15 in PGY 2 and 23 in PGY 3. Participants were from different specialities including psychiatry, internal medicine and radiology. The study was approved by the McMaster Research Ethics Board and supported by a Canada Research Chair.

Materials

Participants were presented with 16 general medicine cases that were a randomly selected subset of cases used in Sherbino et al.,[26] Norman et al. [25] and Monteiro et al. [27] Thus these cases have been tested previously in several experimental studies. The level of case difficulty ranged from rare and difficult to straightforward acute medical conditions. The cases were designed to require equivalent reading times (i.e. they contained roughly the same number of words), however for some cases there may be more time required for the synthesis of data from history, physical and lab results. A sample case is shown in Appendix 1. Additionally, the current study included a 'patient photograph' paired with the primary complaint. Attempts were made to find representative photographs of people to match the patient description (i.e. age and sex). The patient photograph was meant to facilitate identification of specific cases in a subsequent review phase without referring to the complete case report so that, when asked to review a case without the case details present, they had an easily accessible reference to each case. All participants reviewed the same set of cases, but in randomized order. Cases were presented on laptop computers using RunTime Revolution (version 2.8.1; Edinburgh Scotland) software. Case processing time and case diagnoses were recorded by the software and exported as text.

Procedure

Initially participants were told they would be asked to diagnose general medicine cases. Once the experiment started, participants saw a welcome screen on the computer and they entered basic information: name, program name and program year. The primary instructions were delivered in written form on screen. These instructions included general tips for navigating through the program and entering responses as well as a description of how the case information would be presented. The program then randomized participants to 1 of 2 conditions and the relevant instructions for each condition were presented on screen. The instructions were identical except for the emphasis on the amount of time given. Participants given less time were asked to complete all diagnoses in 20 minutes and participants given more time were given 45 minutes. The group asked to proceed quickly is labelled the 'fast' group while the group asked to proceed slowly is labelled the 'slow' group. Unlike previous studies, [25,26] time limits were only suggested in this study and were not imposed by a deadline.

The instructions for the slow group were:

"You will be asked to read and diagnose several cases in 45 minutes. Each case description includes a brief description of the patient and vital statistics, as well as a photograph of the patient and an accompanying diagnostic image when available (e.g. x-ray, ECG, etc.)…Remember that you will not be able to go back to the case file once you have advanced to the diagnosis screen. Thoroughly read the case information as you have 45 minutes."

Once participants had completed diagnosis of all 16 cases they were presented with all the correct diagnoses as a form of feedback. The experiment ended at this point for this group.

The instructions for the fast group were:

“You will be asked to read and diagnose several cases in 20 minutes. Each case description includes a brief description of the patient and vital statistics, as well as a photograph of the patient and an accompanying diagnostic image when available (e.g. x-ray, ECG, etc.)…Remember that you will not be able to go back to the case file once you have advanced to the diagnosis screen. Read the case information completely, but remember to use your time carefully as you only have 20 minutes.”

After a first pass through the cases, participants in this group were then given an opportunity for further reflection and asked to review all 16 cases again and choose between retaining or revising their previous diagnosis. The message presented before proceeding with the review was:

“Thank you for assessing these cases quickly. We would now like you to carefully reconsider every diagnosis. Please re-consider all the evidence, before confirming or changing your initial diagnosis…”

Half of these participants reviewed the full case summaries, while half were only allowed limited access to information: the patient photograph, the primary complaint and the previous diagnosis. From here on, the first phase and initial diagnosis will be referred to as ‘pass 1’, while the second review phase will be referred to as ‘pass 2’.

Scoring

All diagnoses were scored without knowledge of experimental condition or whether it was entered as a first or revised diagnosis. All responses from pass 1 and pass 2 were scored for accuracy on a 3 point system. Incorrect diagnoses received a score of 0, partially correct responses received a score of 1 and correct diagnoses were assigned a 2. A list of correct, partially correct and incorrect diagnoses for all cases were developed by an expert panel of 2 experienced Emergency Medicine and 2 experienced Internal Medicine physicians.[26] This list was expanded using the scores for differential diagnoses and other possible responses collected from Sherbino et al. and Norman et al. Diagnoses for the current study were scored and tallied by the author (SM). Scores are reported as percent correct.

Analysis

All participants completed all 16 cases. Fewer than 0.5% of all responses were not recorded due to participant error (i.e. incorrectly advancing to the next screen).

To determine the effect of time limits on a first response diagnosis, individual case accuracy scores and response times for 16 cases were submitted to a repeated measures ANOVA with one within subject factor of case and one between subject factor of instruction (fast vs. slow). To determine the effect of having an opportunity to revise a quickly derived diagnosis, average case accuracy scores and response times for pass 1 and pass 2 for the fast group, were submitted to a repeated measures ANOVA with two between subject factors: post-graduate year (1, 2 or 3) and access to case details (full or limited). Finally, a negative correlation between time and accuracy was reported in previous research. [25,26] Therefore a correlation study was also conducted to look for a relationship between response times, accuracy from initial and revised diagnoses and post-graduate year.

Results

a) Effect of instructions and experience level

Consistent with previous studies (Monteiro, et al., submitted; Ilgen, et al., 2013; Norman, et al., 2013; Sherbino, et al., 2012), there was no effect of instructions on accuracy scores, as participants asked to use less time were just as accurate (60%) as participants asked to use more time (64%), $F = 1.1$, $p > 0.2$. As expected, there was a main effect of experience (i.e. post graduate year) as PGY 2 (63%) and 3 residents (65%) had higher scores than PGY 1 residents (56%), $F = 4.1$, $p < 0.05$.

Indicated by Figure 1, there was no differential benefit to diagnostic accuracy of slowing down, suggesting that more experienced residents did not benefit further from instructions to take more time. This was confirmed by a non-significant interaction ($F = 0.03$). Average case response times for each post-graduate year and the two conditions are presented in Figure 2.The analysis for response times validated the manipulation of instructions, as participants were faster on average in diagnosing each case (95 s) when allowed 20 minutes compared to participants allowed 45 minutes (107 s), although this difference was marginally significant, $F = 3.6$, $p = 0.06$. This difference in response time is consistent with previous studies (Norman, et al., 2013; Sherbino, et al., 2012; Monteiro, et al., submitted).

To determine if the effect of instructions interacted with experience level in response times, a separate univariate ANOVA was conducted on average response times with 2 between subject factors of instruction and experience level (PGY1, 2 and 3). Experience level had an impact on response time; PGY1 residents took longer on average (119s) to diagnose cases than PGY 2 (96s) and PGY 3 residents (89s), $F = 9.6$, $p < 0.0001$, but there was no interaction with instructions.

b) Effect of the opportunity to reflect

When residents from the fast group were offered the opportunity to review each case again, almost all diagnoses were reviewed a second time (about 75%), indicated by viewing times longer than 3 seconds. However, only 8% (62 out of 746) of all diagnoses were revised, suggesting that residents were generally confident in their initial diagnosis, despite the fact their accuracy only ranged from 58% to 64% on average. Furthermore, only 27 out of 47 participants chose to revise any diagnoses at all in pass 2, suggesting that there may be individual factors affecting the decision to revise a previous diagnosis.

There was a small yet marginally significant effect of reflection on accuracy as average pass 1 diagnoses of the fast cohort were slightly less accurate (60.5%) than average pass 2 diagnoses following reflection (61.5%), $F = 3.9$, $p = 0.06$. This small increase of 1% amounts to a minor improvement in average accuracy for 10 out of 47 participants, while the average accuracy for 32 participants did not change and for 5 participants, revision of a previous diagnosis led to a slight decrease in average accuracy. Table 1 indicates that the diagnoses that received a perfect score in pass 1 and were still revised after reflection experienced a drop in accuracy, while the diagnoses that were incorrect or partially correct did benefit slightly from revision. As indicated in Figure 3, these differences were not due to post graduate year, as there was no interaction between experience level and the opportunity to reflect, $p > 0.05$.

Table 1.

Changes in mean scores for all diagnoses following reflection and revision. There were a total of 746 diagnoses scored in pass 1.

Possible scores from pass 1	Number of diagnoses	Percent revised (number of diagnoses)	Average score in pass 2
0	158	18% (28)	0.09
1	279	10% (28)	1.01
2	309	1.9% (6)	1.99
Overall	746	8%(62)	1.22

As expected participants spent significantly less time reviewing a case in pass 2 (18s) than pass 1 (97s), $F = 424$, $p < 0.0001$. There was a strong influence of case availability on time spent reflecting in pass 2 as participants spent more time reviewing case details when they were available (21s) compared to participants with only the primary complaint and patient photograph (13s), $F = 6.9$, $p < 0.01$. However, the additional reflection time related to case availability did not improve accuracy (Mean without case details = 62%; with case details = 60% $F = 0.2$, $p > 0.6$, $df = 1$).

c) Correlation between response time, post-graduate year and accuracy

The correlation between pass 1 time and accuracy was negative but not significant (R = -0.2, p = .1). However, if we analyze the combined response time and accuracy scores for the fast group (pass 1) and the slow group, the correlation between time and accuracy is slightly stronger, R = -0.3, p = 0.07). These results are not consistent with the view that increased System 2 processing will yield more accurate diagnoses.

Discussion

In the current study, we examined the diagnostic accuracy of participants of different levels of experience given instructions to diagnose fast or slow. As expected, participants with the most experience performed better than less experienced participants overall. We also replicated the finding that instructions to proceed quickly or slowly, while altering the time of diagnosis, did not affect diagnostic accuracy.[25,26,27] Although not significant, there was a slightly negative relationship between response time and accuracy, which is not consistent with the view that reduced deliberate System 2 reasoning is responsible for higher error rates.

Additionally in this study participants in the fast group were then offered an opportunity to re-consider their initial diagnoses. Interestingly, although most participants took the opportunity to review each case, (75% of all cases were reviewed for at least 5 seconds during pass 2) only 8% of all diagnoses were revised. An investigation of the pattern of change in scores indicated that the majority of average scores remained the same, while a small portion either increased or decreased.

There are two possible explanations for the low rate of revision and lack of benefit from reflection in the present study. One possibility is that participants were being influenced by any number of cognitive biases.[9] For instance, rather than reviewing the case to test new hypotheses the majority may simply have confirmed their previous diagnosis, exhibiting a confirmation bias.[3,9] On the other hand, participants may have anchored to their original diagnosis, making it difficult to review the case through an unbiased lens.[3,9] Since these biases have been associated with System 1 processing, any errors should have been limited to diagnoses submitted during pass 1, when participants were encouraged to diagnose fast. According to the dual systems model discussed earlier, these are exactly the types of cognitive biases that should be overcome with reflective, deliberate System 2 reasoning.[3,8,9] Offering participants an opportunity for additional reasoning however did not improve accuracy, suggesting that added reflection alone is not a successful strategy.

Another possibility is that participants had low rates of revision or were unsuccessful at improving their scores because of limits in their knowledge or experience, which is consistent with the observation that participants with the most experience had the highest scores initially. Participants with the knowledge to diagnose a medical case correctly the first time did not need to reflect further, while participants without the required knowledge could not benefit from further reflection. This suggests that diagnostic performance is not modulated by reasoning skills or added reflection, but by experience and knowledge. It is unlikely that instructing residents to always reconsider their initial diagnosis will improve performance as we take the chance of reducing accuracy when the initial diagnosis was correct or forcing residents to make decisions beyond their knowledge and skill. It is unclear at this point whether participants in this study were aware of how accurate their initial diagnoses were as this lack of awareness could also contribute to low rates of revision.

It is likely simplistic to assume that one processing style is consistently better than another, or that any task instruction will guarantee exclusive use of one or the other system.[28] The processing required for diagnosis may be predominantly reflective or reflexive depending on context, past experience and the kind of knowledge structures required. A novice diagnostician for example, may need to rely on deliberate processing and more analytic knowledge structures to keep track of a full patient history, test results and a particular set of symptoms, however, it is reasonable that non-analytic knowledge or pattern recognition guides routine clinical procedures, the interpretation of basic symptomatology and the significance of the primary complaint. [29,30] In other words the novice must rely on System 1 processing for a multitude of tasks in order to have the resources to attend and reflect on the more difficult act of diagnosis. An expert diagnostician on the other hand has worked diligently to rely on non-analytic knowledge for an even greater number of tasks in order to free up resources to think of long-term plans for the patient, possible counter-indications, interactions with other medications and other needs of the patient. Therefore, what seems difficult for the novice is simpler for the expert, allowing experts to quickly recognize symptoms and identify diagnoses. Rather than focusing on the cognitive processes that expert physicians apply during diagnostic reasoning, research should focus on the learning experiences that contribute to expert medical diagnosis. Future research should focus on understanding the differences between physicians who benefit from reflection and physicians who are able to diagnose correctly the first time.

References

1 Evans, J. S. B., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition Advancing the Debate. *Perspectives on psychological science*, 8(3), 223-241.

2 Evans, J. S. B., & Elqayam, S. (2007). Dual-processing explains base-rate neglect, but which dual-process theory and how?. *Behavioral and brain sciences,* 30(03), 261-262.

3 Kahneman D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

4 Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology,* 94(4), 672.

5 Chisholm CD, Weaver CS, Whenmouth L & Giles B. (2011). A task analysis of emergency physician activities in academic and community settings. *Annals of emergency medicine*. 58(2): 117-122.

6 Croskerry P. (2009). Clinical cognition and diagnostic error: applications of a dual process theory of reasoning. *Advances in health sciences education*, 14(1): 27-35.

7 Croskerry, P. (2000). The Cognitive Imperative Thinking about How We Think. *Academic emergency medicine*, *7*(11), 1223-1231.

8 Croskerry P. (2003b). Cognitive forcing strategies in clinical decision making. *Annals of emergency medicine*, 41(1): 110-120.

9 Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. Academic Medicine. 78(8): 775 – 780.

10 Mamede, S., van Gog, T., van den Berge, K.,, Rikers, R.M.J.P., van Saase, J. I. C.M., van Guldener, C. and Schmidt, H.G. 2010. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Journal of the American medical association*. 304(11): 1198-1203.

11 Mamede, S., Schmidt, H. G., Rikers, R. M., Penaforte, J. C., & Coelho-Filho, J. M. (2008). Influence of perceived difficulty of cases on physicians' diagnostic reasoning. *Academic medicine*, 83(12), 1210-1216.

12 Mamede, S., Schmidt, H. and Penaforte, J.C. 2008b. Effects of reflective practice on the accuracy of medical diagnosis. *Medical education*. 42: 468:475.

13 Mamede, S., Schmidt, H.G., Rikers, R.M.J.P, Penaforte, J.C. and Coelho-Filho, J.M. 2007. Breaking Down Automaticity: Case Ambiguity and the Shift to Reflective Approaches in Clinical Reasoning. *Medical education*, 41: 1185 - 1192.

14 Mamede, S. and Schmidt, H. 2004. The Structure of Reflective Practice in Medicine. *Medical education*. 38: 1302 – 1308.

15 Klein, J. G. (2005). Five pitfalls in decisions about diagnosis and prescribing. BMJ: *British Medical Journal*, 330(7494), 781.

16 Redelmeier, D. A. (2005). The cognitive psychology of missed diagnoses. *Annals of internal medicine*, 142(2), 115-120.

17 Graber ML, Gordon R, Franklin N. (2005). Reducing diagnostic errors in medicine: what's the goal? *Academic medicine*. 77(10): 981-992.

18 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *AJM*. 2008; 121 (5 Suppl) S2.

19 Sherbino J, Kulasegaram K, Howey E, Norman G. (2013). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. *Academic emergency medicine*.

20 Sherbino, J., Dore, K.L., Siu, E., Norman, G.R. (2011). The effectiveness of cognitive forcing strategies to decrease diagnostic error: An exploratory study. *Teaching and learning in medicine*, 23 (1): 78-84.

21 Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American psychologist*, 49(8), 725.

22 Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *The teachers college record*, 104(4), 842-866.

23 Schon D. (1983). The reflective practitioner: how professionals think in action: New York, Basic Books.

24 Ilgen, J.S., Bowen, J.L., McIntyre, L.A., Banh, K.V., Barnes, D., Coates, W.C., Druck, J., Fix, M.L., Rimple, D., Yarris, L.M., & Eva, K.W. (2013). The impact of instruction to use first impressions or directed search on candidate diagnostic performance and the utility of vignette-based assessment. *Academic medicine*. 2013.

25 Norman, G. R., Sherbino, J., Dore, K., Wood, T.J., Young, M., Gaissmeier W, Monteiro S, Kreuger S. The etiology of diagnostic errors: A controlled trial of system 1 vs. system 2 reasoning. *Academic medicine*, (in press) 2013.

26 Sherbino, J., Dore, K.L., Wood, T.J., Young, M.E., Gaissmeier, W., Krueger, S., Norman, G.R. (2011). The relationship between response time and diagnostic accuracy. *Academic medicine*, 87: 785-91.

27 Monteiro SM, Sherbino JD, Ilgen JS, Dore KL, Wood TJ, Young ME, Bandiera G, Blouin D, Gaissmeier W, Norman GR. Disrupting diagnostic reasoning: The effect of interruptions on the diagnostic performance of residents and emergency physicians. JAMA (under review) 2013.

28 Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, 30(5), 513-541.

29 Ark T., Brooks LR & Eva K. (2006) Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Academic medicine*, 81, 405-409.

30 Ark, T.K., Brooks, L.R. and Eva, K.W. 2007.The Benefits of Flexibility: The Pedagogical Value of Instructions to Adopt Multifaceted Diagnostic Reasoning Strategies. *Medical education*. 41: 281 - 287.

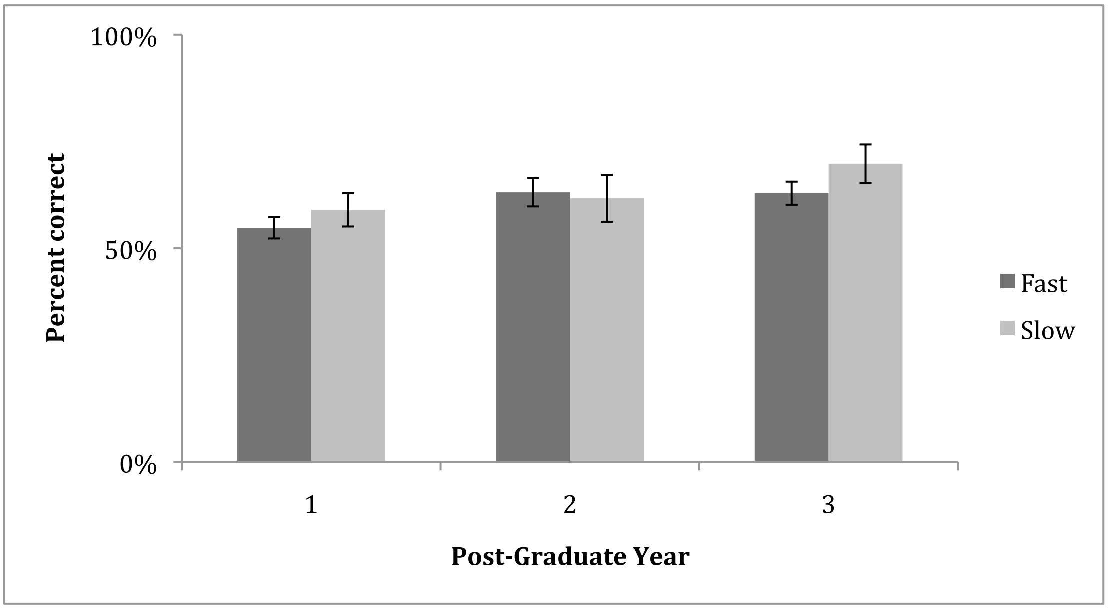


Figure 1: Average percent correct for participants in the fast group and the slow group.

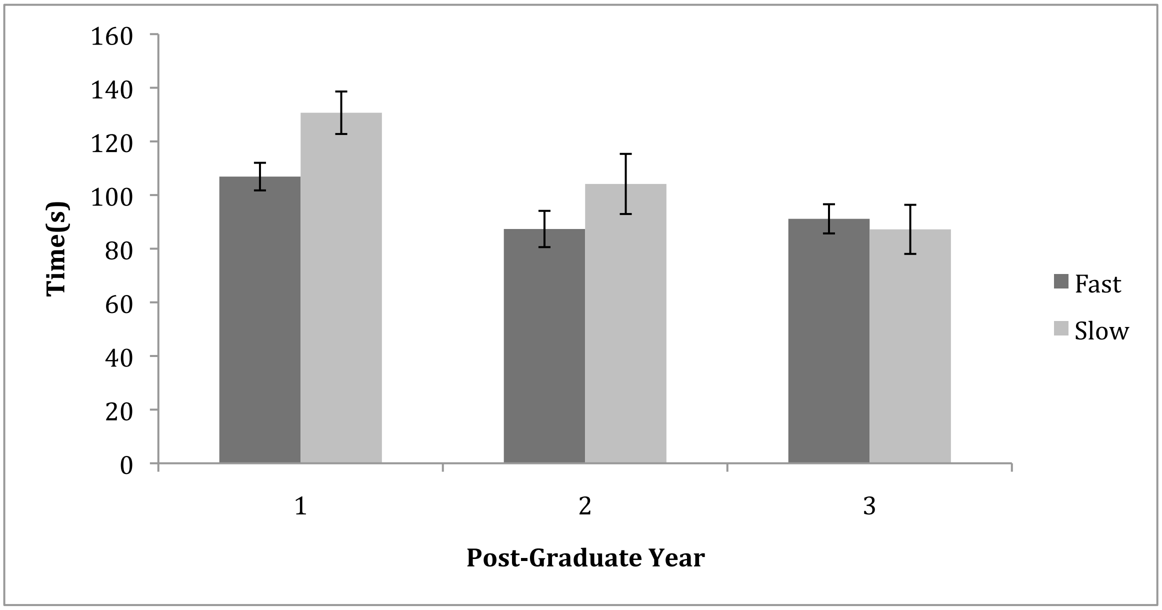


Figure 2: Average case processing times

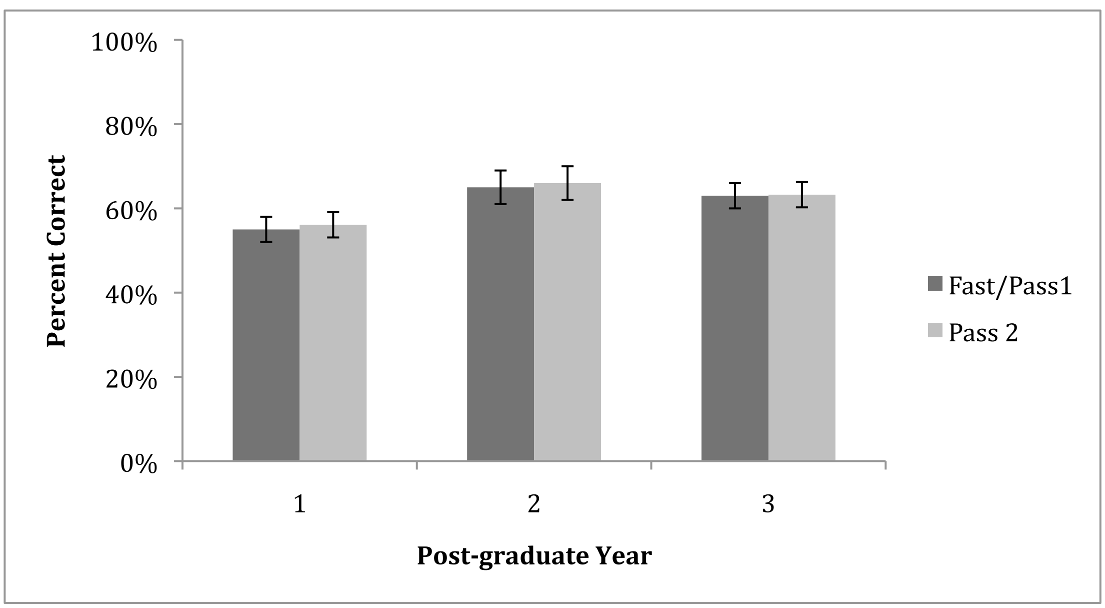
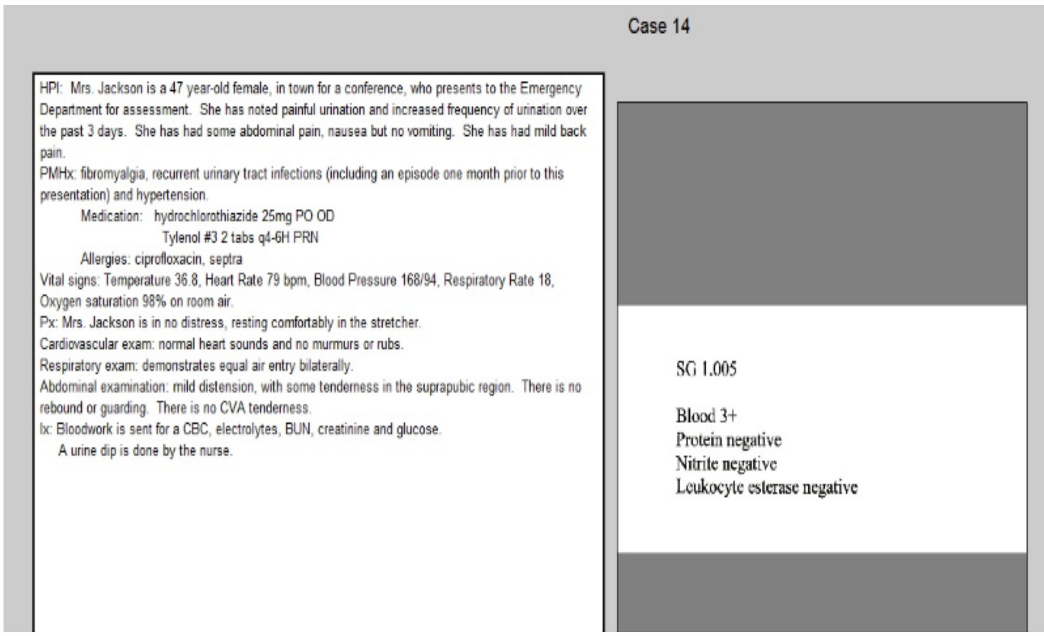


Figure 3: Average percent correct in Pass 1 and 2 for participants in the fast group.

Appendix 1.



Sample case used in diagnostic task. The written case details on the left of the screen and diagnostic image on the right side of the screen remained visible until the participant pressed a key to move on to the next screen. In this example, the diagnostic image are blood test results, although there were also cases with CT scans, rhythm strips and x-rays.

Preface

Chapter 5 presents a new interpretation of a subset of the data reported in chapter 4. This chapter contributes an understanding of when reflection (i.e. taking more time to reach a diagnosis) or revision (i.e. changing a previous diagnosis) is cued. Furthermore, this chapter takes a closer look at the relationship between experience and accuracy.

Chapter 5:

A retrospective analysis of data from a diagnostic task: Do all physicians benefit from reflection equally?

Sandra Monteiro, Geoffrey Norman and Jonathan Sherbino

Corresponding Author

S.D. Monteiro, M.Sc.

McMaster University, Department of Psychology

1280 Main Street West,

Hamilton, Ontario, L8S 4L1, Canada

e-mail: monteisd@mcmaster.ca

phone : 905 525 9140 ext. 23114

fax : 905 572 7099

Abstract

Context

Some authors suggest that reflecting on a diagnosis can prevent diagnostic errors. Although a prescribed form of structured reflection has been shown to be beneficial in some situations, the factors that influence the decision to reflect or revise a diagnosis are not well known.

Objective

The goal of the current study is to examine how experience and a self-directed decision to reflect affect the accuracy of revised diagnoses.

Methods

The current study is a re-analysis of data from Monteiro et al. (in prep). Medical residents (N = 47) from McMaster University diagnosed 16 medical cases (pass 1). Participants were then given the opportunity to reflect on each case (pass 2). Half the participants were permitted to review the full case description and half saw only the primary complaint.

Primary Measures

Individual Pass 1 and 2 scores and response times. Scores and times were coded according to whether they were confirmed or revised in pass 2.

Analysis

Individual scores and response times were submitted to 3 analyses. Individual accuracy scores and response times for revised diagnoses were submitted to separate repeated measures ANOVAs with one within subject factor of pass (pass 1 or 2) and two between subject factors of post-graduate year (PGY 1, 2 or 3) and access to case details (limited or full). Individual accuracy scores and response times were submitted to separate univariate ANOVAs with two between subject factors of post-graduate year (PGY 1, 2 or 3) and decision to revise (revised or confirmed). Individual pass 2 scores and response times were submitted to a similar analysis.

Results

There was a trend for lower accuracy scores and higher rates of revision in pass 2 to be related to longer response times in pass 1 (107sec) than correct diagnosis (96sec). Initial diagnoses that were not revised had the highest accuracy (60%), compared to pass 1 diagnoses that were later revised (32%) and pass 2 revised diagnoses (42%), $F = 9$, $p < 0.001$.

Conclusions

Reflection was used as a strategy when the correct answer was not known, which resulted in some improvement in accuracy. However, these revised diagnoses were not as accurate as initial unrevised diagnoses.

Introduction

The literature on diagnostic errors in medicine presents two very different perspectives about the source of errors and strategies for improving performance. From one perspective, a large proportion of diagnostic errors are preventable, resulting from cognitive biases during the reasoning process.[1,2] Therefore several authors who support this perspective have argued that cognitive errors in medicine can be reduced through strategies to improve reasoning and eliminate biases.[1,2,3,4] One such strategy is for physicians to always reconsider all diagnoses and decisions as if they were incorrect and pursue alternate hypotheses.[4,5,6] This strategy would be independent of any actual intuition physicians might have about their own accuracy. However, this strategy hinges on two assumptions, which are not altogether practical and are largely not supported by research.

First, there is the assumption that rapid pattern recognition is a less sophisticated form of diagnosis than taking more time to reflect and actively pursue alternate explanations. [6,7,8] According to this assumption, a rapidly determined hypothesis should be less accurate than one determined more slowly or following intense thought. This assumption has a common sense appeal and has influenced recent efforts for curriculum change.[3,6] When this assumption was tested experimentally however, the results did not fully support the notion that added reflective thought improves performance. When junior medical residents were instructed to diagnose some cases rapidly and use a very intense reflective method for the remaining cases in a diagnostic task, an overall benefit of reflection was not observed as only a small subset of difficult cases showed a benefit of reflection.[5] However, when a similar structured reflective method was tested in a between subjects design with medical students, junior/senior residents and faculty, there was no overall or selective benefit of reflection, as differences in accuracy scores were influenced more by the participant's experience level.[9] As well, when junior residents were instructed to either diagnose rapidly [10] or more slowly,[11] faster response times were correlated to more accurate diagnoses than slower response times, regardless of case difficulty or instructions. Therefore, there is currently evidence that simply slowing down to reflect will not improve performance.[10,11]

A second assumption, is that all physicians, residents and faculty alike, already possess the required knowledge to ascertain the correct diagnosis for any medical case and therefore errors are caused by misjudging the level of cognitive processing required. For instance, when arguing that added deliberation (i.e. additional reasoning) will reduce diagnostic errors, several authors have linked diagnostic errors to faulty cognitive processes, suggesting that errors are not caused by insufficient knowledge (or inexperience) but by insufficient reasoning (i.e. thinking processes).[1,2] To this end, Custers (2013) has argued that physicians can be trained to classify medical cases by their complexity in order to apply the appropriate level of cognitive processing.[12] This proposal suggests that physicians do not naturally know when a case is difficult, presuming that although they have the knowledge to solve any case, they need special instructions on how best to access their knowledge. Therefore, approaching difficult cases with extra vigilance should improve accuracy. There is some support for this strategy. In one study, junior residents had a higher level of accuracy when given specific instructions to be cautious for medical cases described as problematic, compared to equally difficult cases presented without such warnings.[5] In another study, junior residents were instructed to reflect further on a select group of previously diagnosed misleading cases, and as a result, accuracy for those misleading cases did improve.[13] Possibly the instructions emphasized the probability for error, encouraging residents to consider alternatives and revise their previous diagnosis. Together, the results of these two studies can be interpreted to suggest that heightened awareness about case difficulty or errors can improve performance. However, in those studies, errors occurred in cases that were designed to be misleading, and improvements in accuracy only resulted from very directed and specific instructions for reflection and revision. Whether a generalized strategy to be more cautious can be sustained, through an entire shift for example, remains to be tested. It is possible that such an approach encourages physicians to only consider unusual or rare diagnoses, which may be inappropriate for more common or simpler medical cases. Certainly, a benefit of reflection has yet to be reported for simpler case,[13,14] and in one study, additional reflection and revision of previously correct diagnoses, led to a small decrease in accuracy.[15] Therefore, it may not be prudent for all physicians to treat every case as if it was problematic and it is unreasonable to promote a strategy that has only been demonstrated to work when residents were instructed to use it for specific cases.

At this point, it should be clear that the primary solutions proposed to reducing diagnostic errors due to preconceived biases in reasoning have focused on improving reasoning rather than improving practical or applied knowledge. However, there is currently no evidence that training in generalized reasoning strategies will improve performance. Indeed, such a notion has been discouraged since Elstein (1978).[16]

The alternative to the perspective that diagnostic errors are related to faulty reasoning is the perspective that errors are related to a lack of knowledge or experience.[17,18] While earlier retrospective reports of diagnostic errors have classified observed errors as resulting from bias, [1,2,3] a report by Zwaan et al., classified errors more appropriately as *mistakes* due to a lack of knowledge, inexperience or simply from not realizing an error had been made.

Errors due to a lack of knowledge will likely not be corrected by further reflection. However, even with a lack of knowledge, there may be a realization that an error is likely and reflection may be a naturalistic strategy adopted when there is insufficient knowledge or experience to determine the answer immediately. Certainly longer response times have been linked to less accurate diagnoses, [10,11] suggesting that physicians know their initial diagnosis is incorrect and are therefore trying to improve their accuracy; albeit with little success. Such a finding contradicts the simplistic assumption that all physicians have equal or sufficient knowledge to handle any medical case. Therefore, it should be possible to distinguish between physicians who have the knowledge to determine the correct answer right away and physicians who need to reflect further. As well, it should be possible to distinguish between physicians who can assess their own accuracy and physicians who cannot.

In general, self-assessment skills are considered poor [19] and the factors that contribute to self-assessments of competence are complex.[20] However, we focus quite narrowly on the ability to assess one's diagnostic accuracy. To date, there is little research investigating this ability in medical education. In one study, Friedman et al. compared the diagnostic accuracy and confidence levels of medical students, residents and faculty, concluding that self-assessments of knowledge, or confidence in one's diagnosis, were misaligned with actual accuracy.[21] They reported measurable amounts of over confidence (i.e. being confident in an incorrect diagnosis) and under confidence (i.e. not being confident in a correct diagnosis). However, all participants in that study were confident more often for correct than for incorrect diagnoses. Additionally, Woolley and Kostopoulou assessed the validity of physicians' 'gut feelings' and 'intuition' in patient care, concluding that physicians often respond to unconscious signals that something is wrong, even when established guidelines and standard practice suggest the opposite.[22] In that study, the majority of physicians expressed a lack of confidence in their intuitions primarily because they could not immediately justify their concern. Essentially, they felt that if they could not identify the source of the concern, they would not be taken seriously and should therefore discontinue the search for an alternate answer.

Outside of medical education, undergraduate psychology students were far more accurate (65.9%) on general knowledge questions they answered immediately than for deferred questions they could not answer right away (4.3%). This result is consistent with the suggestion that people make quick judgements about their knowledge and only reflect when they are uncertain or do not have the knowledge.[23] Such a result has not been demonstrated in medical education as many studies that have investigated the role of reflection have studied prescribed reflection as opposed to naturalistic reflection. In the study by Mamede et al. (2010) physicians were assigned cases to revise. There was no indication that the participants in that study had any realization that a mistake had been made prior to the second review stage. Therefore, a reasonable empirical question is whether it is possible to measure self-assessed accuracy in a diagnostic task and whether physicians are more likely to revise or maintain a previous diagnosis when given an opportunity to self-select cases for reflection. As well, in that study, participants had access to the full case during the review process, which may not be comparable to a realistic setting. Physicians are sometimes called upon to reconsider a patient case based on a patient's name, hospital room number or primary complaint, and it is a reasonable empirical question as to whether having access to the full case details provides an advantage over having such limited information.

The Present Study

In the present study we were interested in whether we could measure diagnostic uncertainty, evidenced by a physician's assessment that they made a mistake. In previous studies slower response times were related to lower accuracy scores, suggesting that physicians, who did not know the answer immediately, took longer to diagnose a case.[10,15] In the current study we hypothesized that longer response times would indicate uncertainty and might also predict a decision to revise an incorrect diagnosis. We were also interested in measuring the benefit of a decision to revise and whether more experienced residents were better able to identify and correct errors compared to more junior residents.

Research Questions

1. Can physicians identify which cases are problematic for them? And can they identify their own errors?
2. Does response time predict a decision to revise a diagnosis?
3. Does experience level predict diagnostic accuracy for initial as well as revised diagnoses?

Method

Design

This study is a mixed design comparing between subject accuracy and response times for individual trials based on the decision to revise or confirm a diagnosis. We also compared within subject accuracy comparing initial diagnoses to revised diagnoses.

Setting

Medical residents from the teaching hospitals associated with McMaster University in Hamilton were invited to participate. The test sites were the Juravinski Cancer Centre, St. Joe's Hospital and McMaster Children's Hospital. The study was conducted by SM using laptops set up in conference rooms within each of the test sites.

Participants

Recruitment

The results discussed in this study are part of a larger study. A total of 65 residents in post-graduate years 1, 2 and 3 were recruited from Hamilton health care centers to participate in a study and complete a diagnostic task. Invitations to participate in the study were sent out by e-mail to a general list of all residents in post-graduate year 1, 2 and 3. Residents were not paid for their time, however, as the experiment was most often conducted during early morning or noon hours, most residents were compensated for their time with food.

Through random assignment, 47 residents were assigned to receive instructions to complete a diagnostic task in 20 minutes. The present study focuses on the performance of only these 47 residents: 19 residents in post-graduate year (PGY) 1, 11 in PGY 2 and 17 in PGY 3. Participants were from different specialties including psychiatry, internal medicine and radiology. The study was approved by the McMaster Research Ethics Board and supported by a Canada Research Chair Grant.

Materials

Participants were presented with 16 general medicine cases; a subset of cases used in previous studies.[10,11,15,24] The basic study design was similar to previous studies with cases ranging from rare to straightforward acute medical conditions.[10,11,15,24] Each case contained roughly the same number of words, however cases describing rare conditions may have required more time for synthesizing information from patient history, physical and lab results. A sample case is shown in Appendix 1.

In addition to the basic design, the primary complaint was paired with a 'patient photograph' (i.e. stock images) selected to match each patient description (i.e. age and sex). Cases were presented in random sequence on laptop computers using RunTime Revolution (version 2.8.1; Edinburgh Scotland) software. Case processing time and case diagnoses were recorded by the software and exported as text.

Procedure

All participants were given simple verbal instructions indicating that they would be required to diagnose several general medicine cases. Each participant then entered basic information: name, program name and program year on a welcome screen on a laptop. More detailed instructions were delivered in written form on screen. All participants included in the current study received instructions to proceed quickly and complete diagnosis of all cases in 20 minutes (pass 1). The requirement to proceed quickly was not controlled as in previous studies which used a timer and deadline.[10,11]

The instructions presented during pass 1 were:

"You will be asked to read and diagnose several cases in 20 minutes. Each case description includes a brief description of the patient and vital statistics, as well as a photograph of the patient and an accompanying diagnostic image when available (e.g. x-ray, ECG, etc.)…Remember that you will not be able to go back to the case file once you have advanced to the diagnosis screen. Read the case information completely, but remember to use your time carefully as you only have 20 minutes."

Upon completion of this task, participants were then required to review all cases again. Cases were presented again in random sequence and participants selected how much time to dedicate for the review process and whether or not they revised or confirmed a previous diagnosis. In this stage of the experiment (pass 2), all participants saw their pass 1 diagnosis for each case, however, 21 participants were randomly assigned to have full access to all the case details, while 26 participants saw a restricted amount of information consisting of the patient photograph and primary complaint only.

The instructions presented for pass 2 were:

"Thank you for assessing these cases quickly. We would now like you to carefully reconsider every diagnosis. Please re-consider all the evidence, before confirming or changing your initial diagnosis…"

Scoring

Diagnoses were scored according to a previously established rubric[10] by SM using a 3-point scoring system: 0 for incorrect, 1 for partially correct and 2 for correct. All diagnoses were scored blind to condition and pass. Individual response times and accuracy scores were then coded according to the decision to confirm or revise a diagnosis in pass 2. Therefore, both pass 1 and 2 scores were labelled as confirmed or revised in pass 2.

Analysis

We hypothesized that the decision to revise a diagnosis in pass 2 would indicate an accurate assessment of an error, so pass 1 scores that were later revised should be less accurate than pass 1 scores that were not revised. This uncertainty might also be reflected in response times during pass 1 as greater uncertainty might lead to longer diagnosis times. Additionally, experience level might affect the rate of revisions as well as accuracy in pass 1 and 2. However, only 27 residents chose to revise at least one diagnosis during pass 2, limiting the analysis. Therefore, instead of treating residents as subjects, we treated individual case scores as subjects in order to compare performance for case scores and response times related to revised and unrevised diagnoses.

Individual pass 1 case scores and response times were submitted to separate univariate ANOVAs with 2 between subject factors of post-graduate year (1, 2 or 3) and decision (revised or confirmed). Pass 2 scores were submitted to the same analysis to determine if the opportunity to reflect further on an initial diagnosis provided an overall benefit. Finally, we were interested in determining if revised pass 2 diagnoses had a higher accuracy than the previous pass 1 diagnoses and if the change in score was related to post graduate year and case access. Individual case accuracy scores for revised diagnoses only were submitted to a repeated measures ANOVA with one within-subjects factor of pass (pass 1 or 2) and two between subject factors of case access (full or limited) and post-graduate year (1, 2 or 3).

Results

There were a total of 746 diagnoses successfully entered in pass 1, translating into a 99% completion rate. There were 302 diagnoses entered by

PGY 1 residents, 172 entered by PGY 2 residents and 272 entered by PGY 3 residents. . Table 1 indicates how revisions to pass 1 diagnoses affected average scores.

Overall, only 8% of all pass 1 diagnoses, (62 diagnoses) were revised. Incorrect and partially correct diagnoses were revised more often compared to correct diagnoses suggesting that to some degree, participants were aware of their mistakes, χ^2 2 = 15.6, $p < 0.0001$. However, only 28 of 158 incorrect diagnoses (18%) were revised, and the average score of the revised diagnoses was only .50 out of 2. Similarly, only 28 of 279 (10%) of partially correct diagnoses were revised, and this resulted in only a change of 0.14 in accuracy. and the few revisions that were applied to correct diagnoses, resulted in lower average accuracy. Therefore, although residents were, to some degree, able to identify their own mistakes and made attempts to correct them, the impact of revisions was minimal, especially compared to the overall accuracy of unrevised diagnoses

Table 1. percentage of scores that were revised and the change in score

Pass1 Score	Number of diagnoses with this score that were revised	Total number of diagnoses that received that score	Pass2 Scores (Revised diagnoses only)
0	28	158	0.50
1	28	279	1.14
2	6	309	1.33
Overall	62	746	0.87

We were also interested in whether access to case details affected the rate of revisions. In a previous study [15] access to case details did not affect accuracy in pass 2 (Mean accuracy without case details = 1.25; with case details = 1.21 F = 0.3, p > 0.6, df = 1), however the rate of revisions were different. In this study, when case details were available, participants revised more case diagnoses (39) compared to participants who did not have the full details (23) during pass 2, χ^2 2 = 4.1, p = 0.04. Additionally, with access to case details residents spent more time reviewing the case (42sec) during pass 2 than without case details available (26sec), F = 28.4, p < 0.0001, df = 1, suggesting that they were trying to look for missed information.

Post-graduate year was also related to the number of revisions as PGY1 residents revised more diagnoses (32) compared to PGY2 (17) and PGY3 (13) residents. This difference may not be surprising given the difference in numbers of residents in each post-graduate year, however, it was significant, χ^2 2 = 9.7, p = 0.008. In a previous study [15] senior residents (PGY 2 and 3) were more accurate (1.3) than junior residents (1.1). We were curious as to whether similar differences could be measured between residents at different levels dependent on the decision to revise or not. The pattern indicated that more senior residents had higher accuracy following revision (1.1) compared to PGY 2 (0.88) or PGY 1 (0.78), suggesting that residents with more knowledge were better equipped to correct errors.

The average response times for pass 1 and 2 scores that were revised or confirmed are presented in figure 1. The average percent correct for pass 1 and 2 scores that were revised or confirmed are presented in figure 2.

Pass 1 diagnoses that were later revised had longer average response times (107sec) compared to pass 1 diagnoses that were later confirmed in pass 2 (96sec). Although this difference was not significant, $p = 0.1$, average response times during pass 1 were negatively correlated with accuracy scores in pass 1, $R = -0.23$, $p < 0.001$ and accuracy scores in pass 2, $r = -0.3$, $p < 0.05$.

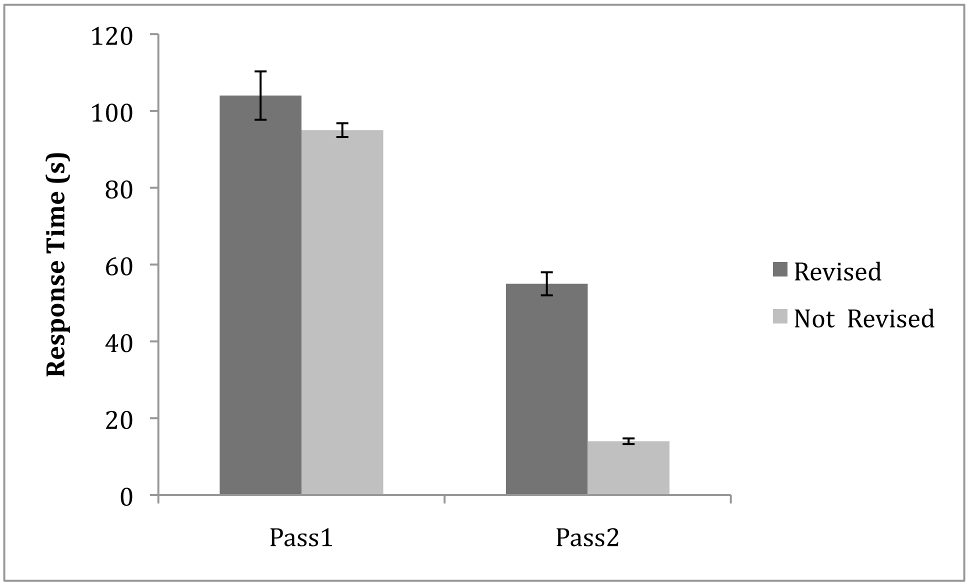


Figure 1. Average response times for diagnoses Pass 1 and 2 diagnoses that were revised or not revised in Pass 2.

The longer response times may indicate uncertainty or lack of knowledge as there was a consistent overall pattern of higher accuracy for confirmed pass 1 diagnoses (mean = 1.3) compared to pass 1 diagnoses that were later revised (mean = 0.65), $F = 29$, $p < 0.001$, df = 1. Revised diagnoses showed a slight, but significant increase in accuracy during pass 2 (mean = 0.87), ($F = 4.3$, $p < 0.05$, df = 1). However, pass 1 confirmed diagnoses (mean = 1.3) were still far more accurate than revised pass 2 diagnoses (mean = 0.87), ($F = 11.04$, $p < 0.01$), df = 1.

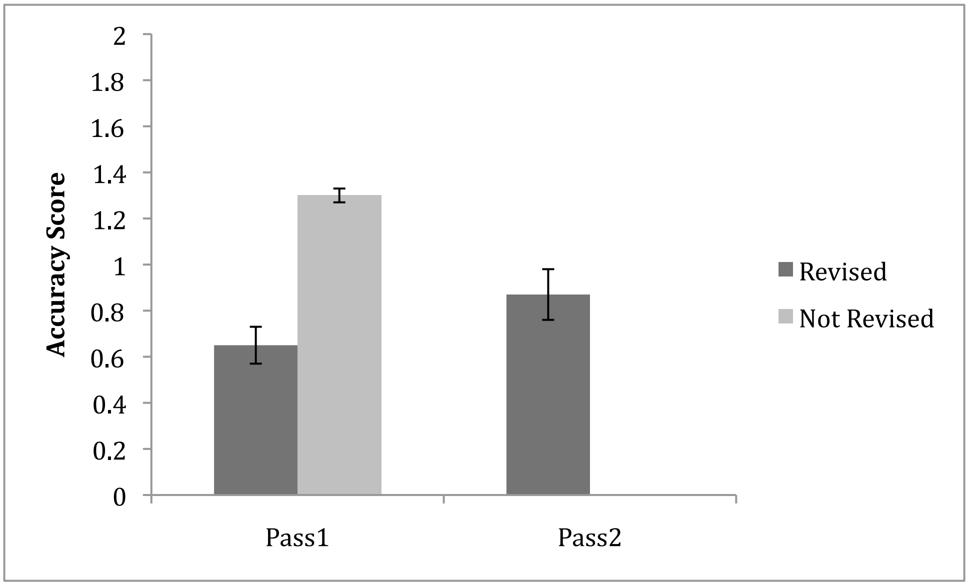


Figure 2. Average case accuracy for pass 1 scores (revised or not revised in pass 2) and pass 2 scores that were revised.

Discussion

The results of the present study demonstrate that if physicians are aware of their diagnostic mistakes they will attempt to correct them by trying to revise incorrect diagnoses. However their ability to detect and correct errors is far from perfect; only 18% of incorrect diagnoses were revised, and the average change was only 0.50. The overall accuracy of revised diagnoses remained much lower than pass 1 diagnoses that were maintained during pass 2.

The assumption that extended reflection during diagnosis or taking a second look at a medical case will necessarily improve accuracy is not supported. These findings discredit the proposed strategy that training physicians to identify errors will improve performance.[3,5,13,14] The current results demonstrate that most errors remain undetected, and simply discovering an error is not sufficient to ensure correction.

The current results are also inconsistent with the results of a previous study that employed a similar experimental design (i.e. compared the accuracy of initial diagnoses to that of revised diagnoses).[13] In that study, however, some medical cases were manipulated to increase the error rate during the first pass. As well, during the second pass, residents were only asked to re-visit the diagnoses of those cases experimentally manipulated to increase error. The design used by Mamede et al., may have influenced how participants in that study reflected on previous diagnoses and, by drawing attention to a subset of cases, may have encouraged them to consider revisions more carefully. This observation is consistent with an earlier study by this group in which residents performed better on cases that were described as very difficult, compared to cases without a similar warning. But in both studies, the identification of cases for further reflection was, to some degree, guided by the experimenter. A second difference between the two studies is that, in the Mamede (2010) study, the errors that were made did not result primarily from knowledge gaps, as the cases themselves were engineered to cue the incorrect diagnosis.

By contrast, in the present study, residents relied on their own judgement, knowledge and experience to decide on revision. Although some cases were more challenging, none were manipulated to be misleading. As such, residents with more knowledge and experience performed better during both passes. It is likely that the physicians who did not benefit from the opportunity to reflect and revise were lacking the requisite knowledge or experience to correct errors.

If this interpretation is correct, error reduction strategies should focus more on improving the knowledge and experience base contributing to the identification of correct diagnoses, rather than the development of metacognitive strategies to detect errors. The evidence from this and other studies, suggests that error detection skills develop in concert with experience, rather than as prescribed techniques. Certainly, there is a difference between residents in this study who did not choose to revise any diagnoses and those who did. When physicians knew the answer and presumably believed that they knew the answer, reflection was unnecessary.

Conclusion

There is currently evidence demonstrating the inadequacy of cognitive forcing strategies.[25,26] Reflection has been shown to improve accuracy but only when highly structured, or used as a prescribed heuristic with junior residents.[13] Reflective practice did not provide any advantage when tested with a larger more diverse group of medical students, residents and faculty.[9] As well, in the present study, an unstructured form of reflective practice that relied on self assessments of accuracy, did not reduce errors significantly.[15] As strategies that focus on additional or improved reasoning are not reliable, it may be time to consider strategies that focus on improved knowledge and experience.

In several retrospective reports, the rate of diagnostic errors linked to knowledge deficits is quite low compared to the rate of errors linked to biased reasoning.[1,2] However, if we define knowledge by its application, then there can be a great deal of variation in how much applied knowledge each physician has acquired. For example, all physicians may have equal knowledge of the primary symptoms of Addison's disease, however some may have diagnosed and treated more patients with the disease, each presenting with variations on the primary symptoms. There is some evidence of the independent role of experiential knowledge in clinical reasoning. In a previous study [24] while years in practice did not predict accuracy for a sample of experienced emergency physicians, current workload, measured in shifts per week and patients per shift, had a 0.29 correlation with accuracy. It is possible then, for physicians to have all the factual medical knowledge necessary, but only limited practical knowledge. In the future, medical education programs should include improved methods for practice in order to increase this practical knowledge.

References

1 Chisholm C.D., Weaver C.S., Whenmouth L. & Giles B. (2011). A task analysis of emergency physician activities in academic and community settings. *Annals of emergency medicine*. 58(2): 117-122.

2 Graber M.L., Gordon R, Franklin N. (2002). Reducing diagnostic errors in medicine: what's the goal? *Academic medicine*. 77(10): 981-992.

3 Croskerry P. (2009). Clinical cognition and diagnostic error: applications of a dual process theory of reasoning. *Advances in health sciences education*, 14(1): 27-35.

4 Redelmeier, D. A. (2005). The cognitive psychology of missed diagnoses. *Annals of internal medicine*, 142(2), 115-120.

5 Mamede, S., Schmidt, H.G., Rikers, R.M.J.P., Penaforte, J.C., Coelho-Filho, J.M. 2008a. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. Influence of Perceived Difficulty of Cases on Physicians' Diagnostic Reasoning. *Academic medicine*. 83(12): 1210-1216.

6 Croskerry P. (2003). Cognitive forcing strategies in clinical decision making. *Annals of emergency medicine*; 41(1): 110-120.

7 Schon D. (1983). The reflective practitioner: how professionals think in action. New York, Basic Books.

8 Kahneman D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

9 Ilgen, J.S., Bowen, J.L., McIntyre, L.A., Banh, K.V., Barnes, D., Coates, W.C., Druck, J., Fix, M.L., Rimple, D., Yarris, L.M., & Eva, K.W. (2013). The impact of instruction to use first impressions or directed search on candidate diagnostic performance and the utility of vignette-based assessment. *Academic medicine*. 2013.

10 Sherbino, J., Dore, K.L., Wood, T.J., Young, M.E., Gaissmeier, W., Krueger, S., Norman, G.R. (2012). The relationship between response time and diagnostic accuracy. *Academic medicine*. 87: 785-91.

11 Norman, G.R., Sherbino, J., Dore, K., Young, M., Gaissmeier, W., Monteiro, S., Kreuger. The etiology of diagnostic errors: A controlled trial of system 1 vs. system 2 reasoning. Submitted to *Academic medicine*, 2012.

12 Custers, E. J., Regehr, G., & Norman, G. R. (1996). Mental representations of medical diagnostic knowledge: a review. *Academic medicine*, 71(10), S55-61.

13 Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M., van Saase, J. L., van Guldener, C., & Schmidt, H. G. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. JAMA: *The Journal of the American medical association*, 304(11), 1198-1203.

14 Mamede, S., Schmidt, H. and Penaforte, J.C. 2008b. Effects of reflective practice on the accuracy of medical diagnosis. *Medical education*. 42: 468:475.

15 Monteiro S., Norman G, Sherbino J, Patel A, Mazzetti I, Gardhouse A and Howey E. Reflecting on diagnostic errors: Does it help to take a second look? In prep.

16 Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical Problem Solving A Ten-Year Retrospective. *Evaluation & the health professions*, 13(1), 5-36.

17 Monteiro, S.D. & Norman, G. In Press. Diagnostic reasoning and remembering: How physicians think and learn. *Teaching and learning in medicine*.

18 Zwaan, L., Thijs, A., Wagner, C., van der Wal, G., & Timmermans, D. R. (2012). Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Academic medicine*, 87(2), 149-156.

19 Eva, K. W., Cunnington, J. P., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in health sciences education*, 9(3), 211-224.

20 Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic medicine*, 80(10), S46-S54.

23 Eva, K. W., & Regehr, G. (2007). Knowing when to look it up: A new conception of self-assessment ability. *Academic medicine*, 82(10), S81-S84.

21 Friedman, C. P., Gatti, G. G., Franz, T. M., Murphy, G. C., Wolf, F. M., Heckerling, P. S., ... & Elstein, A. S. (2005). Do physicians know when their diagnoses are correct?. *Journal of general internal medicine*, 20(4), 334-339.

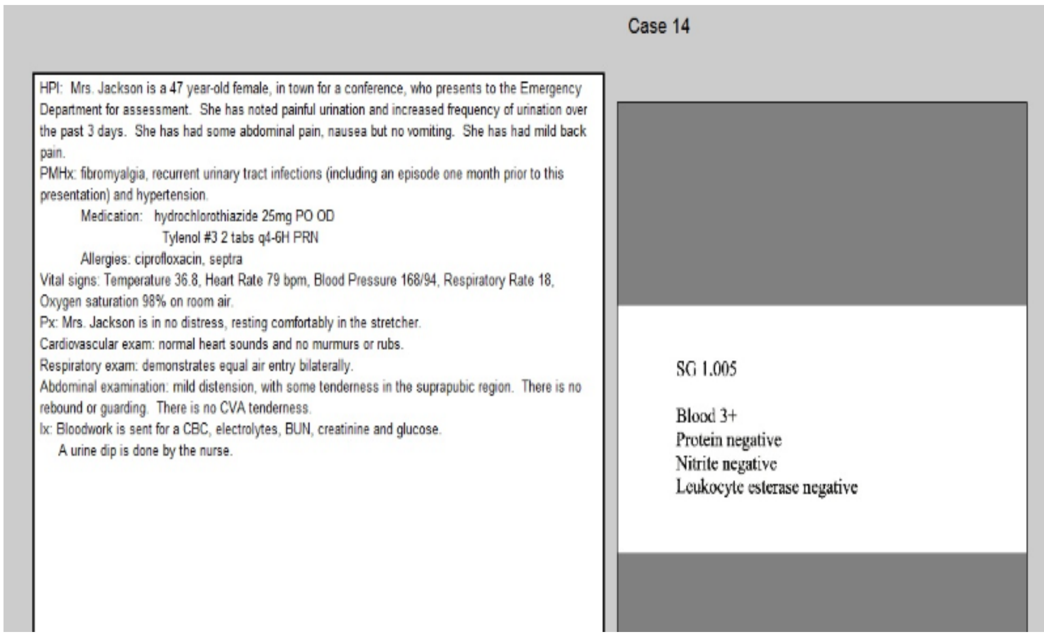
22 Woolley, A., & Kostopoulou, O. (2013). Clinical intuition in family medicine: more than first impressions. *The Annals of family medicine*, 11(1), 60-66.

24 Monteiro, S.D, Sherbino, J.D., Ilgen J.S., Dore, K.L., Wood, T.J., Young, M.E., Bandiera, G., Blouin, D., Gaissmeier, W., Norman, G. and Howey, E. In Prep. Disrupting diagnostic reasoning: The effect of interruptions on the diagnostic performance of residents and emergency physicians.

25 Sherbino, J., Dore, K.L., Siu, E., Norman, G.R. (2011). The effectiveness of cognitive forcing strategies to decrease diagnostic error: An exploratory study. *Teaching and learning in medicine*. 2011; 23 (1): 78-84.

26 Sherbino J, Kulasegaram K, Howey E, Norman G. (2013). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. *Academic emergency medicine*.

Appendix 1.



Sample case used in diagnostic task. The written case details on the left of the screen and diagnostic image on the right side of the screen remained visible until the participant pressed a key to move on to the next screen. In this example, the diagnostic image are blood test results, although there were also cases with CT scans, rhythm strips and x-rays.

Chapter 6: Final Discussion

Preface

In the general introduction to this thesis I discussed two forms of dual process models. A parallel competitive model structure supports research in many domains including cognitive control, categorization and recognition memory, while a default interventionist model structure is associated primarily with research on reasoning. [24] Evans and Stanovich [24] distinguish between these two forms of models in order to clear up misconceptions about dual process reasoning and strategies for improving performance on reasoning tasks.

A parallel competitive model structure,[24] can be applied to classic research in cognition, which focused on the mechanisms that underlie selection of one cognitive process over another. For example, early models of cognitive control proposed a homunculus or selective control mechanism that monitored goals and selected between an automatic and controlled style of processing.[54,55] Such a mechanism is similar to a mechanism that initiates an interventionist System 2 that is capable of overriding automatic associative memory processes.[24] However, current research suggests that the selection mechanism for cognitive control is far more dynamic and flexible.[56] Given the evidence of a more dynamic control selection mechanism, a stage model of reasoning that relies on an inflexible and unidirectional progression from System 1 to System 2 seems less likely. Additionally, a parallel competitive model structure can be applied to classic research in memory, which proposed two processes: familiarity and recollection.[57] The process of recollection is characterized as being more analytic and slower compared to the process of familiarity.[28] However, according to Evans & Stanovich's[24] distinction between dual process model structures, both familiarity and recollection processes would be considered part of the default memory based System 1, which again seems unlikely. This thesis was motivated by the assessment that the default interventionist model of reasoning, which has been highly influential in medical education, is incompatible with current evidence from research in cognitive control and memory processes.

Major Contribution of Data Chapters

In the present thesis, I first examined differences in the progression of medical education research and strategies for diagnostic error reduction based on models of reasoning compared to models of memory. Models of memory suggest that diagnostic performance is related to the manner in which prior experience or knowledge is recruited during a diagnostic reasoning task. In particular, several studies suggest that the similarity of prior experiences to a current problem influences diagnostic accuracy.[11,12,13,14,15,16,17] Alternatively, models of reasoning suggest that diagnostic performance is related to the manner in which logical reasoning is applied to a current problem, regardless of prior experience. This has been interpreted to suggest that performance can be improved through training in reasoning skills.[18,19,20,21,35] Because this particular strategy to improve reasoning skills has been linked to a default interventionist dual process model of reasoning, the primary goal of the present thesis was to test strategies associated with this popular and influential dual process model of reasoning as they pertain to medical diagnosis. [18,19,20,21] Specifically, this thesis addressed the strategies that slowing down, reflecting on a prior decision or removing distractions will improve accuracy.[18,19,20,21] I tested these strategies by relying on a basic assumption of a default interventionist dual process model that increased time and removal of distractions facilitates increased conscious, deliberative analytical processing and that distractions and time pressures reduce reliance on these processes. As such, the significance of this study is that variations in the amount of analytical processing are likely not responsible for variations in diagnostic accuracy. This conclusion is reached after two large studies investigating the impact of instructions to proceed slowly or quickly, experience, distractions and increased reflection on diagnostic accuracy, demonstrated that only experience, measured by years of training or professional practice, influenced accuracy.

With the exception of a main effect of experience reported in the 3 data chapters (i.e. two main studies), this thesis reports a series of null results or support for a null hypothesis of no difference in diagnostic accuracy as a result of experimental manipulations directed at altering the resources supporting analytical processing. Although not a traditional significant contribution, these null results are critical for refuting the assumptions of a default interventionist dual process model. In particular, these null results support the argument that medical education research should not continue to focus on education strategies that promote additional or improved reasoning.

Summary of Data Chapters

In chapter 3, I demonstrated that the processes involved in medical diagnosis are not compromised by time pressure or interruptions to have any measurable effect on diagnostic accuracy. Such a result contradicts the view that, because diagnostic reasoning is more susceptible to cognitive bias within a very busy and distracting environment, errors are more likely. Additionally, a default-interventionist dual process model of reasoning has been interpreted to suggest that increased reliance on System 2 processes is likely to improve performance in reasoning and decision making tasks.[31] If this is an accurate interpretation, then allowing physicians more time and removing environmental distractions should have improved performance. This was not the case. Instead, performance differed primarily as a result of experience level. Conversely, the lack of interference observed from these distractions is consistent with the perspective from the attention literature [57] that such tasks result in “task switching” which may cost some time (as we observed) but are unlikely to influence accuracy as long as the tasks are conceptually distinct.

The studies reported in chapter 4 and 5 demonstrated that offering physicians the opportunity to reflect does not result in a meaningful improvement in diagnostic accuracy. Additionally, similar to the results of the study in chapter

2, more experienced physicians were generally more accurate, regardless of instructions, opportunity to reflect or access to case details. The results of the study reported in chapter 4 suggest that reflection, or additional reasoning, alone cannot offer significant improvement in performance. Furthermore, the analyses conducted for the study reported in chapter 5, suggest that the willingness to reflect on and revise a diagnosis may be driven by a lack of knowledge or an understanding that an error has occurred. Therefore, it was not the case that an overall strategy to reflect improved performance on all cases; instead, reflection was being employed by subjects only when needed in the face of ambiguity. However, the nature of the reflection in the present studies was deliberately designed to possess ecological validity, encouraging clinicians to use whatever resources they chose to review and modify their conclusions. Without specific interventions to retrieve and reconfigure analytical knowledge, no improvement was observed. Importantly, while participants who engaged in reflection around specific cases were a minority of those who actually committed errors, their reflection resulted in only small improvements in accuracy; encouraging physicians to review all cases did not also encourage physicians to reflect equally on all cases. These results are significant particularly in light of previous work by Sylvia Mamede and colleagues. In several studies by Mamede and colleagues [45,47,48,49,50] the use of structured reflective practice resulted in small improvements in diagnostic accuracy. As has already been discussed, there are some methodological concerns with some of these studies, which motivated the study discussed in chapter 4. As well, a replication of Mamede et al. [49,50] showed no difference in performance between participants asked to rely on their first impression and participants asked to follow a more structured reflective method.[51] Critically, in all their studies, Mamede and colleagues[45,47,48,49,50] associated small improvements in accuracy to increased conscious reflection and changes in processing alone. The alternative, that the observed improvement may result from components of structured reflection that require participants to retrieve and reconfigure analytical knowledge, is not addressed by Mamede and colleagues. Further research in medical education should focus on understanding the underlying mechanisms responsible for the recruitment of knowledge or similar prior experience.

Discussion – Theoretical Implications

In this thesis I discussed a dual process model of reasoning, however, the larger issue is the continuing tension between a focus on thinking processes and a focus on organization of knowledge in memory which has characterized the field of medical education from the 1980's. I propose that thinking processes cannot be fully understood without also considering the involvement of memory. In medicine, it has already been demonstrated that training in generalized problem solving skills does not transfer to new cases,[2] so it is surprising that several authors continue to encourage training in generalized reasoning or reflection skills, keeping within the cycle of *diagnostic thinking*. I suggest a new perspective incorporating the concept of *diagnostic memory*.

Traditionally, memory research does not overlap with research on reasoning, logic and problem solving. This is an important distinction to recognize within psychology and within medical education. From a 'reasoning' perspective, it has been acceptable to investigate 'reasoning' processes as they relate to a complex problem, with the assumption that there is only one correct answer to any problem. Critically, the assumption that reliance on System 1 will result in errors is usually supported by research that assumes one correct answer to any complex problem. I propose that there can be more than one correct answer to a complex problem as the solution is entirely dependent on the desired outcome. For instance, while solutions that require a precise calculation, also require slow methodical processing, there are many problems for which an estimate will produce an identical or superior outcome. Studies of reasoning however, compare the performance of faster and slower reasoning processes using scenarios for which the more precise solution is the only one that is considered correct. That is classic research on thinking processes investigated human problem solving using artificial problems that would necessarily lead to an incorrect response when participants relied on estimation or allowed themselves to be influenced by the background context of the problem.[58] In essence, research on reasoning demonstrates that humans tend to rely more often on rapid experience or memory based processes to solve problems and when these processes are experimentally set up to recruit inappropriate information, accuracy will decline.[58] Because the decline in accuracy is experimentally induced, such findings cannot inherently be interpreted to suggest that reliance on experience or memory based processes is inferior.[58] However, there remain many authors who do support this interpretation.[18,19,20,21,31,43,44]

In addition, research on reasoning continues to assume a strictly behaviourist account of human learning and performance that characterizes associative memory processes as largely instinctual and unsophisticated. A classic view of goal oriented human behaviour contrasts a basic ability for learning stimulus-response pairings with a more sophisticated ability for conscious cognitive control. However, research on reasoning often focuses on logic problems and probability calculations, treating each new problem as unique. Participants in these studies are encouraged to consider only the information provided in the problem and the rules of logic while solutions influenced by experience are typically considered faulty.[58] The failure of most test subjects to produce the normative response can be interpreted as a true failure of human behaviour to meet the standards of reasoning, or as a failure of the reasoning model to explain behaviour.

Contemporary research on memory processes however, is not restricted to investigations of basic stimulus-response pairings and has challenged the classic view that humans consciously select either an automatic retrieval of a learned response or a controlled application of goal oriented rules. Current research on attention and memory processes suggests that humans are capable of rapid switches between automatic and controlled processes in an online or task dependent fashion.[27,56] In other words, there is growing evidence suggesting that optimal performance on a given task is supported by online shifts in processing styles, that is largely dependent on the demands of the task, rather than conscious recruitment of logical reasoning.

Conclusion

Within medical education, the heuristics and biases research program[31] has had a great deal of influence on theories about the processes underlying diagnosis. However, another way to interpret both heuristics (i.e. cognitive shortcuts) and bias (i.e. misleading shortcuts) is as memory for prior experience. In this light, it becomes very difficult to separate the processes that contribute to accurate diagnoses from the processes that contribute to inaccurate diagnoses. Prior experience may enable us to recognize that a problem has been encountered before. However the identification may be accurate (and with increasing experience the likelihood of being accurate will also increase) or may be in error. Regrettably, many of the studies of the role of experience and non-analytical reasoning have frequently employed a manipulation whereby prior experimenter-controlled experience will lead the subject astray.[11,13,15,46,53,59] Consequently it is easy to identify evidence that experience can result in bias, however this ignores the specific experimental conditions that led to the generation of errors.

The common theme emerging from the studies reported in this thesis is that experiential knowledge, and the rapid processing that it engenders is not a bias to be avoided or to be conquered by “reflection.” Rather, these experience derived heuristics represent an efficient and effective strategy consistent with the associative nature of human memory.

As expertise develops with increased exposure to a variety of experiences and practice opportunities, the goal of medical education should be to take advantage of memory processes to enhance performance, encourage reliance on memory, and prepare physicians to be able to rely on memory more often: the true sign of expertise. Therefore to reduce errors, medical education must focus on improving practice methods and giving medical students and residents the kinds of experiences that allow them to develop confidence and perform like an expert. The important next steps will be to identify more appropriate theories for learning and practice in medicine.

References for Introduction and Discussion

1 Charlin, B., Tardif, J., & Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Academic medicine*, 75(2), 182-190.

2 Elstein, A. S., Shulman, L. S., Sprafka, S. A., & Allal, L. (1978). *Medical problem solving: an analysis of clinical reasoning* (Vol. 2). Cambridge, MA: Harvard University Press.

3 Groen, G. J., & Patel, V. L. (1985). Medical problem‐solving: some questionable assumptions. *Medical education*, 19(2), 95-100.

4 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(5), 315-322.

5 Norman, G. R., Brooks, L. R., Colle, C. L., & Hatala, R. M. (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.

6 Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. Toward a general theory of expertise: Prospects and limits, Cambridge university press. 93-125.

7 Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and learning in medicine*, 13(2), 110-116.

8 Eva, K. W., Neville, A. J., & Norman, G. R. (1998). Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Academic medicine*, 73(10), S1-5.

9 Charlin, B., Boshuizen, H., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Medical education*, 41(12), 1178-1184.

10 Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Academic medicine*.

11 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.

12 Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of experimental psychology: Learning, memory, and cognition*, 15(6), 1166.

13 Norman, G., Young, M., & Brooks, L. (2007). Non_analytical models of clinical reasoning: the role of experience. *Medical education*, 41(12), 1140-1145.

14 Young, M. E., Brooks, L. R., & Norman, G. R. (2011). The influence of familiar non_diagnostic information on the diagnostic decisions of novices. *Medical education*, 45(4), 407-414.

15 Norman, G. (2005). Research in clinical reasoning: past history and current trends. *Medical education*, 39(4), 418-427.

16 Ark T., Brooks LR & Eva K. (2006) Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Academic medicine*, 81, 405-409.

17 Ark, T. K., Brooks, L. R., & Eva, K. W. (2007). The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Medical education*, 41(3), 281-287.

18 Croskerry, P. (2000). The cognitive imperative: thinking about how we think. *Academic emergency medicine*, 7(11), 1223-1231.

19 Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic medicine*, 78(8), 775-780.

20 Croskerry, P. (2003). Cognitive forcing strategies in clinical decisionmaking. *Annals of emergency medicine*, 41(1), 110-120.

21 Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: origins of bias and theory of debiasing. *British medical journal: quality & safety*.

22 Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of internal medicine*, 165(13), 1493.

23 Chisholm, C. D., Weaver, C. S., Whenmouth, L. F., Giles, B., & Brizendine, E. J. (2008). A comparison of observed versus documented physician assessment and treatment of pain: the physician record does not reflect the reality. *Annals of emergency medicine*, 52(4), 383-389.

24 Evans, J. S. B., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition Advancing the Debate. *Perspectives on psychological science*, 8(3), 223-241.

25 Medin, D. L. (1989). Concepts and conceptual structure. *American psychologist*; 44(12), 1469.

26 Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of experimental psychology: Learning, memory, and cognition*, 24(6), 1411.

26 Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes.

27 Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, 30(5), 513-541.

28 Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46(3), 441-517.

29 Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & reasoning*, 13(4), 321-339.

30 Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual review in psychology*, 59, 255-278.

31 Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

31 Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

32 Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of business*, S251-S278.

33 Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62, 451-482.

34 De Neys, W. (2006). Dual processing in reasoning two systems but one reasoner. *Psychological science*, 17(5), 428-433.

35 Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and brain sciences*, 23(5), 645-665.

36 Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education*,44(1), 85-93.

37 Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic medicine,* 84(8), 1022-1028.

39 Norman, G.R., Sherbino, J., Dore, K., Young, M., Gaissmeier, W., Monteiro, S., Kreuger. (2013) The etiology of diagnostic errors: A controlled trial of system 1 vs. system 2 reasoning. *Academic medicine, In Press*.

38 Sherbino, J., Dore, K.L., Siu, E., Norman, G.R. (2011). The effectiveness of cognitive forcing strategies to decrease diagnostic error: An exploratory study. *Teaching and learning in medicine*, 23 (1): 78-84.

40 Sherbino, J., Dore, K. L., Siu, E., & Norman, G. R. (2011). The effectiveness of cognitive forcing strategies to decrease diagnostic error: an exploratory study. *Teaching and learning in medicine*, 23(1), 78-84.

41 Sherbino J, Kulasegaram K, Howey E, Norman G. (2013). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. *Academic emergency medicine*.

42 Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *The teachers college record*, 104(4), 842-866.

43 Redelmeier, D. A. (2005). The cognitive psychology of missed diagnoses. *Annals of internal medicine*, *142*(2), 115-120.

44 Klein, J. G. (2005). Five pitfalls in decisions about diagnosis and prescribing. *British medical journal*, *330*(7494), 781.

45 Mamede, S., Schmidt, H.G., Rikers, R.M.J.P, Penaforte, J.C. and Coelho-Filho, J.M. (2007). Breaking Down Automaticity: Case Ambiguity and the Shift to Reflective Approaches in Clinical Reasoning. *Medical education*. 41: 1185 - 1192.

46 Mamede, S. and Schmidt, H. (2004). The Structure of Reflective Practice in Medicine. *Medical education*. 38: 1302 – 1308.

47 Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M., van Saase, J. L., van Guldener, C., & Schmidt, H. G. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA: The Journal of the American medical association*, *304*(11), 1198-1203.

48 Mamede, S., Schmidt, H. G., Rikers, R. M., Custers, E. J., Splinter, T. A., & van Saase, J. L. (2010). Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychological research*, 74(6), 586-592.

49 Mamede, S., Schmidt, H.G., Rikers, R.M.J.P., Penaforte, J.C., Coelho-Filho, J.M. (2008)a. Influence of perceived difficulty of cases on physicians'

diagnostic reasoning. Influence of Perceived Difficulty of Cases on Physicians' Diagnostic Reasoning. *Academic Medicine*. 83(12): 1210-1216.

50 Mamede, S., Schmidt, H. and Penaforte, J.C. (2008)b. Effects of reflective practice on the accuracy of medical diagnosis. *Medical Education*. 42: 468:475.

51 Ilgen, J.S., Bowen, J.L., McIntyre, L.A., Banh, K.V., Barnes, D., Coates, W.C., Druck, J., Fix, M.L., Rimple, D., Yarris, L.M., & Eva, K.W. (2013). The impact of instruction to use first impressions or directed search on candidate diagnostic performance and the utility of vignette-based assessment. *Academic medicine*. 2013.

52 Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, 48(4), 516.

53 Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine*, *13*(2), 110-116.

54 Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1-66.

55 Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*(2), 127-190.

56 Bugg, J. M., & Crump, M. J. (2012). In support of a distinction between voluntary and stimulus-driven control: a review of the literature on proportion congruent effects. *Frontiers in Psychology*, *3*.

57 Monsell, S. Task Switching. (2003)*Trends in Cognitive Sciences*,7(3): 134:140.

58 Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology*, *1*(1), 65-82.

59 Hatala, R., Norman, G. R., & Brooks, L. R. (1999). Influence of a single example on subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine*, *11*(2), 110-117.