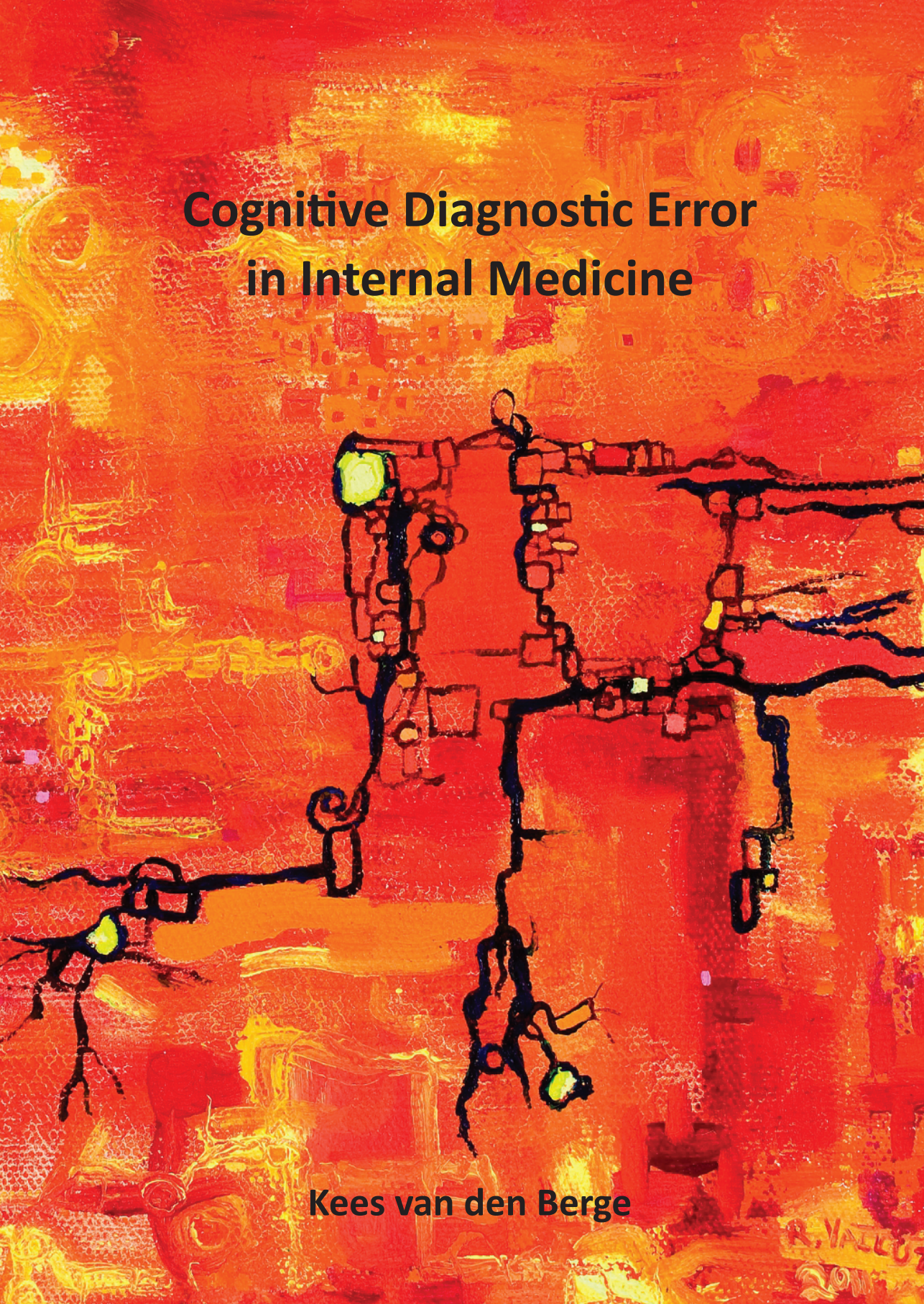
Cognitive Diagnostic Error in Internal Medicine



Kees van den Berge

Cognitive Diagnostic Error in Internal Medicine

Kees van den Berge

2012

The research presented in this thesis was conducted at the Department of Internal Medicine of the Erasmus MC and the Department of Psychology of the Erasmus University, Rotterdam, the Netherlands. Financial support for publication of this dissertation was kindly provided by Astellas Pharma B.V., Servier Nederland Farma B.V., J.E. Jurriaanse Stichting, Abbott B.V.

ISBN 978-94-6169-178-1

Van den Berge, Cornelis Krijn Adriaan
Cognitive diagnostic error in internal medicine
Thesis, Erasmus University Rotterdam, the Netherlands

Cover : Dendritic Echoes by Dr. Regina Valluzzi; The Nerdly Painter (with permission)
Lay-out : Kees van den Berge, Optima Grafische Communicatie
Printing : Optima Grafische Communicatie, Rotterdam

Cognitive Diagnostic Error in Internal Medicine

Cognitieve Diagnostische Fouten in de Interne Geneeskunde

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
donderdag 2 februari 2012 om 15:30 uur

door

Cornelis Krijn Adriaan van den Berge

geboren te Willemstad

PROMOTIECOMMISSIE

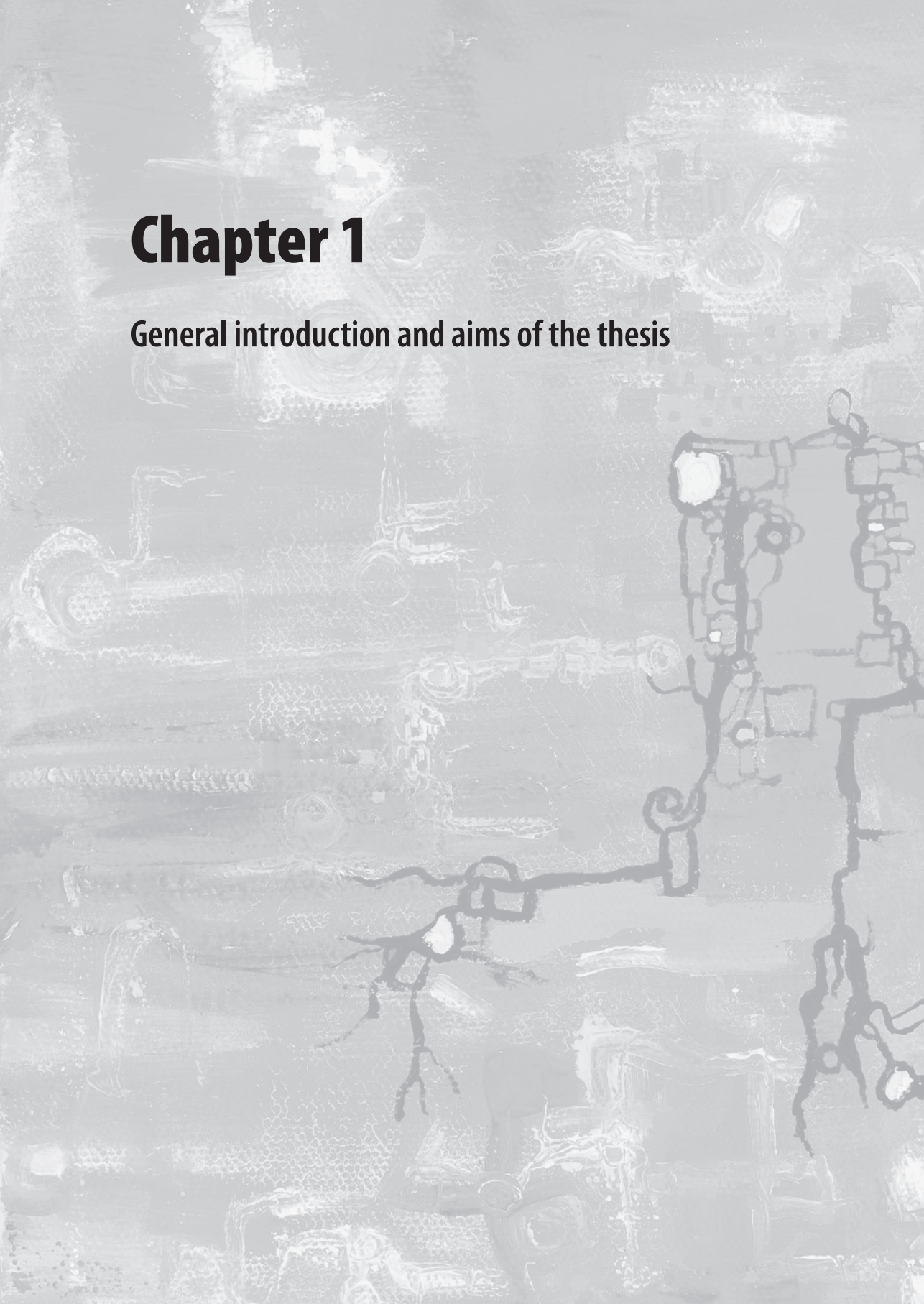
Promotoren:	Prof.dr. J.L.C.M van Saase
	Prof.dr. R.M.J.P. Rikers
Overige leden:	Prof.dr. R.P. Koopmans
	Prof.dr. H.T. van der Molen
	Prof.dr. F.G.W.C. Paas
Copromotor:	Dr. S. Mamede

CONTENTS

Chapter 1	General introduction and aims of the thesis	7
Part I	**Confirmatory tendencies in medical diagnosis**	
Chapter 2	Accepting incorrect diagnostic suggestions by residents: a potential cause of diagnostic error in medicine	25
Chapter 3	Supportive features mediate the tendency to accept diagnostic suggestions	37
Chapter 4	Physicians' accepting of diagnostic suggestions occurs independent of consistency in preceding suggestions	49
Part II	**The influence of availability bias and reflection on medical diagnosis**	
Chapter 5	Exposure to media information causes availability bias in medical diagnosis	63
Chapter 6	Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents	77
Chapter 7	Summary and general discussion	89
Chapter 8	Samenvatting	103
	Appendices	113
	Dankwoord	117
	PhD portfolio	123
	Curriculum Vitae	129

Chapter 1

General introduction and aims of the thesis



COGNITIVE DIAGNOSTIC ERROR IN INTERNAL MEDICINE

This thesis focuses on the subject of cognitive diagnostic error in internal medicine; mistakes resulting from flaws in physicians' reasoning processes. More specifically, this thesis addresses errors caused by confirmation and availability bias. Recently, the potential of cognitive factors to cause faults in diagnosis caught the attention of authors and policy-makers, and the topic is pursued in several position papers.[1-3] Nonetheless, little empirical evidence supporting a relationship between cognitive factors and medical error exists. In this doctoral thesis, five experimental studies will be presented that focus on this theme.

THE PROBLEM OF MEDICAL ERRORS

After the landmark Institute of Medicine (IOM) report "To Err is Human" on patient safety issues was published,[4] medical errors caught the attention of researchers and discussion on this subject increased.[5,6] In the IOM report, medical errors were found to be important causes of death in the United States. It was estimated that up to 98.000 Americans die annually as a result of those errors.[7,8] To illustrate, this figure surpasses the annual number of deaths due to motor vehicle accidents and Alzheimer's disease,[9] and medication errors alone cause more deaths than workplace injuries in the US.[10] Moreover, medical errors resulting in injury are a considerable financial burden. A rough estimation showed that those mistakes cost the American society between $17 billion and $29 billion annually.[11] Although most descriptive studies on medical error have been conducted in the US, the problem is not limited to the American healthcare system. On the contrary, medical error is ubiquitous, with figures and consequences worldwide comparable to the US.[12-16]

Besides errors related to treatment (e.g., medication errors, surgery-related errors), which are responsible for the majority of all mistakes, faults related to the diagnostic process are considered to contribute substantially to medical errors. Elstein estimates the rate of diagnostic error in clinical medicine in the order of 15%,[17] and his view is supported by findings from autopsy-studies, which indicate that such errors occur in approximately 10-15% of cases.[18,19] This rate is highest in the clinical specialties of internal medicine and emergency medicine.[7,8,20] The susceptibility for diagnostic mistakes in these specialties is believed to be related to domain-specific requirements, such as complex decision making in contexts of high uncertainty.[21] That is, in those specialties making a diagnostic decision is frequently a complex endeavor, because the spectrum of problems is much larger than in other specialties. In addition, in these specialties a diagnosis can often only be made if many findings are explained and interpreted in the context of an individual patient. For example, findings in one patient could be related as well to the correct diagnosis as to other, but incorrect, diagnoses. Furthermore, normal variations in test results can complicate the diagnostic process. For instance, the value of serum creatinine,

a routinely measured protein that may reflect kidney disease, is influenced, among others, by body mass, age and pregnancy.[22] The same value, then, could be considered as normal, borderline, or abnormal, depending on other aspects of the patient. Finally, adding up to uncertainty, clinical decisions, especially in emergency medicine, are often made under time pressure, which, for example, may impair thorough data-gathering, or are affected by other disturbances, such as telephone calls, which may divert attention from a patient.

COGNITIVE DIAGNOSTIC ERROR

Graber, Franklin, and Gordon provide highly informative data on diagnostic errors.[23] Their study describes 100 cases of diagnostic error in internal medicine identified retrospectively from five academic medical centers over a 5-year period. They defined diagnostic error as "a diagnosis that was unintentionally delayed, wrong, or missed, as judged from the eventual appreciation of more definitive information". They found diagnostic errors to usually result from multiple causes and typically involve both system-related and cognitive factors. System-related mistakes, for instance, due to technical failure, equipment problems and organizational flaws, were found to contribute to errors in 65% of cases. An even greater proportion of mistakes, however, could be attributed to faults in individual physicians' cognitive processes (74%). These so-called cognitive diagnostic errors may arise from inadequate knowledge, faulty data gathering, inaccurate clinical reasoning and erroneous verification of diagnostic hypotheses.[24] For example, a physician may focus on or overweigh clinical findings that are in line with an (incorrect) diagnostic hypothesis he has in mind, which could increase the chance of accepting this diagnosis. Several position papers stress the potential effect of cognitive errors on medical diagnosis,[1-3] and observational studies have provided indirect evidence that faults in physicians' cognitive processes may in fact be involved in the majority of missed or delayed diagnoses.[23,25] Such studies have indicated that faulty reasoning, more than knowledge gaps, may be the most frequent cause of cognitive errors.

The mechanisms underlying faulty clinical reasoning are, however, still subject of discussion.[26] Research on medical expertise, which describes how clinicians make judgments in clinical encounters with patients, suggests that clinicians' reasoning may be susceptible to bias.[27-29] For example, this research has shown that clinicians generate hypotheses early in the clinical encounter, mainly through pattern recognition: similarities in clinical characteristics between the current and previously seen patients quickly bring one or a few diagnostic hypotheses to the physician's mind, which are subsequently used to guide the search for additional information. This so-called non-analytical mode of reasoning, which tends to be a largely automatic, and therefore relatively effortless, process, is the dominant reasoning mode when clinicians deal with routine problems.[29] As it occurs largely without conscious control, generation of hypotheses based on pattern-recognition could be influenced by multiple factors that remain

unnoticed, making physicians more prone to bias and, consequently, to errors.[1,2] Experimental studies to establish a relationship between physicians' clinical reasoning and diagnostic error are, however, scarce.

REASONING MODES AND DIAGNOSTIC ERROR

Within the literature on clinical reasoning a debate is ongoing whether physician's reasoning modes predispose them to make cognitive diagnostic errors.[26,30] Some authors believe that an excessive reliance on non-analytical reasoning may lead to errors,[1,3] which could have been avoided if a strategy consisting of analytical and non-analytical reasoning would have been used.[31] In addition, there is evidence that diagnostic accuracy on complex clinical problems may improve if physicians rely on reflective, or analytical, reasoning.[32] Analytical reasoning is rule-based and systematic, and therefore costs more time and effort than non-analytical reasoning, which is based on holistic pattern-recognition rather than on intensive deliberation. The two modes of reasoning are the representatives of the so-called dual process view on reasoning, with non-analytical reasoning referred to as System 1 and analytical reasoning as System 2,[33,34] which is the dominant theoretical framework concerning research on clinical reasoning, judgment and decision making.[35,36]

The anatomical and physiological co-existence of two distinct reasoning systems is supported by experimental findings. For instance, an event-related fMRI-study showed that non-analytical thinking is associated with ventral medial prefrontal cortex activity and analytical thinking is reflected by activity in the right inferior prefrontal cortex.[37] Likewise, physiological evidence of the dual process view has been presented based on the assumption that analytical reasoning requires more glucose, the primary fuel for brain processes, than non-analytical reasoning. In their study, Masicampo and Baumeister show how blood glucose interacts with the reliance on non-analytical reasoning.[38]

The two modes of reasoning are simultaneously possible for clinicians, and research has shown case ambiguity (i.e., a case that is consistent with the typical pattern of a disease but also includes features consistent with alternative diagnoses) to be a determinant of reasoning strategy: more ambiguous clinical cases are associated with a shift from non-analytical to analytical approaches in clinical reasoning.[39] It is, however, unclear whether errors in medical diagnosis are associated with either analytical or non-analytical reasoning.[30] In addition, certain specific cognitive tendencies, such the inclination towards confirmation might, regardless of the interplay of reasoning modes, lead to diagnostic errors.

CONFIRMATORY TENDENCIES IN REASONING

People's inclination towards confirmation of hypotheses is often pointed out as an important cause of cognitive diagnostic errors.[1,2] This tendency has been studied for several decades in psychological research. If errors are made because data are gathered or interpreted to confirm a hypothesis rather than refute it, confirmation bias may be pointed out as the cause.[40] Research on confirmation bias started with an observation made by Wason.[41] He found that when a rule explaining the link between several numbers had to be discovered during problem solving, people think in terms of confirmation rather than disconfirmation. Confirmation bias is considered as one of the primary causes of error in the literature on reasoning[42] and, since the phenomenon was first demonstrated by Wason, its influence has been shown in several settings, such as judicial reasoning, forensic science, and gambling.[43-45]

This domain-specific research on confirmation bias is sometimes initiated by high profile cases, which are thought to be subject to the bias. For example, in forensic science, "The Mayfield Case" initiated such research.[44,46] In this case, the bias arose in the identification of fingerprints in the terrorist train bombing in Madrid in 2004 when the FBI misidentified a Mr. Mayfield as the source of the crime scene prints. That is, after re-evaluation, Spanish authorities exonerated Mayfield because they found the prints actually belonged to an Algerian citizen. To analyze why their investigation had failed, the FBI initiated an external audit. In the final report, the error was ascribed in part to confirmation bias (i.e., once the FBI's first forensic examiner linked the prints to Mayfield, he saw what he expected to see during his own re-examination. The subsequent second reviewer knew a positive identification had already been made, and could therefore also have been influenced by confirmation bias). Moreover, detection of mistakes was constrained due to the laboratory culture where critique was not embraced: as the report noted: "To disagree was not an expected response."[46]

Another study focused on confirmatory tendencies in social psychology.[45] This study showed a biased evaluation of outcomes by gamblers. In one of the experiments described in this study, gamblers took more time to explain away their losses than to explain their wins. In addition, gamblers generally discounted their losses and augmented their wins. This tendency could account for the continuation of gambling despite persistent failure; gamblers convince themselves that their chances of winning were higher than they actually are. Those findings can be seen as examples of an overweighting of supportive evidence and an underweighting of opposing evidence. Pyszczynski and Greenberg interpreted these observations as supportive of the idea that people in general require less hypothesis-consistent information to accept a hypothesis than hypothesis-inconsistent information to reject a hypothesis.[47]

CONFIRMATORY TENDENCIES IN CLINICAL REASONING

In the medical domain, research on physicians' confirmatory tendencies was mainly done with visual stimuli, and focused on the influence of diagnostic suggestions on the quality of diagnostic decisions.[48-49] This research showed that students and physicians could be biased towards a correct diagnosis or a plausible alternative diagnosis, by having them first evaluate the plausibility of either the correct or the alternative diagnosis. In one study,[49] medical students and residents were shown photographs of patients accompanied by a case history that was supportive of a subsequently presented suggested diagnosis. This diagnostic suggestion was either the correct or a plausible alternative diagnosis. Next, participants were asked to rate the likelihood of the suggested diagnosis. Results showed that participants who were first exposed to the correct diagnosis accurately decided that this diagnosis was indeed correct in 77.2% of the cases. When they were suggested the incorrect diagnosis, they erroneously decided that this one was correct in 65.8% of the cases. Results further showed that participants' difficulty with rejecting incorrect diagnoses probably arose because their identification and interpretation of clinical features was influenced by the suggested diagnosis.[48,49] For example, in one scenario, participants misinterpreted tanned skin as jaundice more often when biased toward liver cancer (which is frequently associated with jaundice) than when biased toward stomach cancer (which is less frequently associated with jaundice), despite the patient's white sclerae (which indicated that there definitely was no jaundice, since jaundiced skin is by definition accompanied by jaundiced sclera).

THE INTERDEPENDENCE OF FEATURE IDENTIFICATION AND DIAGNOSIS

A main goal for specialists involved in the training of residents and students is the prevention of mistakes in clinical reasoning of their apprentices. Having this in mind, they often recommend students to gather all information before making a diagnosis. This complete gathering of information, including the meticulous listing of the patient's presenting features, is believed to reduce the chances of premature closure.[1] However, as the abovementioned studies point out, the detection of clinical features may not be a clear-cut process.[48,49] In fact, some authors believe clinical features to be anything but independent cues that are processed and used as pieces of evidence in order to arrive at a diagnosis.[48-50] Instead, they argue that features are extracted and interpreted in light of the diagnoses that the physician has in mind.

For instance, research that used radiographs to study clinical decision making has shown that tentative diagnoses can increase diagnostic accuracy by drawing attention to features that might otherwise be missed.[51] In another study using chest radiographs, prior clinical histories for ambiguous cases of bronchiolitis affected not only the diagnoses but also the identification of the presence of particular clinical features.[50] Likewise, a study using ECGs showed that a

biasing history influenced both diagnostic accuracy and the identification of features.[52] In this study, a suggestive history prior to evaluating the ECGs guided internal medicine residents towards the diagnosis suggested by the history. Additionally, when a history supportive of the correct diagnosis was suggested, residents mentioned more features compatible with this diagnosis and fewer features compatible with the plausible alternative diagnosis. In contrast, when a history supportive of a plausible alternative diagnosis was given, residents were more likely to leave out features compatible with the correct diagnosis, and report more features that were compatible with both the correct and alternative diagnoses. Similar results have been found in studies using photographs of patients.[48,49] This tendency to identify and interpret clinical features in light of a suggested diagnosis may lead to diagnostic errors through susceptibility to confirmation bias.

The previous studies showed that diagnostic decisions may be influenced by diagnostic suggestions, and that the tendency to confirm a suggestion may be caused by a reinterpretation or misidentification of clinical features. In these studies, however, the evaluation of the suggested diagnosis was preceded by a biased (i.e., supportive) case history, which is likely to have led to participants' difficulties with rejecting incorrect diagnoses. In fact, because different case histories accompanied the same photograph, in these studies, the cases *themselves* can be viewed upon as being different.

In addition, those studies used highly visual materials, such as pictures of patients, which may be more affected by interpretation. That is, visual images are probably more powerful in opening space for noticing features that would probably not be expressed in a written case description. For instance, a tanned skin shown in a patient's picture could be misinterpreted as jaundice, if a wrong diagnosis of liver carcinoma is suggested. But in a written case, the "tanned skin" would not be mentioned, rather either the sclerae would be described or nothing would be said. Therefore, the results of previous studies[48-52] cannot be generalized to other types of medical stimuli, such as internal medicine cases. That is, it remains to be investigated whether such confirmatory tendencies also occur when physicians evaluate suggested diagnoses on *written* clinical internal medicine cases, since doctors are known to rely on case histories,[48] and the specialty of internal medicine is often implicated as being affected by diagnostic error.[7,8,20]

THE EFFECT OF THE EASE OF INFORMATION-RETRIEVAL FROM MEMORY

Besides confirmatory tendencies, other potential cognitive pitfalls exist. For example, if information from an unlinked source influences medical diagnosis and causes a diagnostic mistake, the physician has fallen prey to a cognitive error called availability bias.[1] Psychological research has shown that such a flaw in reasoning processes is often evoked by the use of heuristics, which are defined as mental shortcuts that are invoked, largely unconsciously, by physicians to accelerate the decision-making process.[53] Heuristics may be helpful in most situations but

they may also lead to biased decisions.[54] The availability heuristic leads people to assess the likelihood or probability of an event by the ease with which instances can be retrieved from memory.[54,55] As Tversky and Kahneman pointed out, this heuristic is generally rather useful, as events that occur more frequently are easier to recall or imagine than less frequent ones.[55] However, the ease with which instances come to one's mind is also influenced by other issues, such as someone's own recent experiences. Consequently, the use of this heuristic may generate an *availability bias*, resulting in errors of judgment.

In psychology, several demonstrations of this bias are given. An illustration comes from an experiment that is described in a seminal paper by Tversky and Kahneman.[55] The authors present a study in which participants heard a list of names of both sexes. Next, they were asked to judge whether the list contained more names of men or women. Different groups of participants listened to different lists. In some of the lists the men were more famous than the women, and in others the women were more famous than the men. In each of the lists, the subjects erroneously judged that the sex that had the more famous individuals was the more numerous. This indicates that such frequency judgments are mediated by availability, since famous names are generally easier recalled than not so well-known names.

Another study in psychological research showed that the dramatic nature of certain events could also cause availability bias in reasoning.[56] In this experiment participants were asked to estimate the frequencies of particular causes of death. The causes of death were presented in pairs, with one member being very dramatic (e.g., floods) and the other less spectacular (e.g., asthma). The authors hypothesized that the frequency of the more dramatic causes would be overestimated, and that of less dramatic causes would be underestimated because the media's underreporting of these causes of death would have led to a biased knowledge base in the participants' memory. Indeed, the more dramatic causes were overestimated in this study, contrary to the objective frequencies, again indicating that frequency judgment of lethal events were influenced by availability.

Such biased reasoning may have substantial consequences. This is illustrated by an availability related effect of the September 11 terrorist attacks: people's diminished enthusiasm to fly because of dread risk.[57] Dread risk is defined as people's inclination towards avoiding circumstances in which many people could die simultaneously, while being relatively unmoved by risky situations in which deaths are less condensed. Ironically, avoiding the dread risk of flying after the September 11 attacks may have led to an elevated number of traffic fatalities. In the three months after the 9/11 attacks, the number of fatal traffic accidents was intensely elevated relative to baseline rates from other years during the same period, apparently because people decided to avoid air travel due to fear for a repetition of the September 11 tragedy. This illustrates that the availability bias might have caused numerous indirect fatalities due to the influence of this dramatic event on people's judgment and decisions.[58]

AVAILABILITY BIAS AND CLINICAL REASONING

The medical literature positions the availability bias as a main contributor to errors of judgment in diagnosis, despite the scarcity of empirical evidence encountered in this literature.[1-3,59-61] At least one non-experimental study provides a clue that availability bias may influence medical diagnosis.[62] This study suggested that physicians may indeed be influenced by what they read in the newspaper or saw on the television. During outbreaks of West Nile fever in the United States and Israel, an association between the frequency of reports in the media about the virus and the number of ordered lab-tests for disease was shown. The authors proposed that this rise in testing requests was due to an availability bias in the physicians involved, caused by the media. Yet, ordering such a test does not necessarily imply that the physician is biased. That is, if West Nile fever is prevalent, it somewhat raises the odds that a patient with symptoms similar to those of that disease actually has West Nile fever. Testing for the virus would then not be a demonstration of bias; it could also denote a careful physician. In addition, it is unsure whether the physicians concerned were actually convinced that their patients had the disease. That is, extensive media coverage might just as well have influenced their patients who could have believed that they were infected, and therefore could have claimed testing. Such interpretations would be clarified by a controlled experimental study.

STRATEGIES TO MAKE PHYSICIANS LESS PRONE TO COGNITIVE DIAGNOSTIC ERROR

Equally scarce as proof for cognitive factors' ability to cause diagnostic errors is empirical evidence in support of educational strategies aimed at making physicians less susceptible to bias. The medical literature has raised attention for "debiasing strategies" and metacognitive training as a method to reduce error and improve diagnostic decision making.[1,21] For example, enthused by parallels between the fields of medicine and aviation,[63] Singh, Petersen, and Thomas suggested that enhancing situational awareness (i.e., a shared understanding of "what's going on" and "what is likely to happen next") might reduce diagnostic errors.[64] Others believe that stimulating physicians' awareness of possible biases under conditions of non-analytical reasoning would make them less vulnerable to errors.[1] However, there is more descriptive research advocating such strategies as a means to counteract bias than experimental evidence that they actually *do*.

A few experimental studies have suggested that reflective reasoning may be able to counteract bias.[32,65] Reflective reasoning is defined as the critical appraisal one's own reasoning and decisions. It involves clinicians to take the time to generate possible alternative explanations in response to an unfamiliar problem. After generating alternative hypotheses, reflective reasoning requires the physician to actively think about the consequences of these hypotheses.

That is, the reflective physician must consider the possible signs and symptoms that should be present if one of these hypotheses were correct. Then, the hypotheses should be tested against the problem at hand. Next, the physician must be willing to reflect when facing a complex or unfamiliar problem. Finally, reflective clinicians should possess the ability to meta-reason, that is, the ability to think about one's own thinking processes and to critically review own assumption or beliefs regarding a problem.[66]

In two experimental studies,[32,65] where reflection was operationalized, and its effects on diagnostic performance could be measured, internal medicine residents diagnosed simple and complex written clinical cases, either through pattern recognition (non-analytical reasoning) or reflective (analytical) reasoning. The results indicated that structured reflection increases diagnostic accuracy on complex cases for internal medicine residents, and creates rationale for the hypothesis that reflection may decrease the chances of diagnostic error caused by bias.

AIMS AND OUTLINE OF THE DISSERTATION

At present, there is hardly any experimental evidence that cognitive factors are able to cause errors in medical diagnosis. The research presented in this doctoral thesis is designed to investigate the relation between cognitive factors (i.e., confirmatory tendencies and availability bias) and diagnostic errors using an experimental approach. The main hypothesis is that these cognitive factors have the potential to cause diagnostic mistakes in physicians. The studies presented in **Chapters 2** through **6** can be roughly divided into two parts. **Chapters 2, 3,** and **4** focus on the potential of diagnostic suggestions to cause errors in diagnostic decisions due to confirmatory tendencies in physicians' reasoning. The studies in **Chapters 5** and **6** tested whether diagnostic performance can be impaired by availability bias while physicians diagnose clinical cases under conditions of non-analytical reasoning. Additionally, the studies in these final 2 chapters also investigated the potential of reflective reasoning to counteract bias and to improved diagnostic performance.

THE STUDIES

Chapter 2 presents the results of an experiment, designed to investigate the influence of diagnostic suggestions on physicians' diagnostic conclusions, using written clinical cases. It was hypothesized that physicians would tend to go along with a suggested diagnosis and, therefore, would have more difficulty rejecting incorrect diagnostic suggestions than accepting correct ones. If this happens, suggested diagnoses may evoke confirmatory tendencies and consequently may lead to diagnostic errors, if the suggestion happens to be incorrect. The study presented in **Chapter 3** attempts to explain confirmatory tendencies in clinical

decision-making. It is hypothesized that physicians tend to focus on and hence report more clinical features that support the suggested diagnostic hypotheses, while ignoring features that speak against those suggestions. In addition, to check whether erroneous decisions may have been caused by the tendency to gather data in line with the suggestion, we analyze whether participants report a higher percentage of features supportive of incorrect diagnostic suggestions when they accept them than when they reject them. **Chapter 4** investigates whether residents' inclination towards accepting diagnostic suggestions depends on consistency in preceding diagnostic suggestions. That is, whether acceptation of diagnostic suggestions could be influenced if physicians would first encounter a number of correct suggestions followed by a number of incorrect suggestions, and vice versa. It was hypothesized that more incorrect suggestions would be accepted if participants had first evaluated a series of correct suggestions. On the other hand, it was hypothesized that, if participants had first evaluated incorrect suggestions, they would become more critical about the suggestions, and hence will become more inclined to reject correct suggestions.

The study presented in **Chapter 5** investigates whether exposure to media information about a disease would cause availability bias in internal medicine residents, resulting in diagnostic errors, and whether reflection would counteract the bias. In **Chapter 6**, we investigate whether recent experiences with patients have the potential to generate an availability bias in non-analytical clinical reasoning, thereby provoking errors. In addition, since non-analytical reasoning develops in association with experience, we hypothesize the effect of the bias to be greater in more experienced physicians. In addition, this study investigates whether reflective reasoning may counteract such bias.

Finally, **Chapter 7** provides a summary and general discussion of the main findings reported in the studies of this dissertation and gives directions for future research.

REFERENCES

1. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775-80.
2. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *Brit Med J* 2005;330:781-4.
3. Redelmeier DA. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115-20.
4. Kohn KT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
5. Stelfox HT, Palmisani S, Scurlock C, Orav EJ, Bates DW. The "To Err is Human" report and the patient safety literature. *Qual Safety Health Care* 2006;15:174-78.
6. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal Trends in Rates of Patient Harm Resulting from Medical Care. *New Engl J Med* 2010;363:2124-34.
7. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study 1. *New Engl J Med* 1991;324:370-6.
8. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;38:261-71.
9. Heron M. Deaths: Leading causes for 2006. *National Vital statistics Reports* 2010;14:7.
10. Phillips DP, Christenfeld N, Glynn LM. Increase in US medication-error deaths between 1983 and 1993. *The Lancet* 1998;351:643-44.
11. Thomas EJ, Studdert DM, Newhouse JP, Zbar BI, Howard KM, Williams EJ, Brennan TA. Costs of Medical Injuries in Utah and Colorado. Inquiry 1999;36:255-64.
12. Weingart SN, Wilson RMcL, Gibberd RW, Harrison B. Epidemiology of medical error. *Brit Med J* 2000;320:774-777.
13. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *Brit Med J* 2001;322:517-519.
14. Shaw R, Drever F, Hughes H, Osborn S, Williams S. Adverse events and near miss reporting in the NHS. *Qual Saf Health Care* 2005;14:279-283.
15. Zegers M, De Bruijne MC, Wagner C, et al. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. *Qual Saf Health Care* 2009;18:297-302.
16. Aranaz-Andrés JM, Aibar-Rémon C, Vitaller-Burillo J, et al. Impact and preventability of adverse events in Spanish public hospitals: results of the Spanish National Study of Adverse Events (ENEAS). *Int J Qual Health Care* 2009;21(6):408-414.
17. Elstein AS. Clinical reasoning in medicine. In: Higgs J, Jones MA, eds. *Clinical Reasoning in the Health Professions*. Woburn, Mass: Butterworth-Heinemann;1995. p. 49-59.
18. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three different eras. *New Engl J Med* 1983;308:1000-5.
19. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy detected diagnostic errors over time. *JAMA* 2003;289:2849-56.
20. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Australia* 1995;163:458-71.
21. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121(5):S2-S23.
22. Perrone RD, Madias NE, Levey AS. Serum creatinine as an index of renal function: New insights into old concepts. *Clin Chem* 1992;38(10):1933-53.

23. Graber ML, Franklin N,. Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493-9.
24. Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: What's the goal? *Acad Med* 2002;77:981-92.
25. Singh H, Thomas EJ, Khan MM, & Peterson LA. Identifying diagnostic errors in primary care using an electronic screening algorithm. *Arch Intern Med* 2007;167:302-8.
26. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* 2010;44:94-100.
27. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise - Theory and implications. *Acad Med* 1990;65:611-21.
28. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev* 1993;5:205-21.
29. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ* 1997;2:173-84.
30. Norman GR. Dual-processing and diagnostic errors. *Adv Health Sci Educ* 2009;14:37-49.
31. Eva KW, Hatala RM, LeBlanc VR, Brooks LR. Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ* 2007;*41*:1152–58.
32. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008;42:468-75.
33. Evans JSBT. In two minds: Dual-process accounts of reasoning. *Trends Cogn Sci* 2003;10:454-59.
34. Evans JSBT. Dual processing accounts of reasoning, judgment and social cognition. *Ann Rev Psychol* 2008;59:255-78.
35. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;1:98-106.
36. Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ* 2007;41:1140-45.
37. Goel V, Dolan RJ. Explaining modulation of reasoning by belief. *Cognition* 2003;87:B11-B22.
38. Masicampo EJ, Baumeister RF. Toward a Physiology of Dual-Process Reasoning and Judgment: Lemonade, Willpower, and Expensive Rule-Based Analysis. *Psychol Sci* 2008;3:255-60.
39. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ* 2007;41:1185-92.
40. Nickerson RS. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev Gen Psychol* 1998;2:175-220.
41. Wason PC. On the failure to eliminate hypotheses in a cognitive task. *Q J Exp Psychol* 1960;12:129-40.
42. Evans JSBT. Bias in human reasoning: causes and consequences. NJ, Erlbaum; 1989.
43. Kuhn D, Weinstock M, Flaton R. How well do jurors reason? Competence dimensions of individual variations in a juror reasoning task. *Psychol Sci* 1994;5:289-96.
44. Dror IE, Charlton D, Péron AE. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci Int* 2006;156:74-8.
45. Gilovich T. Biased evaluation and persistence in gambling. *J Pers Soc Psychol* 1983;44:1110-26.
46. Stacey RB. "Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case." *J Forensic Ident* 2004;54(6):706-18.
47. Pyszczynski T, Greenberg J. Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis testing model. *Adv Exp Soc Psychol* New York: Academic Press; 1987:297-340.
48. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med* 2001;76:S18-20.

49. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med* 2002;77:S67-9.
50. Norman GR, Brooks LR, Coblentz CK, Babcook CJ. The correlation of feature identification and category judgments in diagnostic radiology. *Mem Cognition* 1992;4:344–355.
51. Berbaum KS, Franken EA, Dorfman DD, Barloon T, Ell SR, Lu CH, Smith W, Abu-Yousef MM. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. *Invest Radiol* 1986;21:532–539.
52. Hatala RM, Norman GR, Brooks LR. The impact of a clinical scenario upon electrocardiogram interpretation. *J Gen Intern Med* 1999;14:126–129.
53. McDonald CJ. Medical heuristics: The silent adjudicators of clinical practice. *Ann Intern Med* 1996;124:56-62
54. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185:1124-31.
55. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychol* 1973;5:207-32.
56. Lichtenstein S, Slovic P, Fischhoff B, Layman M, Combs B. Judged frequency of lethal events. *J Exp Psychol Human* 1978;4(6):551-78.
57. Slovic P. Perception of risk. *Science* 1987;236(4799):280-85.
58. Gigerenzer G. Dread Risk, September 11, and Fatal Traffic Accidents. *Psychol Sci* 2004;15:286-7.
59. Bornstein BH, Emler AC. Rationality in medical decision making: A review of the literature on doctors' decision making biases. *J Eval Clin Pract* 2001;7(2):97-107.
60. Dawson NV, Arkes HR. Systematic errors in medical decision making: Judgment limitations. *J Gen Intern Med* 1987;2:183-7.
61. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *Brit Med J* 2002;324:729-32.
62. Brezis M, Halpern-Reichert D, Schwaber MJ. Mass media-induced availability bias in the clinical suspicion of West Nile fever. *Ann Inter Med* 2004;140(3):234-5.
63. Durso FT, Drews FA. Health care, aviation, and ecosystems: A socio-natural systems perspective. *Curr Dir Psychol Sci,* 2010;19:71-5.
64. Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: A lesson from aviation. *Qual Safety Health Care* 2006;15:159-64.
65. Mamede S, Schmidt HG, Rikers RMJP, Custers EJFM, Splinter TAW, Van Saase JLCM. Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychol Res* 2010;74:586-592.
66. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ* 2004;38:1302-8.

Part I

Chapter 2

Accepting Diagnostic Suggestions by Residents: A Potential Cause of Diagnostic Error in Medicine[1]

[1] This Chapter is in press as: Van den Berge K, Mamede S, Van Gog T, Romijn JA, Van Guldener C, Van Saase JLCM, Rikers RMJP. Accepting Diagnostic Suggestions by Residents: A Potential Cause of Diagnostic Error in Medicine. *Teaching and Learning in Medicine*, 2012.

ABSTRACT

Background: Psychological research has shown that people tend towards accepting, rather than refuting hypotheses. Diagnostic suggestions may evoke such confirmatory tendencies in physicians which may lead to diagnostic errors. *Purpose*: This study investigated the influence of a suggested diagnosis on physicians' diagnostic decisions on written clinical cases. It was hypothesized that physicians would tend to go along with the suggestions and therefore would have more difficulty rejecting incorrect suggestions than accepting correct suggestions. *Methods*: Residents ($N = 24$) had to accept or reject suggested diagnoses on 6 cases. Three of those suggested diagnoses were correct, and 3 were incorrect. *Results*: Results showed the mean correct evaluation score on cases with a correct suggested diagnosis ($M = 2.21$, $SD = 0.88$) was significantly higher than the score on cases with an incorrect suggested diagnosis ($M = 1.42$, $SD = 0.97$), meaning physicians indeed found it easier to accept correct diagnoses than to reject incorrect diagnoses ($t(23) = 2.74$, $p < 0.05$, $d = 0.85$), despite equal experience with the diagnoses. *Conclusion*: These findings indicate that suggested diagnoses may evoke confirmatory tendencies and consequently may lead to diagnostic errors.

INTRODUCTION

An admitted patient is often accompanied by a diagnostic suggestion from a medical professional. For example, a nurse or the general practitioner may have already formed an opinion on the diagnosis and share this hypothesis with the doctor in the emergency room. Although generally helpful, such suggested diagnoses could be a source of error. Research in psychology has shown that people tend to accept, rather than refute hypotheses.[1] So when the diagnostic suggestion would happen to be incorrect, but the physician would be inclined to accept it, a diagnostic error with potential harmful effects could be made. This experimental study investigated the influence of suggested diagnoses on internal medicine residents' diagnostic decisions.

According to a report by the Institute of Medicine, medical errors are among the leading causes of death in the United States.[2] Besides errors related to treatment (e.g., medication errors), which cover the majority of all mistakes, faults related to the diagnostic process are considered to contribute to a high proportion of medical errors. The rate of those diagnostic errors is approximately 10-15% as shown by autopsy studies[3,4] and is highest in the specialties of internal medicine, emergency medicine, and family medicine.[5-7]

Diagnostic errors usually result from multiple causes, but a substantial proportion of mistakes can be attributed to faults in individual physicians' cognitive processes.[8] These so-called cognitive diagnostic errors may arise from inadequate knowledge, faulty data gathering, inaccurate clinical reasoning and erroneous verification of diagnostic hypotheses.[9,10] The potential effect of cognitive errors on medical diagnosis has been stressed in position papers,[11-13] and observational studies have provided indirect evidence that faults in physicians' cognitive processes may in fact be involved in the majority of missed or delayed diagnoses.[8,14] Such studies have indicated that faulty reasoning, more than knowledge gaps, may be the most frequent cause of cognitive errors.

The mechanisms underlying faulty clinical reasoning are still subject of discussion,[15] but research on medical expertise[16-18] suggests that doctors' reasoning may be susceptible to bias. This research has shown that doctors generate hypotheses early in the clinical encounter, mainly through pattern recognition: similarities in clinical characteristics between the current and previously seen patients quickly bring a diagnosis to the doctor's mind, determining the course of action that the doctor will take. This so-called non-analytical mode of reasoning, which tends to be a largely automatic, and therefore relatively effortless process, is the dominant reasoning mode when clinicians deal with routine problems.[18] As it occurs largely without conscious control, generation of hypotheses based on pattern-recognition could be influenced by multiple factors that remain unnoticed, making physicians more prone to bias and, consequently, to errors.[11-12]

When patients are admitted, they often arrive along with diagnostic suggestions of a colleague (e.g., the general practitioner, a nurse, or the ambulance personnel). If physicians would

tend towards accepting such suggestions, that is, would be prone to confirmation bias, this would lead to errors when these suggested diagnoses are wrong, which is not unusual.[1] Confirmation bias is attributed to people's tendency to seek or interpret data to confirm rather than disconfirm their hypotheses.[11] This tendency has been studied for over half a century, starting with the observation made by Wason[19] that when a rule explaining the link between several numbers had to be discovered during problem solving, people think in terms of confirmation rather than disconfirmation. Confirmation bias is pointed out as one of the primary causes of error in the literature on reasoning.[20]

Since this tendency was first shown by Wason,[19] its adverse impact has been demonstrated in several settings, such as judicial reasoning,[21] forensic science[22] and gambling.[23] For example, Gilovich[23] showed a biased evaluation of outcomes by gamblers. In one of the experiments described in his study, gamblers took more time to explain away their losses than to explain their wins. In addition, gamblers generally discarded their losses and augmented their wins. This tendency could account for the continuation of gambling despite persistent failure; they convince themselves that their chances of winning are higher than they actually are. Those early findings demonstrate an overweighting of supportive evidence and an underweighting of opposing evidence. Pyszczynski and Greenberg[24] interpreted these observations as supportive of the idea that people in general require less hypothesis-consistent information to accept a hypothesis than hypothesis-inconsistent information to reject a hypothesis.

Studies within the medical domain indicate that the influence of a previous diagnosis on physicians' judgments may in fact open the door for confirmation bias.[25-27] LeBlanc et al.[26] showed that students and residents in family medicine can be biased towards a correct diagnosis or a plausible alternative diagnosis by having them first evaluate the plausibility of either the correct or the alternative diagnosis. In this study, medical students and residents were shown photographs of patients accompanied by either a correct or incorrect diagnostic suggestion preceded by a case history that was supportive of the suggested diagnosis. Participants were subsequently asked to rate the likelihood of the suggested diagnosis. Results showed that participants who were first exposed to the correct diagnosis accurately decided that this diagnosis was indeed correct in 77.2% of the cases. When they were suggested the incorrect diagnosis they erroneously decided that this one was correct in 65.8% of the cases. Their data further showed that participants' difficulty with rejecting incorrect diagnoses probably arose because their interpretation of clinical features was influenced by the suggested diagnosis.

In the LeBlanc et al. studies, however, the evaluation of the suggested diagnosis was preceded by a biased case history which accompanied the photograph. This case history was supportive of the subsequently presented suggested diagnosis, which may have led to participants' difficulties with rejecting incorrect suggested diagnoses. In fact, because the same photograph was accompanied by a different case history in each condition (i.e., corroborative of a subsequently presented suggested diagnosis), one could say it was not only the accuracy of the suggested diagnosis that differed but also the cases themselves were different. In addition,

those studies used highly visual materials, such as pictures of patients, and can therefore hardly be generalized to other types of medical stimuli, such as internal medicine cases.[25,26]

The present study investigates whether the influence of a suggested diagnosis occurs in clinical cases within the domain of internal medicine and without the addition of other potentially biasing information (i.e., different histories that may either support or speak against the suggested diagnosis). It is hypothesized that physicians (i.e., residents in internal medicine) will perform less well on cases with an incorrect suggested diagnosis than on cases with a correct suggested diagnosis, because they will have more difficulty rejecting an incorrect diagnostic suggestion than accepting a correct suggested diagnosis and accordingly, are at risk for errors due to confirmation bias.

METHOD

Participants

A convenience sample of 24 first-and second-year internal medicine residents (17 female) from the Leiden University Medical Centre volunteered to participate in this study. The ethics review committee from the Department of Psychology, Erasmus University Rotterdam, provided approval for this study. Debriefing and feedback were provided when the study was completed, and participants' written consent was obtained.

Materials

A set of 6 written clinical cases was used in this study (see Appendix for an example of a case used in the study). All cases were based on real patients and had a confirmed diagnosis. They were designed and validated independently by 2 experts in internal medicine and had been previously used in studies with internal medicine residents.[28] The cases were presented to the participants in a booklet, showing one case per page. Each case started with a diagnostic suggestion, which was to be evaluated by the participants. Directly after each case description they had to indicate whether they agreed or disagreed with the suggested diagnosis (The Appendix shows how the task was presented to participants).

There were two versions of the booklet. That is, the cases were the same and were presented in the same order in every booklet, but the diagnostic suggestions that preceded them differed: Half of the participants started with a correct suggested diagnosis, the other half started with an incorrect suggested diagnosis, after which the correctness of the suggested diagnoses alternated (i.e., an incorrect suggested diagnosis was always followed by a correct suggested diagnosis and vice versa, see Table 1 for correct and incorrect suggested diagnoses). The two versions of the booklet were randomly assigned to the participants.

After finishing the evaluation of all diagnoses, participants were asked to self-report their level of experience with all 12 suggested diagnoses (i.e., the 6 correct and the 6 incorrect

Table 1: Cases used in the study with either the correct or an incorrect diagnostic suggestion.

Case	Correct diagnosis	Incorrect diagnosis
1	Aortic dissection	Myocardial infarction
2	Viral infection	Q fever
3	Pneumococcal pneumonia	Legionnaires' disease
4	Inflammatory Bowel Disease	Irritable Bowel Syndrome
5	Clostridium Colitis	Ulcerative Colitis flare-up
6	Liver cirrhosis	Liver metastasis

suggestions) on a 7-point Likert scale, ranging from 1 (no experience with the disease) to 7 (highly experienced with the disease). Note that depending on the version of the booklet they had been assigned, they had either seen the correct or the incorrect alternative while evaluating suggested diagnoses.

Procedure

The study was conducted during an educational session at the Leiden University Medical Centre. It was introduced as an opportunity for the residents to contribute to the design of a new educational program in which learning based on clinical cases would play a pivotal role. The instruction for evaluation of the cases was provided in the booklet: "Read the following cases quickly but carefully and indicate whether you agree or disagree with the suggested diagnosis." Participants were given 75 seconds to evaluate each diagnosis, which a pilot study had shown to be sufficient for reading all case information. Time was kept by an experiment-leader who told the participants to continue to the next case after every 75 seconds. After finishing the evaluation of all 6 suggested diagnoses, demographic questions were asked (gender, months of clinical experience). The whole experiment took about 15 minutes.

Data Analysis

All 24 participants finished the experiment and provided consent. The data were scored as follows: for each correct evaluation (i.e., agreeing to the correct suggested diagnosis or disagreeing with the incorrect suggested diagnosis) a score of 1 point was assigned. So, 3 points could be obtained for rejecting incorrect diagnoses and 3 points for confirming correct diagnoses. A paired samples *t*-test was used to compare this score on cases accompanied by a correct suggested diagnosis with this score on cases accompanied by an incorrect suggested diagnosis. This test was also used to compare participants' experience with the diagnoses presented as correct suggestions with their experience with the diagnoses presented as incorrect suggestions). For all analyses, a significance level of 0.05 is used. For the *t*-test, *d* is reported as a measure of effect size, with values of 0.20, 0.50, and 0.80 corresponding to small, medium, and large effect sizes, respectively.[29] The scoring procedure was straightforward and was therefore conducted by a single rater.

RESULTS

Experience with suggested diagnoses

Differences in reported experience (range: 6-42) with diagnoses presented as correct suggestions ($M = 26.25$, $SD = 3.27$) and incorrect suggestions ($M = 25.71$, $SD = 3.00$), were not significant $t(23) = 0.73$, $p = 0.47$, $d = 0.17$.

Diagnostic decisions

The participants' performance concerning diagnostic decisions are shown in Figure 1. In line with the hypothesis, it was found that participants' mean correct evaluation score on the 3 cases with a correct suggested diagnosis was significantly higher ($M = 2.21$, $SD = 0.88$) than that on the 3 cases with an incorrect suggested diagnosis ($M = 1.42$, $SD = 0.97$), $t(23) = 2.74$, $p < 0.05$, $d = 0.85$. This implies that the odds of making an incorrect diagnostic decision more than double, when participants are confronted with incorrect suggestions. On average, 53% of incorrect suggestions were accepted (i.e., (3 – 1.42) / 3) and 26% of correct suggestions were rejected (3 – 2.21) / 3).

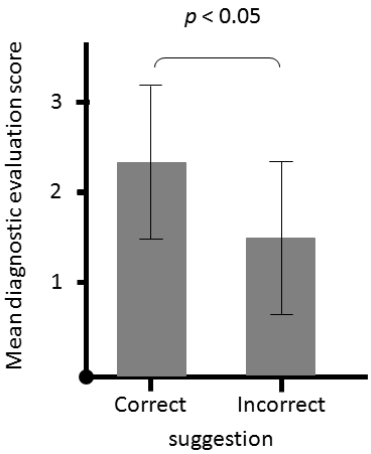


Figure 1: Mean diagnostic evaluation scores and standard deviations for cases with correct and incorrect diagnostic suggestions.

DISCUSSION

This study investigated the effect of suggested diagnoses on diagnostic decisions by internal medicine residents. All participants were exposed to the same cases; only the suggested diagnosis differed (i.e., correct or incorrect). Results showed that in line with the hypothesis,

residents had greater difficulty rejecting incorrect diagnoses than accepting correct diagnoses, despite equal experience with the suggested diagnoses.

The present study built on studies that used visual stimuli and in which the evaluation of the suggested diagnoses was preceded by a biased case history,[25,26] which was expected to influence the participants' diagnostic decisions. One could argue that physicians would reasonably tend to favor the suggested diagnosis after having received the biasing story with information that supported that diagnostic hypothesis. However, our study shows that participants are biased towards accepting suggested diagnoses even when no such history is present. In other words, the cases were exactly the same, only the accuracy of the diagnostic suggestion differed, and this had an impact on diagnostic decisions. In particular, this study showed that approximately 53% of the incorrect suggested diagnoses were accepted, that means, were considered to be correct. The high percentage of accepted incorrect diagnoses is an important finding since the process of establishing a correct diagnosis is significant for every clinician and is core-business for internists. If even fairly experienced doctors are influenced by a suggested diagnosis in such way that they have difficulties to refute a wrong diagnosis, their judgments may be susceptible to confirmation bias. A somewhat more complex finding concerns the rejection of 26% of the correct suggestions. One could argue that when "all the ducks are aligned," that is, when all described signs and symptoms would support the suggested diagnosis, these rejections would not have happened. However, to avoid ceiling effects, the selected cases posed some challenges for the participants since there must be room for mistakes. However, it is important to note that the obtained performance on those cases is similar to that of previous studies.[25,26]

Study limitations

One question that is more difficult to answer is why physicians' diagnostic decisions are influenced by diagnostic suggestions. Research in social psychology has shown that people generally need less hypothesis-consistent information to accept a hypothesis than hypothesis inconsistent-information to reject an incorrect hypothesis.[24] In other words, people tend to accept their own and other's hypotheses easier than they reject them. In medicine, this implies that when a diagnosis has been suggested, case features may be interpreted in light of that diagnosis, or perhaps, the features that are in line with the diagnosis tend to receive more consideration. Future research should try to explore how this tendency to confirm incorrect diagnoses occurs. Neutral features within a clinical case, that is, features that can be attributed to both the correct and the incorrect diagnosis, might play an important role here, as they may be taken as evidence of the correctness of the hypothesis and restrain the process of creating a differential diagnosis.

In medicine it is common practice that a colleague offers a diagnostic hypothesis with little further explanation. For example, a nurse in the emergency room may introduce a patient with a history of myelodysplasia and associated fatigue as a probable case of worsening of

the myelodysplasia, because she saw stable leukocytosis in the laboratory results. This suggestion is in itself not necessarily damaging, but might become so when the emergency room doctor (often a resident, as in this study), focuses on confirming the proposed diagnosis, and as a consequence misses alternative diagnoses (e.g., sepsis). One might argue that in real-life, the tendency of doctors to go along the garden path will be smaller than in our experimental setting, in other words, that doctors will become more reflective or critical when the stakes are high. However, research has suggested that in everyday decision making, people tend to reuse strategies that have been successful in the past.[1,30,31] In the example of the myelodysplasia case, if the emergency room nurse has a reputation of being correct most of the time, the resident will be less likely to doubt the proposed diagnosis because of this reputation and the pressure to quickly act in this context.

Besides speculating on the causal mechanisms and the potential of diagnostic suggestions to influence everyday clinical decision making, it could be argued that the implications of this study are limited, due to the small-sized sample. However, most studies on the development of medical expertise, medical and non-medical alike, encounter the problem that expertise is scarce, and hence it is difficult to find and involve many participants. On the other hand, as in the present study, effects are often large, and a small number of participants is often sufficient to identify significant differences. With an effect size of 0.85 on the diagnostic decisions results it seems that the study was sufficiently poweredFuture studies could concentrate on the effect of diagnostic suggestions on diagnostic conclusions in even more experienced physicians. This issue is important, especially given the debate on the relationship between experience, age and the vulnerability for cognitive errors and performance in medicine.[32,33] Some authors argue that experience does not always go hand in hand with better performance. For example, Eva and Cunnington[34] have shown that older doctors tend to overweigh information that is presented early in the case. Whether the potential of incorrect suggested diagnoses to cause diagnostic errors stretches to older internists, who are known to rely heavily on non-analytical reasoning[17] might be interesting for future research. In addition, it might be appealing to investigate how physicians can overcome the potentially negative influence of a suggested diagnosis. For instance, research on reflective practice has shown that actively reflecting upon a problem is a promising approach to better deal with complex cases[35] and can also act as a counteracting strategy for bias.[28,36]

The findings from our study are also significant from an educational perspective. Simply telling students and residents to be aware or to ignore diagnostic suggestions is not likely to work.[25] Besides, diagnostic suggestions are probably largely correct most of the time. It would therefore be inefficient and even unreasonable to ignore them. A more promising approach is to create a default mode in which students are taught always to consider and pursue plausible alternative diagnoses. In this way, chances of making a correct diagnosis increases. As has been shown in recent studies, such a reflective approach not only had a positive effect on diagnostic performance in complex clinical cases,[35] but it also helped clinicians to overcome errors due to

bias.[28] Further exploration of the reflection paradigm, that is the translation of this paradigm into concrete instructional approaches, should be the next step to increase doctors' defenses against diagnostic errors.[15]

In conclusion, this study suggests that physicians are influenced by a suggested diagnosis in such way that they have difficulties in rejecting a wrong diagnosis, which may open the door for confirmation bias and, consequently, for diagnostic errors.

REFERENCES

1. Nickerson RS. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 1998;2:175-220.
2. Kohn KT, Corrigan JM, Donaldson MS. Editors. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 1999.
3. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three different eras. *New England Journal of Medicine* 1983;308:1000-5.
4. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy detected diagnostic errors over time. *Journal of the American Medical Association* 2003;289:2849-56.
5. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study 1. *New England Journal of Medicine* 1991;324:370–6.
6. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Medical Care. 2000;38:261-71.
7. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Medical Journal of Australia* 1995;163:458–71.
8. Graber ML, Franklin N,. Gordon R. Diagnostic error in internal medicine. *Archives of Internal Medicine* 2005;165:1493-9.
9. Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: What's the goal? *Academic Medicine* 2002;77:981-92.
10. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *British Medical Journal* 2002;324:729-732.
11. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine* 2003;78:775-80.
12. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *British Medical Journal* 2005;330:781-4.
13. Redelmeier DA. The cognitive psychology of missed diagnoses. *Annals of Internal Medicine* 2005;142:115-20.
14. Singh H, Thomas EJ, Khan MM, & Peterson LA. Identifying diagnostic errors in primary care using an electronic screening algorithm. *Archives of Internal Medicine* 2007;167:302-8.
15. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Medical Education* 2010;44:94-100.
16. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise - Theory and implications. *Academic Medicine* 1990;65:611-21.
17. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educational Psychology Review* 1993;5:205-21.
18. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Advances in Health Sciences Education* 1997;2:173-84.
19. Wason PC. On the failure to eliminate hypotheses in a cognitive task. *Quarterly Journal of Experimental Psychology* 1960;12:129-40.
20. Evans JSBT. Bias in human reasoning: causes and consequences. NJ, Erlbaum; 1989.
21. Kuhn D, Weinstock M, Flaton R. How well do jurors reason? Competence dimensions of individual variations in a juror reasoning task. *Psychological Science* 1994;5:289-96.
22. Dror IE, Charlton D, Péron AE. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International* 2006;156:74-8.

23. Gilovich T. Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology* 1983;44:1110-26.
24. Pyszczynski T, Greenberg J. Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis testing model. *Advances in Experimental Social Psychology*. New York: Academic Press; 1987:297-340.
25. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Academic Medicine*. 2001;76:S18-20.
26. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Academic Medicine*. 2002;77:S67-9.
27. McLaughlin K, Heemskerk L, Herman R, Ainslie M, Rikers RM, Schmidt HG. Initial diagnostic hypotheses bias analytic information processing in non-visual domains. *Medical Education* 2008;42:496-502.
28. Mamede S, van Gog T, van den Berge K, Rikers RMJP, van Saase JLCM, van Guldener C et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Journal of the American Medical Association* 2010;304:1198-1203.
29. Cohen J. Statistical power analysis for the behavioural sciences (2nd Ed.). Hillsdale, NJ: Erlbaum; 1988.
30. Betsch T, Haberstroh S, Hohle C. Explaining routinized decision making: A review of theories and models. *Theory & Psychology* 2002;12:453–88.
31. Betsch T, Haberstroh S, Glockner A, Haar T, Fiedler K. The effects of routine strength on adaptation and information search in recurrent decision making. *Organizational Behavior and Human Decision Processes* 2001;84:23–53.
32. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Annals of Internal Medicine* 2005;142:260-73.
33. Norman GR, Eva KW. Clinical experience and the quality of healthcare, comments and responses. *Annals of Internal Medicine* 2005;143:85-6.
34. Eva KW, Cunnington JPW. The difficulty with experience: Does practice increase susceptibility to premature closure? *Journal of Continuing Education in the Health Professions* 2006;26:192–8.
35. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education* 2008;42:468-75.
36. Schmidt HG, Mamede S, Van den Berge K. van Gog T. Van Saase JLCM, Rikers RMJP. (submitted for publication). Exposure to media information causes availability bias in medical diagnosis.

Chapter 3

Supportive Features Mediate the Tendency to Accept Diagnostic Suggestions[1]

[1] This Chapter is submitted as: Van den Berge K, Mamede S, Van Gog T, De Graaf J, Van Saase JLCM, Rikers RMJP. Supportive Features Mediate the Tendency to Accept Diagnostic Suggestions.

ABSTRACT

Background: A substantial portion of diagnostic errors can be attributed to faults in physician's cognitive processes. Research on medical expertise suggests that physician's reasoning may be susceptible to bias, such as confirmation bias (i.e., the tendency to seek information to support rather than refute a hypothesis).

Objective: This study attempts to explain confirmatory tendencies in medical diagnosis. It is hypothesized that residents tend to focus on and therefore report more clinical features that support a diagnosis that has been suggested to them, while ignoring features that speak against the suggested diagnosis.

Method: Thirty-eight residents evaluated suggested diagnoses on four written cases. Two of those suggestions were correct and two were incorrect. After evaluating each suggestion, participants reported the features they would mention to a supervisor while discussing the patient.

Results: Participants had more trouble rejecting incorrect suggestions (evaluation score on cases with incorrect suggestions $M = 0.63$, $SD = 0.59$) than accepting correct suggestions (evaluation score on cases with correct suggestions $M = 1.13$, $SD = 0.70$), $t(37) = 2.84$, $p = 0.007$, $d = 0.77$. Irrespective of the correctness of the suggested diagnosis, a higher percentage of features supporting the suggestions ($M = 39.05$, $SD = 12.70$) was reported than features supporting the alternatives ($M = 25.25$, $SD = 13.33$), $t(37) = 4.40$, $p < 0.001$, $d = 1.06$. Furthermore, when incorrect suggestions were accepted, more features supportive of these suggestions were reported than when they were rejected, $t(49) = 1.88$, $p = 0.03$, $d = 0.53$, indicating erroneous decisions may have been caused by the tendency to gather data in line with the suggestion.

Conclusions: These findings demonstrate that the tendency to confirm suggested diagnoses is mediated by a focus on features that support these suggestions. This tendency may lead to diagnostic errors when the suggestion is incorrect.

INTRODUCTION

A recent study demonstrates that residents tend to accept suggested diagnoses on written clinical cases irrespective of their correctness and that such a tendency may lead to diagnostic errors.[1] It is, however, unclear why physicians exhibit confirmatory tendencies while evaluating diagnostic suggestions on written cases. One explanation might be that a diagnostic suggestion leads physicians to focus on clinical case information that favors the suggestion, while ignoring evidence that speaks against it. The present study investigates whether clinicians indeed tend to focus on information that supports diagnostic suggestions.

A landmark study showed that medical errors are important causes of death in the United States.[2] Amid medical errors, treatment errors are considered the main culprit, but the rate of diagnostic errors is also substantial,[3,4] and is highest in the domains of internal medicine and emergency medicine.[5-7] A sizeable portion of diagnostic errors can be attributed to faults in physicians' cognitive processes.[8] Such cognitive diagnostic errors may arise from inadequate knowledge, faulty data gathering, inaccurate clinical reasoning, and erroneous verification of diagnostic hypotheses.[9] The potential of cognitive factors to cause error in medical diagnosis has been stressed in several position papers,[10-12] and observational studies have provided indirect evidence that faults in physicians' cognitive processes, most importantly faulty reasoning, may in fact be involved in the majority of diagnostic errors.[8,13]

Research on medical expertise suggests that cognitive bias might be the mechanism that causes faulty reasoning.[14-17] This research has shown that physicians automatically and largely unconsciously generate diagnostic hypotheses early in the diagnostic process, mainly through pattern recognition. This non-analytical and automatic mode of reasoning costs, therefore, little effort, and is the dominant reasoning mode when clinicians deal with routine problems.[17] As it occurs largely without conscious control, generation of hypotheses based on pattern-recognition could be influenced by multiple factors that remain unnoticed, making physicians more vulnerable to bias and consequently to errors.[11,12]

A main goal for clinical teachers is the prevention of mistakes in clinical reasoning of their students. Therefore, they often recommend students to gather all information before making a diagnosis. This includes meticulous listing of the patient's presenting features, to reduce the chances of an incorrect diagnosis due to premature closure.[10] However, the detection of clinical features has shown to be no clear-cut process, since some authors believe clinical features to be anything but independent cues that are processed and used as pieces of evidence in order to arrive at a diagnosis.[18-20] Instead, they believe features are extracted and explained in light of the diagnoses that the physician has in mind.

For instance, research that used radiographs to study clinical decision-making, has shown that tentative diagnoses can increase diagnostic accuracy by directing attention to features that might otherwise be missed.[21] Another study using ECGs showed that a biasing history influenced both diagnostic accuracy and the identification of features.[22] In this study, a

suggestive history prior to evaluating the ECGs, guided internal medicine residents towards the diagnosis suggested by the history. Additionally, when a history supportive of the correct diagnosis was suggested, residents mentioned more features compatible with this diagnosis and fewer features compatible with plausible alternative diagnoses. In contrast, when a history supportive of a plausible alternative diagnosis was given, residents were more likely to leave out features compatible with the correct diagnosis and call more features that were compatible with both the correct and alternative diagnoses. Similar results have been obtained in studies using photographs of patients.[19,20] An intriguing observation from one study[20] concerned the reinterpretation of features in the light of a suggested diagnosis. For example, in one scenario, participants misinterpreted tanned skin as jaundice more often when biased toward liver cancer than when biased toward stomach cancer, despite the patient's white sclerae.

This tendency to identify and interpret clinical features in light of a suggested diagnosis may lead to diagnostic mistakes through susceptibility for confirmation bias. This bias, defined as peoples' inclination to seek and interpret data to confirm, rather than refute a hypothesis,[23] is believed to cause diagnostic errors in medicine.[10-12] As reported above, experimental evidence on this subject in the medical domain is scarce and consists mainly of studies that used visual materials.[18,19,20,22] It is not known whether this also applies to non-visual materials, such as cases often encountered in internal medicine; a domain that is thought to be prone to diagnostic errors.[5-7] An important difference between visual materials and written cases, is that in written cases, features are explicitly mentioned in the case descriptions, so they are less likely to be missed. Moreover, visual materials may introduce irrelevant features that would probably not be described in a written case (e.g., a tanned skin could be misinterpreted as jaundice in a picture when that would be in line with a diagnostic suggestion, but it would probably not be mentioned in a written case). In addition, in the studies with visual materials the evaluation of diagnostic suggestions was preceded by a biased (i.e., corroborative of the diagnostic suggestion) case history.[19,20,22] In other words, these cases were developed with the aim to create a lure for the participants. So, in those studies not only the accuracy of the suggested diagnoses differed but also the cases themselves were different. In fact, it is likely that the corroborative histories that accompanied the suggested diagnoses in those studies may have been very persuasive in drawing participants' attention to features that supported the suggestions.

In a recent study,[1] similar results with regard to diagnostic conclusions were shown with written cases. In contrast to previous studies, this study did not use corroborative case histories but asked physicians to evaluate a suggested diagnosis that preceded the written clinical cases, all of which were based on real patients and had a confirmed diagnosis. Although this study showed that confirmatory tendencies exist, also when evaluating suggested diagnoses with written cases, it remains unclear *how* suggested diagnoses evoke confirmatory tendencies while evaluating them on written clinical cases. Even though -as mentioned above- written cases are fundamentally different from visual stimuli, a similar mechanism may be at work while

evaluating diagnostic suggestions on written cases: the suggestion may result in a search that extracts features in favor of the suggested diagnosis.

This study aims to investigate the hypothesis that physicians indeed focus on features that are in line with a suggested diagnosis, while evaluating that suggestion for written clinical cases, leading them to mainly report supportive features, rather than features that would support alternative diagnoses.

METHOD

Participants

Thirty-eight internal medicine residents (28 female, age: *M*=30.00, *SD*=2.63 years) from the Radboud University Nijmegen Medical Centre, Faculty of Medicine, volunteered to participate in this study. The ethics review committee from the Department of Psychology, Erasmus University Rotterdam, provided approval for this study. Participants' written consent was obtained and they were debriefed when the study was completed.

Materials

Cases: A set of 4 written clinical cases was used for this study (see Appendix for an example). All cases were based on real patients and had a confirmed diagnosis. They were designed and validated independently by two experts in internal medicine and had been previously used in studies with internal medicine residents.[24,25] The cases were presented to the participants in a booklet, showing one case per page. Each case started with a diagnostic suggestion, which was to be evaluated by the participants. Participants evaluated the suggested diagnosis by indicating, directly after each case description, which was presented immediately after the suggested diagnosis, whether they agreed or disagreed with the suggested diagnosis. On the next page, participants were asked to write down the features they would report to their supervisor when they would discuss the patient for further management. The order in which the cases were presented in the booklet was randomized, and whether a correct or incorrect suggested diagnosis preceded a case was also randomized (see Table for correct and incorrect diagnoses). In other words, all participants evaluated the same four cases, two with a correct suggested diagnosis and two with an incorrect suggested diagnosis; both the suggested diagnoses and the order of presentation of the cases differed between participants.

Evaluation of experience: Participants were asked to indicate their level of experience with all eight possible suggested diagnoses (i.e., they rated both the suggested diagnoses they actually saw, as well as the alternatives) on a seven point Likert-scale, ranging from (1) "no experience with the disease", to (7) "high experience with this disease".

Procedure

The study was conducted during an educational session. The booklets were randomly distributed among the participants. The instruction for evaluation of the cases was provided in the booklet: "Read the following cases quickly but carefully and indicate whether you agree or disagree with the suggested diagnosis". Participants were given 75 seconds to evaluate each diagnosis. A previous study had shown this time to be sufficient for reading all case information and making a diagnostic decision.[1] After 75 seconds the experiment leader instructed the participants to move on to the next page, and write down the features they would report as if they were discussing the patient with their supervisor. They were not allowed to check the case descriptions while writing down the features. After 45 seconds, participants were instructed to move to the next case, until all four cases were completed. Next, they were asked to report their level of experience with all possible suggested diagnoses. Finally, participants filled out questions regarding demographic data (gender and age). The whole experiment took about 10 minutes.

Data Analysis

Diagnostic decisions: For each correct decision (i.e., agreeing with the correct suggested diagnosis or disagreeing with the incorrect suggested diagnosis) a score of 1 point was assigned. So, a maximum of 4 correct diagnostic evaluations could be made (i.e., 2 rejections of incorrect diagnostic suggestions and 2 acceptations of correct diagnostic suggestions). Paired samples *t*-tests were used to compare the score on cases accompanied by a correct suggested diagnosis with the score on cases accompanied by an incorrect suggested diagnosis.

Experience: Average experience ratings with the 4 correct and the 4 incorrect suggested diagnoses were calculated, resulting in experience scores ranging from 1 (i.e., no experience) to 7 (i.e., high experience) for both correct and incorrect suggested diagnoses. Paired samples *t*-tests were used to compare participants' self-reported experience with the diseases presented as correct diagnostic suggestions to those presented as incorrect diagnostic suggestions.

Case features: The features participants wrote down were scored by assigning each reported feature to one of three categories: (1) features supportive of the correct suggested diagnosis, (2) features supportive of the incorrect suggested diagnosis, and (3) neutral features that could not be exclusively assigned to either the correct or incorrect diagnosis. All features present in the clinical cases were analyzed and labeled independently by two experts in internal medicine prior to the study. The inter-rater agreement for the labeling of features was 99%. Next, the number of reported features supportive of either correct or incorrect suggestions were converted to percentages of existing features in a case, in order to correct for differences between cases in terms of features. That is, for some of the cases, the number of features supportive of the correct suggested diagnosis was not equal to the number of features supporting the incorrect suggested diagnosis (see Table 1 for number of features per case).

Table 1: Correct (cases) and incorrect diagnostic suggestions used in the study

Correct diagnosis (number of supportive features in case)	Incorrect diagnosis (number of supportive features in case)
Aortic dissection (3)	Myocardial infarction (3)
Neurosyphilis (6)	Wernicke's encephalopathy (3)
Primary sclerosing cholangitis (5)	Viral hepatitis (3)
Acute bacterial endocarditis (4)	Tuberculosis (4)

To check whether a suggested diagnosis influenced the reporting of features, the percentage of features supportive of suggested diagnoses was compared to the percentage of features supportive of alternative diagnoses, which participants identified when they received the suggestions. In addition, to check whether erroneous decisions may have been caused by the tendency to gather data in line with the suggestion, it was analyzed whether participants reported a higher percentage of features supportive of the incorrect diagnostic suggestions when they accepted them than when they rejected them.

All reported features were assigned the same weight (i.e., 1 point for each mentioned feature) and comparisons were analyzed by paired samples *t*-tests. For all analyses, a significance level of 0.05 is used, and *d* is reported as a measure of effect size, with values of 0.20, 0.50, and 0.80 corresponding to small, medium, and large effect sizes, respectively.[26]

RESULTS

Diagnostic evaluation scores and experience on diagnoses presented as suggested diagnoses

The mean evaluation score (range 0-2) on cases with a correct suggestion was significantly higher (M=1.13, SD=0.70) than on cases with an incorrect suggested diagnosis (M=0.63, SD=0.59). This indicates that participants were generally inclined towards accepting the suggestions even when they were incorrect, $t(37)=2.84$, $p=0.007$, $d=0.77$.

Reported experience (range: 1-7) with diagnoses presented as correct suggestions (M=3.11, SD=0.97) proved to be significantly lower than experience with diagnoses presented as incorrect suggestions (M=3.77, SD=0.87), $t(34)=5.12$, $p<0.001$, $d=0.72$.

Reporting of features

A significantly higher percentage of features supporting the suggested diagnoses (M=39.05, SD=12.70) was reported than features supportive of the alternative diagnosis (M=25.25, SD=13.33), $t(37)=4.40$, $p<0.001$, $d=1.06$. This reporting was irrespective of the participants' diagnostic decisions. When participants accepted incorrect suggestions, the percentage of reported features supporting these incorrect suggestions (M=50.10, SD=23.52) was higher than when participants rejected those suggestions (M=36.85, SD=26.16), meaning their erroneous

decisions may have been caused by the tendency to gather data in line with the suggestion, $t(49)=1.88$, $p=0.03$, $d=0.53$.

DISCUSSION

The present study has tried to explain physicians' confirmatory tendencies while evaluating diagnostic suggestions using written cases.[1] It was hypothesized that the diagnostic suggestions direct physicians' attention to the features that support the suggestion. Indeed, our findings indicate that participants reported more features supportive of the diagnostic suggestions than those that were not in line with the suggestion. This indicates that a diagnostic suggestion results in a tendency to identify features that are in favor of the suggested diagnosis, which may explain physicians' inclination towards accepting the suggestion.

Moreover, when participants made an error (i.e., decided to accept an incorrect or reject a correct suggestion), their faulty decisions seem to have been caused by the tendency to gather data in line with the suggestion. That is, in case of an incorrect suggestion, more features supportive of the incorrect suggestion were reported when participants accepted this suggestion, than when they rejected it. This preference for information that corroborates physicians' beliefs, above information that falsifies them, may prevent a critical evaluation of diagnoses, possibly leading to diagnostic errors caused by confirmation bias.[10-12] This conclusion is consistent with previous studies on the relationship between diagnostic decisions and feature identification, where suggested diagnoses also directed diagnostic decisions and the reporting of features towards the suggestions.[19,20,22]

However, our study differs in several aspects from previous studies. For instance, it did not rely on visual materials but used clinical case descriptions. Consequently, the features were explicitly mentioned in the case descriptions and therefore they were less likely to be missed by participants than features depicted in a picture or an ECG. Furthermore, visual images are probably more powerful in opening space for noticing features that would probably not be expressed in a written case description. For instance, a tanned skin shown in a patient's picture could be misinterpreted as jaundice if a wrong diagnosis of liver carcinoma is suggested. But in a written case, the "tanned skin" would not be mentioned, rather either the sclerae would be described or nothing would be said. In addition, the motivation for reporting features also differed from previous studies.[19,20] In those studies, participants wrote down all relevant features. In the present study, participants listed features they would report to their supervisor while discussing the patient prior to management. This approach to the reporting of features mimics professional interaction in everyday clinical decision making, and illustrates the potential consequences of decisions on a patient-management level. Namely, if the supervisor accepts the resident's erroneous diagnosis because of all the reported evidence supporting it, the error could jeopardize patients' health.

A finding that is more difficult to clarify is the participants' relatively low score while evaluating correct suggestions. Only 57% of these suggestions were accepted and this number is lower than in similar studies.[1,19,20] Possibly, the use of less straightforward cases in the present study, which was necessary to allow some space for error, can explain this finding. Additionally, participants could have thought there must have been a catch in most of the cases, leading them to reject some correct suggestions. Finally, participants lower experience with diagnoses presented as correct suggestions may partly account for the rejection of correct suggestions.

Based on this study, we cannot exclude that these results would be less dramatic or even disappear if participants would have been more experienced. On the other hand, studies with visual materials showed that even experienced physicians are influenced by diagnostic suggestions,[20,21] and a recent study has provided evidence that more experience may actually open the door for biased decisions even further.[25] Future studies should therefore focus on the relationship between experience and cognitive diagnostic errors. That is, if experience does not protect patients from their physicians' cognitive errors, counteracting strategies have to be designed and taught to medical professionals.

Which brings us to the main didactical question: how to deal with diagnostic suggestions in everyday practice? In this experiment, the suggestions were intentionally incorrect half of the time. In practice however, they will often be correct, and may facilitate diagnosis.[21] On the other hand, accepting incorrect diagnoses due to a search for supportive features may also lead to diagnostic errors,[10,11] and is a process that has to be restrained.[10] A remedy for this type of error may be instructing residents to try to seriously consider alternative hypotheses as much as possible. In a recent study, reflection has shown to be a promising approach to deal with complex cases,[27] and an effective way to counteract diagnostic errors caused by bias.[25,28] Therefore, future studies in clinical reasoning could investigate what will happen with diagnostic conclusions and characteristics of reported features if physicians are asked to reflect upon cases with a suggested diagnosis. In line with this recommendation, supervising specialists should be aware that in professional communication there is always a risk of confirmatory tendencies. Therefore, while discussing patients, both supervisors and residents must remain alert and reflective. Moreover, supervisors should require residents to think of alternative diagnoses in order to avoid diagnostic errors, and such discussions could also contribute to a more efficient educational environment.

In conclusion, this study indicates that the tendency to confirm a suggested diagnosis is mediated by a focus on features that support it. This inclination to concentrate on data to confirm a diagnosis may lead to diagnostic errors. Counteracting such problems in clinical practice remains a challenge, but promising strategies are emerging.

REFERENCES

1. Van den Berge K, Mamede S, Van Gog T, et al. Accepting Diagnostic Suggestions by Residents: A Potential Cause of Diagnostic Error in Medicine.Teach Learn Med. In press 2012.
2. Kohn KT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
3. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three different eras. New Engl J Med. 1983;308:1000-5.
4. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy detected diagnostic errors over time. JAMA. 2003;289:2849-56.
5. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study 1. New Engl J Med. 1991;324:370–6.
6. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care. 2000;38:261-71.
7. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. Med J Australia. 1995;163:458–71.
8. Graber ML, Franklin N,. Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165:1493-9.
9. Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: What's the goal? Acad Med. 2002;77:981-92.
10. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78:775-80.
11. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. Brit Med J. 2005;330:781-4.
12. Redelmeier DA. The cognitive psychology of missed diagnoses. Ann Intern Med. 2005;142:115-20.
13. Singh H, Thomas EJ, Khan MM, Peterson LA. Identifying diagnostic errors in primary care using an electronic screening algorithm. Arch Intern Med. 2007;167:302-8.
14. Norman GR, Eva KW. Diagnostic error and clinical reasoning. Med Educ. 2010;44:94-100.
15. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise - Theory and implications. Acad Med. 1990;65:611-21.
16. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. Educ Psychol Rev. 1993;5:205-21.
17. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. Adv Health Sci Educ. 1997;2:173-84.
18. Norman GR, Brooks LR, Coblentz CK, Babcook CJ. The correlation of feature identification and category judgments in diagnostic radiology. Mem Cognition. 1992;4:344–355.
19. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. Acad Med. 2001;76:S18-20.
20. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. Acad Med. 2002;77:S67-9.
21. Berbaum KS, Franken EA, Dorfman DD, et al. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Invest Radiol. 1986;21:532–539.
22. Hatala RM, Norman GR, Brooks LR. The impact of a clinical scenario upon electrocardiogram interpretation. J Gen Intern Med. 1999;14:126–129.
23. Nickerson RS. Confirmation bias: A ubiquitous phenomenon in many guises. Rev Gen Psychol. 1998;2:175-220.
24. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM, Influence of perceived difficulty of cases on physicians' diagnostic reasoning. Acad Med. 2008;12:1210-1216.

25. Mamede S, van Gog T, van den Berge K, et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. JAMA. 2010;304:1198-1203.
26. Cohen J. Statistical power analysis for the behavioural sciences (2nd Ed.). Hillsdale, NJ: Erlbaum; 1988.
27. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ. 2008;42:468-75.
28. Schmidt HG, Mamede S, Van den Berge K. van Gog T. Van Saase JLCM, Rikers RMJP. Exposure to media information causes availability bias in medical diagnosis. Submitted.

Chapter 4

Physicians' Accepting of Diagnostic Suggestions Occurs Independent of Consistency in Preceding Suggestions[1]

[1] This Chapter is submitted as: Van den Berge K, Mamede S, Van Gog T, C, Van Saase JLCM, Rikers RMJP. Physicians' accepting of diagnostic suggestions occurs independent of consistency in preceding suggestions.

ABSTRACT

Background: Studies in clinical reasoning suggest that physicians tend to accept diagnostic suggestions, which could lead to diagnostic errors if the suggestion happens to be incorrect. Those studies did not take into account that physicians in clinical practice will mainly encounter correct suggestions before they are confronted with an incorrect suggestion. The present study investigated physicians' diagnostic performance if they would first encounter a number of correct suggestions followed by a number of incorrect suggestions, and vice versa. It was hypothesized that more incorrect suggestions would be accepted if participants had first evaluated a series of correct suggestions. On the other hand, it was hypothesized that, if participants had first evaluated incorrect suggestions, they would become more critical about the suggestions, and hence will become more inclined to reject correct suggestions.

Method: Internal medicine residents ($N = 38$) evaluated suggested diagnoses on 8 written clinical cases. Four of those suggested diagnoses were correct, and 4 were incorrect. Half of the participants first evaluated 4 correct suggestions and then evaluated 4 incorrect suggestions (C/I-condition). The other half started with the 4 incorrect suggestions followed by the correct suggestions (I/C-condition).

Results: Our findings show that there was no significant main effect of case order and phase. Evaluation score in the C/I condition ($M = 2.87$, $MSE = 0.14$) equaled that in the I/C condition ($M = 2.66$, $MSE = 0.14$), $F(1,36) = 1.09$, $p = 0.30$, *ns*. Evaluation score on the first 4 cases ($M = 2.68$, $MSE = 0.14$) equaled that score on the last 4 cases ($M = 2.84$, $MSE = 0.17$), $F(1,36) < 1$, meaning that consistency in preceding suggested diagnoses did not influence the tendency to accept subsequent diagnostic suggestions. There was, however, a significant interaction effect between condition and phase, $F(1,36) = 11.82$, $p = 0.001$, $\eta_p^2 = 0.25$, demonstrating that, in both conditions, the score on cases with correct suggestions was higher than the score on cases with incorrect suggestions.

Conclusion: These findings indicate that consistency in preceding correct or incorrect diagnostic suggestions did not influence the tendency to accept or reject subsequent suggestions. However, overall physicians still showed a tendency to accept diagnostic suggestions, which may lead to diagnostic errors if the suggestion is incorrect.

INTRODUCTION

Several studies suggested that physicians tend to accept diagnostic suggestions for clinical cases.[1-3] Such an inclination towards confirmation might, if the suggestion is incorrect, lead to diagnostic errors. Those studies, however, did not take into account that in everyday clinical practice, physicians are likely to encounter many correct suggestions before they are confronted with an incorrect suggestion. This study investigates whether the inclination towards accepting diagnostic suggestions is influenced by consistency in correctness of preceding diagnostic suggestions.

The publication of the Institute of Medicine report led to an increase in research on medical errors, and stimulated discussion on patient safety issues.[4-6] The report showed that medical mistakes, the majority of which are related to treatment, cause many preventable deaths in the United States. Besides treatment-errors, diagnostic mistakes account for a substantial portion of medical errors. The rate of such diagnostic errors lies within the 10-15% range,[7,8] and the clinical specialties of internal medicine and emergency medicine are believed to be most affected by them.[9-11]

Diagnostic errors have many causes but a substantial number of mistakes seem to stem from faults in physicians' cognitive processes.[12] These so-called 'cognitive diagnostic errors' may occur due to insufficient knowledge, but other factors, such as faulty gathering or interpretation of clinical data, and flawed verification of diagnostic hypotheses, have been pointed out as the main culprits.[12] Several authors have discussed the potential of cognitive factors to cause diagnostic errors,[13-15] and observational studies suggest that clinicians' thinking errors may actually be involved in the majority of missed or delayed diagnoses.[12,16]

The discussion about the causes of cognitive diagnostic errors is ongoing,[17] and may benefit from medical expertise research. This research suggests that diagnostic reasoning may be vulnerable to bias.[18-20] For example, it has been demonstrated that physicians generate hypotheses in the beginning of patient contact, mainly through pattern recognition: similarities between the current and previously seen patients quickly bring one or a couple of diagnostic hypotheses to the physician's mind, which are used to guide the search for additional evidence. This mainly automatic, non-analytical mode of reasoning occurs relatively effortless, and is the chief mode of reasoning when clinicians deal with routine problems.[20] It is usually effective but, as it occurs largely without conscious control, generation of hypotheses based on pattern-recognition may be influenced by multiple factors that remain unnoticed, making physicians more prone to bias and, consequently, to errors.[14,15]

Admitted patients often come with the diagnostic considerations of another medical professional (e.g., the general practitioner, a nurse, or the ambulance personnel). If physicians would tend towards accepting such suggestions, correct suggestions could facilitate fast and accurate diagnosis. However, even though such suggestions will often be correct, they may sometimes be wrong, and in that case, accepting diagnostic suggestions may lead to errors.[21,22] A recent

study showed that physicians indeed tend to accept diagnostic suggestions for written clinical cases.[1] In that study, residents in internal medicine evaluated diagnostic suggestions for subsequently presented case-descriptions, which were all based on real patients and had a verified diagnosis. Half of the diagnostic suggestions were correct, and half of them were incorrect. Results showed that participants found it harder to reject an incorrect suggested diagnosis than to accept a correct suggested diagnosis. However, in that study, the correct and incorrect suggestions alternated, which is unlikely to happen in everyday clinical practice. That is, in clinical practice, the correctness of diagnostic suggestions is unlikely to alternate that often.

Based on research in medical expertise, it can be assumed that perceiving a consistent series of diagnostic suggestions might influence diagnostic decision making on subsequent cases. For instance, seeing a consistent series of correct diagnostic suggestions might lead to the expectancy that a next suggestion is also likely to be correct, and hence increases the chances that it is accepted even when incorrect.[23-25] On the other hand, it is known that when physicians encounter inconsistencies or complexity in cases, they may return to a more deliberate mode of diagnostic reasoning.[23,25] Accordingly, it can be hypothesized that, when inconsistencies between the suggested diagnosis and the findings in a case are noticed, this is likely to evoke a more critical approach towards such suggestions.

In the present study, it is hypothesized that more incorrect suggestions would be accepted if participants have first evaluated a number of correct suggestions than when no correct suggestions were evaluated prior to evaluating incorrect suggestions. Conversely, it is hypothesized that, if participants had first evaluated incorrect suggestions, they would become more critical about the suggestions, and hence will become more inclined to reject correct suggestions than when no incorrect suggestions were evaluated prior to evaluating correct suggestions.

METHOD

Participants

Thirty-eight internal medicine residents (mean age = 30.00, *SD* = 3.06 years; 23 women) from a university hospital in the Netherlands voluntarily participated in this study. The ethics review committee from the Department of Psychology, Erasmus University Rotterdam, approved this study. Participants were debriefed after the study.

Materials

A set of eight written clinical cases, which were based on real patients and had confirmed diagnoses, was used in the study (see Appendix for an example). They were designed and validated independently by two experts in internal medicine and had been previously used in studies with internal medicine residents.[1,26] The cases were presented to the participants in a booklet, showing one case per page. A diagnostic suggestion, that had to be evaluated

by the participants by indicating, directly after each case description whether they agreed or disagreed with the suggestion, preceded each case description.

The cases consisted of two series of four cases. One series with four correct suggested diagnoses and the other four with incorrect suggested diagnoses. Within each series, the cases were presented in a fixed order. There were two versions of the booklet; the cases were the same in both versions but the two series of cases were presented in a different order: half of the participants evaluated four correct suggested diagnoses, followed by four incorrect suggested diagnoses (C/I-condition). The other half first evaluated the four incorrect suggested diagnoses, and then evaluated the four correct suggestions (I/C-condition, see Table 1 for an overview of the materials).

Table 1: Suggested diagnoses and correct diagnoses for the cases used in the study

C/I-condition		I/C-condition	
Suggested diagnosis	**Correct diagnosis**	**Suggested diagnosis**	**Correct diagnosis**
Aortic dissection	Aortic dissection	Q fever	Viral infection
Inflammatory bowel disease	Inflammatory bowel disease	Legionnaire's disease	Pneumococcal pneumonia
Neurosyphilis	Neurosyphilis	Ulcerative colitis flare-up	Clostridium colitis
Primary sclerosing cholangitis	Primary sclerosing cholangitis	Liver metastasis	Liver cirrhosis
Q fever	Viral infection	Aortic dissection	Aortic dissection
Legionaire's disease	Pneumococcal pneumonia	Inflammatory bowel disease	Inflammatory bowel disease
Ulcerative colitis flare-up	Clostridium colitis	Neurosyphilis	Neurosyphilis
Liver metastasis	Liver cirrhosis	Primary sclerosing cholangitis	Primary sclerosing cholangitis

After evaluating suggested diagnoses, participants evaluated their experience with the diagnoses that were presented as suggested diagnoses on a seven point Likert-scale, ranging from (1) "no experience with the disease", to (7) "highly experienced on the disease". A short demographic questionnaire concerning gender and age completed the materials.

In order to ensure that cases with a correct suggestion did not differ in complexity from cases with an incorrect suggestion, a pilot study was conducted using the same cases. In this pilot, the cases were randomly presented to 15 participants in a booklet, showing one case per page. The participants were asked to read the case quickly but carefully and write, directly after each case text, their diagnosis. They were allowed 75 seconds to diagnose each case. For each correct diagnosis, a score of 1 point was assigned. When the diagnosis was incorrect, no points were given. Results showed the diagnostic performance on cases that were presented with correct suggestions in the main study ($M = 2.80$, $SD = 0.77$) did not significantly differ from

diagnostic performance on cases that were accompanied by an incorrect suggestion ($M = 3.27$, $SD = 1.03$) in the main study, $t(14) = 1.39$, $p > 0.05$.

Procedure

The study was conducted during a bimonthly educational session, which is part of the internal medicine residency training program in the Netherlands. These educational sessions that last one day, consist of lectures and discussions on a range of topics in internal medicine. Participation is voluntary and involvement of the attending residents is generally 100%. The instruction for evaluation of the cases was provided in the booklet: "Read the following cases quickly but carefully and indicate whether you agree or disagree with the diagnosis". Based on a previous study, participants were given 75 seconds to evaluate each diagnosis.[1] Time was kept by an experiment-leader who told the participants to continue to the next case after every 75 seconds. The whole procedure took about 20 minutes.

Data analysis

Mean experience ratings with the four correct and the four incorrect suggested diagnoses were calculated, resulting in experience scores ranging from 1 (i.e., no experience) to 7 (i.e., high experience) for both correct and incorrect suggested diagnoses. An independent samples *t*-test was used to compare experience with the diseases presented as (in)correct diagnostic suggestions between conditions. A paired samples *t*-test was used to compare experience with the diseases presented as (in)correct diagnostic suggestions within conditions. Participants' data on diagnostic decisions were scored as follows: for each correct evaluation (i.e., agreeing to the correct suggested diagnosis, disagreeing with the incorrect suggested diagnosis) a score of 1 point was obtained. So, a maximum score of 8 points could be obtained: 4 points for rejecting incorrect diagnoses and 4 points for accepting correct diagnoses. Data on diagnostic decisions were submitted to a mixed-design 2 x 2 analysis of variance (ANOVA) with case order (i.e., C/I or I/C) as a between subjects factor and the phase of the experiment (i.e., the first four cases compared with the last four cases) as a repeated measure.

For all analyses, a significance level of 0.05 is used. For the ANOVA, η_p^2 is reported as a measure of effect size with values of 0.01, 0.06, and 0.14, corresponding to small, medium, and large effect sizes respectively. For *t*-tests, *d* is reported as a measure of effect size with values of 0.20, 0.50, and 0.80, corresponding to small, medium, and large effect sizes, respectively.[27]

RESULTS

Participants' characteristics and experience on suggested diagnoses

Participants' characteristics were similar between conditions and they are shown in Table 2. Experience (range: 1-7) with diagnoses presented as incorrect diagnostic suggestions in the

C/I-condition equaled that experience in the I/C-condition. Experience with the diagnoses presented as correct suggestions was also similar between conditions. Within both conditions, experience with diagnoses presented as incorrect suggestions (C/I: $M = 3.90$, $SD = 0.96$; I/C: $M = 3.79$, $SD = 0.74$) exceeded experience with the diagnoses presented as correct suggested diagnoses (C/I: $M = 3.04$, $SD = 0.96$; I/C: $M = 3.07$, $SD = 0.74$) with, $t(18) = 4.43$, $p < 0.05$, $d = 0.92$, in the C/I-condition, and $t(18) = 4.43$, $p < 0.05$, $d = 0.96$ in the I/C-condition.

Table 2: Characteristics of the participants

Characteristic	C/I-condition	I/C-condition	*p*-value
Age (y)	$M = 30.26$, $SD = 3.07$	$M = 29.74$, $SD = 3.11$	.60
Average experience correct suggestions	$M = 3.04$, $SD = 0.96$	$M = 3.07$, $SD = 0.96$	.93
Average experience incorrect suggestions	$M = 3.92$, $SD = 0.96$	$M = 3.79$, $SD = 0.77$	.64

Diagnostic scores

There was no significant main effect of case order: the diagnostic evaluation score in the C/I-condition ($M = 2.87$, $MSE = 0.14$) equaled that score in the I/C-condition ($M = 2.66$, $MSE = 0.14$), $F(1.36) = 1.09$, $p = 0.30$. There was also no main effect of phase: Evaluation score on the first 4 cases ($M = 2.68$, $MSE = 0.14$) equaled that score on the last 4 cases ($M = 2.84$, $MSE = 0.17$), $F(1,36) < 1$, meaning that consistency in preceding suggested diagnoses did not influence the tendency to accept subsequent diagnostic suggestions.

There was, however, a significant interaction effect between case order and phase (see Figure 1), demonstrating that within both conditions the score on cases with correct suggestions (C/I: $M = 3.21$, $MSE = 0.20$, I/C: $M = 3.16$, $MSE = 0.25$) was higher than the score on cases with incorrect suggestions (C/I: $M = 2.53$, $MSE = 0.25$, I/C: $M = 2.16$, $MSE = 0.20$), $F(1,36) = 11.82$, $p = 0.001$, $\eta_p^2 = 0.25$.

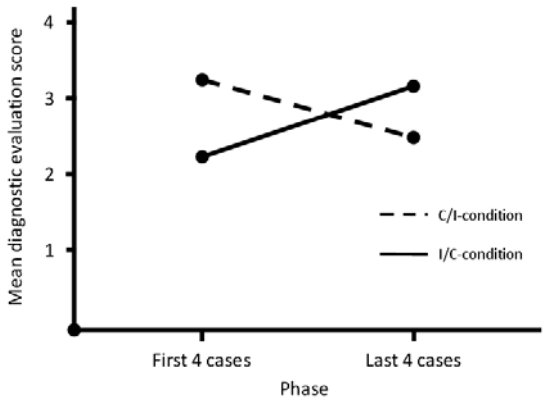


Figure 1: interaction between condition and phase of case-presentation.

DISCUSSION

The present study investigated whether physicians' tendency to accept diagnostic suggestions on written clinical cases could be influenced by a more consistent presentation of the suggestions.[1] It was hypothesized that, if a series of diagnostic suggestions would prove consistent with subsequently read case-descriptions (i.e., the suggestions were correct) this would lead to an increased confidence in the source of the suggestions, resulting in increased accepting of subsequently presented incorrect suggestions.[24] This tendency would lead to diagnostic mistakes, revealed by a lower diagnostic evaluation score on cases with incorrect suggestions than that of participants who were not first exposed to correct suggestions. Conversely, it was hypothesized that if participants would experience inconsistencies between a diagnostic suggestion and subsequent case findings, this would lead to a more critical appraisal of subsequently presented cases with correct suggestions,[23,25] resulting in a lower score on these cases by participants who had first seen incorrect suggestions than participants who were not first exposed to incorrect suggestions.

In contrast to these hypotheses, participants did not accept more incorrect suggestions after encountering a series of correct suggestions than when no prior correct suggestions had been encountered. Likewise, when participants had first evaluated incorrect suggestions they did not make more mistakes on subsequent cases with correct suggestions than their colleagues who first saw those cases with correct suggestions. Therefore, consistency in diagnostic suggestions does not seem to contribute to diagnostic errors in later cases. The significant interaction effect between condition and phase, however, showed that the rate of accepted incorrect diagnoses was, although equal between conditions, substantial within both conditions. That is, 52% of incorrect diagnoses and 80% of correct suggestions were accepted. This tendency to accept diagnostic suggestions may lead to diagnostic errors if the suggestion happens to be incorrect.[1,14,15]

It could be argued that the tendency to accept incorrect suggestions in this study results from differences in case complexity, because different cases accompanied correct and incorrect diagnostic suggestions. However, a pilot study among similar participants revealed no differences in diagnostic performance on those cases, indicating differences in case complexity are unlikely to explain this finding. In addition, differences in experience with the diagnoses presented as correct and incorrect diagnoses can also not account for participants' greater trouble with rejecting incorrect suggestions. That is, participants experience with diagnoses presented as incorrect suggestions even exceeded experience with diagnoses presented as correct suggestions. Therefore, participants are potentially able to reject these incorrect diagnostic suggestions. This implies that the tendency to accept diagnostic suggestions indeed might be hard to resist.[1]

The question is why consistency in preceding diagnostic suggestions did not influence diagnostic decisions on subsequently presented suggestions, is not easy to answer. A potential

explanation might be that exposing participants to only four cases to build up confidence or distrust, might have been insufficient. However, studies on routine behavior have shown engaging in as little as two repetitive tasks could be enough to persuade naïve participants to "stick to the routine".[28] In addition, since the diagnostic decision score on cases with correct suggestions was not perfect (i.e., approximately 80%), it could be argued that participants were not as confident in their case evaluations as we anticipated. However, scores of about 80% on cases with correct suggested diagnoses are consistent with findings in previous studies,[1-3]. Still, the cases that were used were not simple, which may explain the score on cases with correct diagnostic suggestions. Perhaps the use of very uncomplicated cases would have increased the score on cases with correct suggestions, possible resulting in higher confidence in the suggestions.

Future studies could attempt to directly measure participants' confidence in their diagnostic conclusions. Although several experimental studies have addressed physicians' confidence, [2,3,29] direct insights in physicians' confidence in their diagnostic conclusions on cases with suggested diagnoses and the actual accuracy of their diagnoses has, to the best of our knowledge, not been experimentally investigated, and might lead to further improvement of our understanding of the handling of diagnostic suggestions.

The present study has important implications for clinical practice. Consistency in diagnostic suggestions did not influence the acceptation of subsequently presented diagnostic suggestions. However, in both conditions, a substantial number of incorrect suggestions were accepted. Still, in practice diagnostic suggestions are probably correct most of the time, ignoring them would be ineffective and even unwarranted. It would therefore be much better to train physicians to identify those situations in which a diagnostic suggestion might be faulty. Research on the role of reflection in clinical practice can play an important role to help physicians to identify those situations.[26]

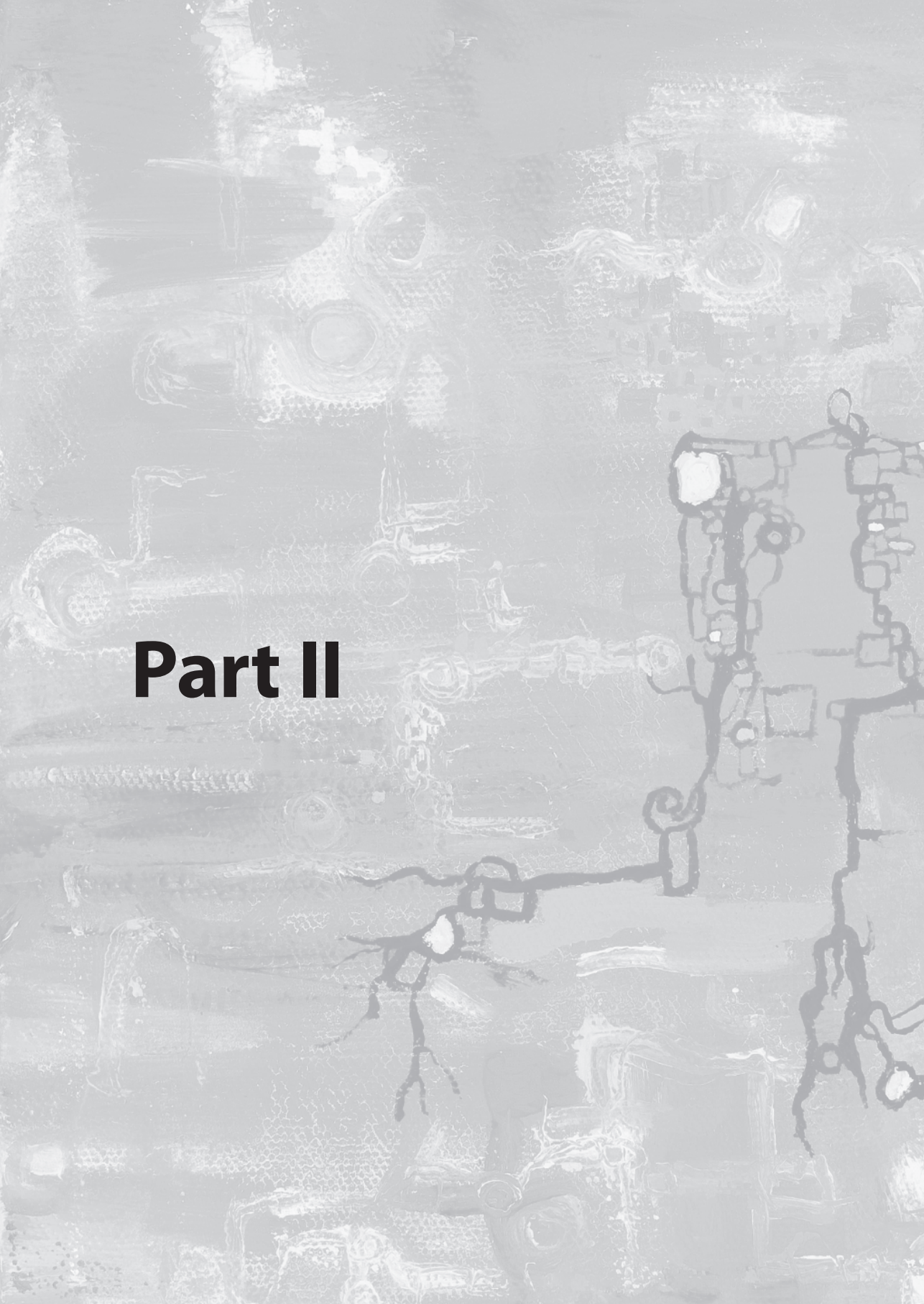
In conclusion, this study showed that physicians' tendency to accept diagnostic suggestions is independent of the correctness of preceding suggestions. Since the inclination towards accepting suggestions can, if the suggestions are incorrect, lead to errors, further study of causal and protective mechanisms should be conducted.

REFERENCES

1. Van den Berge K, Mamede S, Van Gog T, et al. Accepting Diagnostic Suggestions by Residents: A Potential Cause of Diagnostic Error in Medicine. *Teach Learn Med* In press 2012.
2. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med*. 2001;76:S18-20.
3. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med* 2002;77:S67-9.
4. Kohn KT, Corrigan JM, Donaldson MS. Editors. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 1999.
5. Stelfox HT, Palmisani S, Scurlock C, Orav EJ, Bates DW. The "To Err is Human" report and the patient safety literature. *Qual Safe Health* 2006;15:174–178.
6. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal Trends in Rates of Patient Harm Resulting from Medical Care. *New Engl J Med* 2010;363:2124-34.
7. Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three different eras. *New Engl J Med* 1983;308:1000-5.
8. Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy detected diagnostic errors over time. *JAMA* 2003;289:2849-56.
9. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study 1. *New Engl J Med* 1991;324:370–6.
10. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;38:261-71.
11. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Australia* 1995;163:458–71.
12. Graber ML, Franklin N,. Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493-9.
13. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775-80.
14. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *Brit Med J* 2005;330:781-4.
15. Redelmeier DA. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115-20.
16. Singh H, Thomas EJ, Khan MM, & Peterson LA. Identifying diagnostic errors in primary care using an electronic screening algorithm. *Arch Intern Med* 2007;167:302-8.
17. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* 2010;44:94-100.
18. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise - Theory and implications. *Acad Med* 1990;65:611-21.
19. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev* 1993;5:205-21.
20. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ* 1997;2:173-84.
21. Barrows HS, Feightner JW, Neufeld VR, Norman GR. *Analysis of the clinical methods of medical students and physicians*. Hamilton, Ontario, Canada: McMaster University School of Medicine; 1978.
22. Groopman JE. *How doctors think* (e.g., pp: 72-73). New York, NY: Mariner Books; 2007.
23. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ* 2004;38:1302-8.
24. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121:S2-S23.
25. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: Case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ* 2007;41:1185-92.

26. Mamede S, van Gog T, van den Berge K, Rikers RMJP, van Saase JLCM, van Guldener C et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;304:1198-1203.
27. Cohen J. Statistical power analysis for the behavioural sciences (2nd Ed.). Hillsdale, NJ: Erlbaum; 1988.
28. Betsch T, Haberstroh S, Glockner A, Haar T, Fiedler K. The effects of routine strength on adaptation and information search in recurrent decision making. *Organ Behav Hum Dec Proc* 2001;84:23–53.
29. Friedman CP, Gatti GC, Franz TM, Murphy GC, Wolf FM, Heckerling PS, et al. Do physicians know when their diagnoses are incorrect? Implications for decision support and error reduction. *J Gen intern Med* 2005;20:334-339.

Part II



Chapter 5

Exposure to Media Information Causes Availability Bias in Medical Diagnosis[1]

[1] This Chapter is submitted as: Schmidt HG, Mamede S, Van den Berge K, Van Gog T, Van Saase JLCM, Rikers RMJP. Exposure to Media Information Causes Availability Bias in Medical Diagnosis.

ABSTRACT

Availability bias has been pointed out as one of the culprits for wrong diagnoses in medicine, but there is little empirical evidence supporting this hypothesis. Correlational studies have suggested that media-distributed information may result in availability bias, leading doctors to overestimate the likelihood of a certain diagnosis, but this effect has not yet been experimentally demonstrated. We investigated whether exposure to media information about a disease would cause availability bias in internal medicine residents, resulting in diagnostic errors, and whether reflection would counteract such bias. First, the 38 residents were randomly assigned to read the Wikipedia entry about one of two diseases. Six hours later, as part of a seemingly unrelated study, they diagnosed eight clinical cases; two of which superficially resembled the disease in the Wikipedia entry they had read (bias expected), and two of which resembled the other disease they had not read about (bias not expected). Then, they again diagnosed the cases subject to bias, but using structured reflection. Results showed that the exposure to a Wikipedia entry led to diagnostic error, with the average number of cases mistakenly diagnosed as the 'Wikipedia disease' being significantly higher than when those cases were diagnosed without previous exposure. However, reflection restored diagnostic performance to non-biased levels.

INTRODUCTION

Diagnostic errors and patients' safety have always been a concern in health care, but have attracted even more attention from researchers, policy makers, as well as the public in the last decade, after the Institute of Medicine reported in 1999 that between 48000 and 98000 people die annually in the US as a result of preventable medical mistakes.[1] Although the majority of these mistakes related to therapy, diagnostic errors also constituted a substantial proportion.[1] Diagnostic mistakes occur in every specialty, although error rates tend to be higher in clinical specialties that require complex decision making in contexts of high uncertainty, such as internal medicine and emergency medicine.[2,3] In medical practice, multiple factors may interact to produce diagnostic errors, but it seems that a significant proportion of such errors can be attributed, at least in part, to individual physicians' cognitive processes. Graber, Franklin, and Gordon investigated 100 cases of diagnostic errors in internal medicine in five American academic hospitals and found that cognitive factors were present in 74% of the cases.[4] Knowledge deficiency was implicated in only a minor proportion; the great majority of these errors derived from flaws in the physicians' reasoning processes.

The medical literature is rife with suggestions that cognitive biases play an important role in errors of judgment in medical diagnosis,[5-7] based on psychological research which has shown that an important cause of flaws in reasoning processes is the application of particular heuristics, which may lead to bias.[8] For example, the *availability* heuristic, on which the present study focuses, leads people to assess the likelihood or probability of an event by the ease with which instances come to mind.[8,9] As Tversky & Kahneman point out, this heuristic is quite useful in general, as events that occur more frequently are easier to recall or imagine than less frequent ones.[9] However, the ease with which instances come to one's mind is also affected by other factors, such as the person's own recent experiences. As a result, the use of this heuristic may generate availability bias, resulting in errors of judgment with potentially severe adverse consequences.

Making judgments based on shortcuts in reasoning such as the availability heuristic, though subject to bias, is a powerful adaptive strategy in light of the characteristics of human thinking. In dual-process theories of reasoning, for example, two generic modes of cognitive function are distinguished: a non-analytical mode (or System 1), in which judgments and decisions are made through a rapid, effortless, largely unconscious, and contextualized process, and an analytical mode (or System 2), which is a slow, effortful, logical and decontextualized process that takes places under conscious control.[10,11] Studies on the nature of expertise have shown that experienced professionals tend to rely largely on non-analytical reasoning to make routine decisions rather than engaging in a conscious, effortful analysis of the problem at hand. They are able to do so thanks to the large storage of examples of previously solved problems that they have in memory. Indeed, while early research on human problem-solving sought general strategies,[12] study of expertise in domains like chess and medicine revealed that a central

feature of expertise is deliberate practice, in which the expert consciously masters many specific strategies designed to deal with the myriad of possible problem presentations in the area. One example is chess, where it is estimated that a master, in the course of 10,000 hours of deliberate practice toward mastery, has acquired 50,000 separate chess strategies.[13] The availability of such rehearsed strategies in turn reduces problem solution to one of recall of an appropriate learned solution.

Developing expertise in medicine has similar characteristics. We tend to think of medical experts in terms of their deep understanding of biological processes and their relation to health and disease. In fact, during the roughly 6-12 years of training a specialist will undergo prior to being permitted independent practice (after undergraduate education, a physician engages in 3-6 years of specialty training aimed at a board certification as a specialist in that medical domain), only the first 1.5 -2 years are devoted to primary understanding of biological processes. The remaining time occurs in clinical care settings, where formal conceptual learning is secondary and haphazard, and the primary goal is to acquire extensive experience with the many ways disease can present. Thus, much of medical training, while perhaps lacking the deliberative and carefully structured nature of more formal practice, nevertheless remains devoted to acquisition of numerous examples. Such examples allow expert physicians to diagnose most of the cases encountered in daily practice by relying on pattern-recognition, that is, on recognition of similarities between the case at hand and examples of previous cases stored in memory (as concrete instances;[14] or in the form of illness scripts,[15-17]). Reliance on non-analytical reasoning has many benefits, as it is not only highly effective (i.e., it will very often lead to accurate diagnoses), but also highly efficient (i.e., it saves time, as well as effort). However, it has also been suggested that, since it takes place without conscious control, non-analytical reasoning makes individuals more prone to bias, which could be prevented or counteracted by analytical reasoning.[10]

The medical literature has frequently claimed that cognitive bias is an important cause of errors of judgment in medical diagnosis; however, this literature is mainly descriptive in nature.[2,3,5-7,18-21] Some experimental evidence exists for confirmation bias (i.e., the tendency to focus on evidence that confirms a belief or hypothesis). For example, Brooks, LeBlanc, and Norman and LeBlanc and colleagues showed that when students and physicians were provided with a suggested diagnosis, they tended to seek and interpret evidence (case features) in favor of that diagnosis, regardless of whether it was correct or incorrect.[22-24]

For availability bias in medical diagnosis, there is much less experimental evidence. A recent study by Mamede, Van Gog et al.[25] showed that second-year internal medicine residents who had been previously exposed to a disease (i.e., they were asked to confirm a diagnosis provided for a case, e.g., acute viral hepatitis) subsequently provided that disease as a diagnosis for cases about patients with different diseases that had similar signs and symptoms (e.g., liver cirrhosis or primary sclerosing cholangitis) more often than when they had not been previously exposed to that disease.

Note that both the studies by LeBlanc et al. and Mamede, Van Gog et al.[23-25] used a biasing task that is rather close to the diagnostic task (i.e., evaluating given diagnoses) that was part of the same experimental context as the diagnostic task, with only a very short or no time interval in between. The question is, then, whether such bias can also arise from information encountered in another context, from other kinds of sources, such as the media, and over an extended period of time.

One non-experimental study provided some indirect evidence that availability bias may indeed occur in medical diagnosis as a result of exposure to media-distributed information on a disease:[26] during outbreaks of West Nile fever in Israel and the USA, an association was demonstrated between the number of weekly reports in the mass media about the disease and the number of laboratory samples submitted to test for West Nile virus. The authors suggested that this increase in the number of lab tests ordered was due to an availability bias in the doctors concerned, caused by the media. However, a doctor ordering such tests is not necessarily biased. If West Nile fever is around, it slightly raises the chance that this particular patient (in front of me) with symptoms resembling those of that disease actually has Nile fever. Testing for the West Nile fever would then not be a manifestation of bias but simply an act of cautiousness. In addition, it is not certain that the doctors involved really *believed* that their patients had the disease. It may also have been their *patients* who were influenced by the media coverage and believed that they might have the disease, therefore insisting on testing (indeed, there is some evidence that priming can lead people to report higher frequencies of physical symptoms[27]). Such interpretational degrees of freedom are inevitable in non-experimental studies. An experimental approach is required to clarify these issues.

One question that arises if empirical research would indeed show that availability bias occurs due to media information, is how to prevent or counteract it. The medical education literature has frequently called for "debiasing strategies" and training of metacognitive or reflective strategies as a means to reduce mistakes.[2,3,28] For example, inspired by an analogy between the domains of aviation and medicine,[29] Singh, Petersen, & Thomas suggested that enhancing situational awareness might reduce diagnostic errors.[30] Others have argued that increasing physicians' awareness of potential biases under conditions of non-analytical reasoning would make them less susceptible to errors.[28] Again, however, there seems to be much more descriptive research advocating instructional strategies to induce analytical thinking as a means to counteract bias than experimental research providing evidence that they actually do. A few experimental studies showed that instructions to induce reflection improved physicians' diagnostic performance when they were confronted with complex clinical cases[31,32] and that undergraduate students benefited from combining non-analytical and analytical reasoning while learning how to interpret ECGs.[33] The study by Mamede, Van Gog et al.[25] was the first to experimentally investigate whether reflection could counteract bias, with positive results.

In sum, experimental evidence for the occurrence of availability bias in medical diagnosis is scarce, and so is evidence that instructions aimed to induce analytical reasoning by means of

reflection would help in counteracting bias. One experimental study has suggested that availability bias might occur.[25] However, this study did not investigate the effects of media distributed information. So even though there are some correlational data that suggest availability bias might occur due to media distributed information,[34] this has not yet been experimentally established. It is important to address this question not only for medical practice, because media information is ubiquitous; it is also interesting from a scientific perspective, as media distributed information is very different in nature from diagnosis and hence from previous tasks that have been shown to induce bias, such as evaluating a given diagnosis.[24,25] Moreover, in previous studies, the task that induced bias (e.g., evaluating a given diagnosis) was usually given in the same context in which the diagnoses subsequently had to be made, with a very short or no time interval in between, so an open question is whether or not bias would occur due to information encountered in another context, with a substantial time interval in between.

The present study addresses that question, investigating whether mere exposure to media-distributed disease information would bias doctors into using that information in a seemingly unrelated context after a substantial time interval, leading to diagnostic errors. The second purpose of our study was to investigate whether inducing analytical reasoning by means of instructions to engage in structured reflection could restore performance to non-biased levels, as the study by Mamede et al.[25] suggested.

METHOD

Participants and Design

Thirty-eight residents in internal medicine (mean age 28.97 years, *SD* = 2.25; 23 female) from four Dutch teaching hospitals volunteered to participate in this study. The teaching hospitals were associated with the Faculties of Medicine of the Erasmus University Rotterdam (2), Radboud University Nijmegen (1), and Maastricht University (1). Residents are physicians in training to become a specialist, and are responsible to a large extent for first-aid and in-patient care in teaching hospitals.

This study had a three-phases design (see Table 1): 1) *exposure to disease information reported by the media*; participants were randomly assigned to evaluate the accuracy of the Wikipedia entry either for Legionnaires' disease or for Q fever; 2) *non-analytical diagnosis* of 8 clinical cases. All cases had a diagnosis different from the diseases in Phase 1, but 2 of those 8 cases had signs and symptoms similar to Legionnaires' disease and 2 resembled Q fever; 3) *reflective diagnosis* of the 2 cases from Phase 2 that resembled the disease they had been exposed to in Phase 1. Phase 1 and Phases 2 and 3 were presented as two separate and unrelated studies from different institutes (see materials and procedure section for more details), with Phase 1 taking place 6 hours before the two other phases. Availability bias was expected in Phase 2, for the 2 cases similar to the disease that the participant had encountered in Phase 1. If such bias occurred,

Table 1. Overview of Study Design and Materials

Phase 1: Wikipedia article	Phase 2: Diagnosing 8 cases non-analytically	Phase 3: Diagnosing 4 cases reflectively
Legionnaires' disease (*n* = 19)	• Pneumococcal pneumonia (bias expected) • Community-acquired pneumonia (bias expected) • Acute bacterial endocarditis (bias unlikely) • viral respiratory infection (bias unlikely) • 4 filler cases	• Pneumococcal pneumonia • Community-acquired pneumonia • 2 filler cases
Q fever (*n* = 19)	• Acute bacterial endocarditis (bias expected) • Viral respiratory infection (bias expected) • Pneumococcal pneumonia (bias unlikely) • Community-acquired pneumonia (bias unlikely) • 4 filler cases	• Acute bacterial endocarditis • Viral respiratory infection • 2 filler cases

Note: In Phases 2 and 3 the cases were presented in random order

participants would tend to misdiagnose those cases looking similar to Legionnaire's disease or Q fever as being Legionnaire's disease or Q fever, thus reducing diagnostic accuracy for those cases. The instructions for structured reflection upon the possibly biased diagnoses in Phase 3 was predicted to override the bias.

The recruitment of participants was conducted via the coordinators of the residency programs, who informed their residents that two requests for participation in studies had arrived, forwarding the "official" letters from two different institutes (different logos, et cetera) and asking them to indicate whether or not they would be interested in participating in these studies. The majority (ca. 85%) of residents approached in this manner volunteered to participate in both studies.

Materials and Procedure

Table 1 provides an overview of the materials used in the different Phases of the study. Phase 1 took place 6 hours before the two subsequent phases and was presented as an unrelated study. Participants were informed that patients tend to consult Internet sources before visiting their doctor and that it is therefore important that these sources contain accurate information. The purpose of the study, they were told, was to gain insight into the accuracy of information provided by one of the most often consulted Internet sources, the Wikipedia encyclopedia. Participants received a paper copy of the Dutch Wikipedia entry (retrieved on April 7, 2009 from http://nl.wikipedia.org) on either Legionnaires' disease or Q fever depending on their assigned condition. They were instructed to underline correct statements about epidemiology, transmission, symptoms, and therapy encountered in the text, as well as to judge the accuracy, completeness, how up-to-date it was, and the clarity of the information by rating their

agreement with one statement about each of these aspects on a 5-point scale ranging from (1) 'strongly disagree' to (5) 'strongly agree'.

Phases 2 and 3 took place 6 hours later in a session which was presented as another, unrelated study (i.e., on medical diagnosis), was led by a different experimenter and had materials presented in the letterhead of a research institute different from the materials provided in Phase 1. During the 6-hour time interval between the 'two studies', participants went back to their routine duties. In Phase 2, participants were presented with a booklet containing 8 clinical cases in random sequence. The cases consisted of a written description of a patient's medical history, signs and symptoms, and tests results (see Appendix 1 for an example). All cases were prepared by experts in internal medicine based on their experience with real patients and had a confirmed diagnosis. Two of those eight cases concerned diseases with signs and symptoms similar to those frequently encountered in patients with Legionnaires' disease (i.e., pneumococcal pneumonia and community-acquired pneumonia), and two others concerned diseases with signs and symptoms similar to those frequently encountered in patients with Q fever (i.e., acute bacterial endocarditis and viral respiratory infection). The remaining four cases were fillers (aortic dissection, acute alcoholic pancreatitis, acute viral pericarditis, and appendicitis). Participants were instructed to read the case and write down the first diagnosis that comes to mind.

In Phase 3, participants were presented with a booklet that contained 4 of the 8 cases from Phase 2 in random order: 2 of the filler cases and the 2 cases that resembled the case from the Wikipedia entry from Phase 1 (i.e., pneumococcal pneumonia and community-acquired pneumonia for participants who received the Wikipedia entry on Legionnaires' disease, and acute bacterial endocarditis and viral respiratory infection for participants who received the Wikipedia entry on Q fever). Participants were instructed to: 1) read the case again; 2) write down the diagnosis given for the case in Phase 2; 3) list the findings in the case description that support this diagnosis; 4) list the findings that speak against this diagnosis; 5) list the findings that would be expected to be present if this diagnosis were true but that were not described in the case; 6) list possible alternative diagnoses and repeat the first five steps for each alternative diagnosis; 7) rank the alternative diagnoses in order of likelihood and select one's final diagnosis.[25,31,32]

After completing Phase 3, participants were asked whether they had been aware of any connection between the 'two studies' but none of the participants indicated they had been. They were later debriefed that the 'two studies' were actually one and were provided with information about the purposes and theoretical background of the study.

Data Analysis

Two experts in internal medicine independently assessed the diagnoses provided by the participants by comparing them to the confirmed diagnoses of the cases. Participants' diagnoses were

scored as *correct*, *partially correct* or *incorrect*, and assigned 1, 0.5, or 0 points, respectively. A diagnosis was considered correct whenever the core diagnosis was cited by the participant. When the core diagnosis was not mentioned but a constituent element of the diagnosis was cited, the diagnosis was scored as partially correct (e.g., in the case of community-acquired pneumonia, "pneumonia " was scored as correct, and "dehydration" as partially correct). The experts agreed upon 88% of the diagnoses, and disagreements were resolved through discussion.

For each participant, the mean diagnostic accuracy was computed for the 2 cases in Phase 2 and 3 that had similarities with the disease encountered while evaluating the Wikipedia entry in Phase 1 (i.e., on which an availability bias could be expected to occur, reducing diagnostic accuracy), and for the 2 cases in Phase 2 that resembled the other disease which they did not read about. Even though they did not see this Wikipedia entry, they might still provide the Wikipedia disease as a diagnosis in Phase 2 because of the similarities between cases. Therefore, the number of times the Phase 1 Wikipedia disease diagnosis was provided as a diagnosis for those cases was also counted and analyzed. If availability bias occurs, the Phase 1 Wikipedia disease diagnosis should be provided more often for the similar cases in Phase 2 by participants who did encounter the similar disease in Phase 1 than by participants who did not.

RESULTS

For all analyses reported here, a significance level of .05 is used (two-tailed). Cohen's *d* is reported as a measure of effect size, with .20, .50, and .80 corresponding to small, medium, and large effect sizes, respectively.[35]

A paired *t*-test showed that in line with our hypothesis, diagnostic accuracy (max. = 2) in Phase 2 was significantly lower when participants had been exposed in Phase 1 to information about the disease similar to the to-be-diagnosed cases ($M = 0.54$, $SD = 0.24$) than when they had not ($M = 0.67$, $SD = 0.33$), $t(37) = 2.30$, $p = .027$, $d = .45$. Also in line with our hypothesis, this decrease in diagnostic accuracy was the result of availability bias: A paired *t*-test showed that participants provided a higher number of wrong diagnoses of Q fever or Legionnaires' disease when they had read the Wikipedia information on this disease ($M = 0.30$, $SD = 0.27$) than when they had not ($M = 0.14$, $SD = 0.23$), $t(37) = 3.14$, $p = .003$, $d = .63$.

However, as we hypothesized, a paired *t*-test comparing diagnostic performance in Phase 2 and 3 on the two cases that resembled the Wikipedia entry, showed that reflection (Phase 3) could significantly improve diagnostic accuracy on those cases that were affected by availability bias in Phase 2 (from $M = 0.54$, $SD = 0.24$ in Phase 2 to $M = 0.68$, $SD = 0.33$ in Phase 3), $t(37) = 2.52$, $p = .016$, $d = .48$. Note that reflection increased performance to the same level ($M = 0.68$, $SD = 0.33$) as performance on the test cases that were not subject to bias in Phase 2 ($M = 0.67$, $SD = 0.33$). This improvement was due to a counteraction of bias: After reflection, the number of biased diagnoses (i.e., Q fever or Legionnaires' disease depending on the Wikipedia information

they had read) was significantly reduced (from $M = 0.30$, $SD = 0.27$ in Phase 2 to $M = 0.18$, $SD = 0.32$ in Phase 3), $t(37) = 2.16$, $p = .037$, $d = .41$.

DISCUSSION

These findings provide experimental evidence that availability bias may emerge from exposure to disease information in the media that was irrelevant to the task at hand, encountered in a different context, several hours before. The size of the bias was substantial; reading about the diseases in the Wikipedia entries increased the number of cases mistakenly diagnosed as one of those diseases by 100%. Studies on the availability heuristic among naïve participants, for example, with frequency-of-occurrence judgments,[36] self-judgments of assertiveness,[37] or vulnerability to heart disease,[38] have shown much smaller effects.

Even more important than the size of the bias, however, was the fact that this the effect emerged from a task carried out several hours earlier in a context entirely different from the diagnostic task. Most studies on the availability heuristic requested participants to make judgments about an event immediately after performing a task, which is expected to make that event easily retrievable from mind.[9,25,36,37] In our study, there was a timelag of several hours that separated the two tasks, during which the physicians were engaged in their routine clinical duties, encountering several patients with a diversity of problems. Our findings provide, therefore, an important addition to the literature on availability bias, by demonstrating that this bias may occur not only after a short time interval when the biasing information is still fresh in mind and therefore easier to retrieve, but also after a substantial 6-hour time interval during which participants' minds were occupied with their regular duties.

Moreover, the exposure to the biasing information occurred out of the context of the diagnostic task, that is, both sessions were presented as different studies, were led by different experimenters, and the materials contained letterheads of different institutions. This reduced the chance that participants used the information to which they had been previously exposed simply because they might consider it to be relevant somehow, something they might consider (despite instructions to the contrary) when the exposure is part of the same session. Finally, whereas participants in most studies on the availability heuristics were naïve, e.g., undergraduate students,[8,9,36,37] the present study showed bias to also occur among fairly experienced residents, who were shown to possess sufficient knowledge to counteract bias.

Presumably, fairly experienced physicians are so sensitive to information seemingly irrelevant to the task at hand because they rely heavily on non-analytical reasoning processes, making extensive use of pattern-recognition.[39,40] Non-analytical reasoning processes are usually efficient in routine situations, but they are also liable to bias. In particular, there is an accumulation of evidence that retrieval of a similar episode from memory may be influenced by factors other than its logical relevance to the problem at hand. For example, in one study,

resident physicians' diagnostic accuracy in reading ECGs was strongly influenced by a recent prior case with the same occupation and age (e.g. a 46 year old banker).[41] In another study, novice diagnosticians' probability judgments were influenced by the similarity of a name (MacIntosh vs. McKinley) and occupation (teacher, professor).[42] It is known that details of a case previously seen, even if irrelevant to the problem at hand, may influence the diagnosis of a subsequent similar looking case.[14] This is hardly surprising, in view of the fact that such non-analytical processes are automatic, and not subject to conscious introspection.[43] Bias emerged in these studies, however, in the context of previous experience with similar patients. One could argue that these potential biases may occur infrequently in what is a basically adaptive mental strategy - reasoning from prior cases. However, the present study extends the retrieval process beyond previous cases to other, only peripherally related, kinds of knowledge - in this case, a brief text read several hours earlier in an unrelated context - and shows that prior exposure to such media information can have large consequences on diagnostic accuracy.

We hypothesize that media information about a disease has this effect because having been exposed to that information makes some of the features in a similar looking case become more salient for a physician than they would be if such information about a disease had not been encountered. Such salient features will get the physician's attention, presumably to the expense of attention to other (more relevant) features, which leads to errors. The method of structured reflection that was used "forced" physicians to focus on all features before making a decision, and once their attention was focussed on the relevant features, they were more likely to make the correct decision. This also shows they do have the required knowledge about the disease, in line with the conclusions by Graber et al.[4] These are assumptions about mechanisms of media induced bias and reflection though, that should be directly investigated in future research.

A limitation of this study related to the above is that it is unclear whether it were our specific instructions for structured reflection, or the fact that additional time and effort was spent on cases that led to performance improvement in Phase 3. Analytical reasoning approaches are by definition slower and more effortful, and the reverse could, but does not necessarily have to be true, that is, more investment of time or effort might or might not lead to prevention/correction of errors. We assume that it is also important in what processes time and effort is invested (i.e., that the performance improvement resulting from our structured reflection instructions is not so much due to the additional time investment, as it is to the fact that participants are not just asked to generate alternative plausible diagnoses, but are also focussed on the features supporting those diagnoses). Prior research seems to support the notion that just giving physicians more time might not prevent errors, as findings show that physicians often become anchored in their initial hypothesis, looking for confirming evidence to support their initial diagnosis, underestimating evidence against it, and therefore failing to adjust their initial impression in light of all available information.[3,6] In addition, premature closure, that is, failing to continue considering reasonable alternatives after reaching an initial diagnosis, was identified as the most common cause of cognitive errors in the study by Graber et al.[4]

Despite the fact that we are not able to disentangle this issue based on our data, our findings do show that a more analytical reasoning approach seems to be an effective way to restore performance. This is reassuring, because when the effect of availability bias is so substantial and so easily produced among fairly experienced doctors, one must fear for the societal consequences. Our findings seem to lend credibility to the claims that cognitive errors account for large proportion of all medical mistakes and are directly implicated in adverse outcomes and deaths,[4-6] but also to claims that these biases are mainly associated with non-analytical thinking.

Regarding implications for medical education and practice, our findings support suggestions that physicians should be encouraged to use both analytical and non-analytical reasoning.[21] The question is how to train them to do so. Non-analytical reasoning is highly efficient for routine situations, however, when cases appear similar to often encountered ones but are not, it may also lead to errors. The paradox is of course that people are not able to recognize when this happens (or bias would not occur). In other words, it is unlikely that physicians will know when a more analytical approach would be appropriate. If analytical reasoning, for example by means of structured reflection, would always be used, the number of errors due to bias could potentially be reduced; however, due to the time and effort required this would not be very practical in clinical settings –unless this process of structured reflection itself could somehow be internalized and automated through training, in which case it could be rapidly executed. It would therefore be important to establish in future research whether this is a skill in itself that can be trained, whether extended periods of such training would speed up the structured reflection process without compromising its essence, and whether it could be transferred from educational to practical situations.

REFERENCES

1. Kohn KT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
2. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121(5):S2-S23.
3. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775-80.
4. Graber ML, Franklin N,. Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493-9.
5. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *Brit Med J* 2005;330:781-4.
6. Redelmeier DA. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115-20.
7. McDonald CJ. Medical heuristics: The silent adjudicators of clinical practice. *Ann Intern Med* 1996;124:56-62
8. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185:1124-31.
9. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychol* 1973;5:207-32.
10. Evans JSBT. Dual processing accounts of reasoning, judgment and social cognition. *Ann Rev Psychol* 2008;59:255-78.
11. Stanovich KE, West RF. Discrepancies between normative and descriptive models of decision making and the understanding acceptance principle. *Cognitive Psychology* 1999;38:349-385.
12. Newell A., Simon HA. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972.
13. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;79:S70-S81.
14. Brooks LR, Norman GR, Allen SW. Role of specific similarity in a medical diagnostic task. *J Exp Psychol Gen* 1991;120(3):278-287.
15. Boshuizen HPA, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Sci* 1992;16:153-184.
16. Custers EJFM, Boshuizen HPA, Schmidt HG. The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Mem Cognition* 1996;24:384-399.
17. Van Schaik P, Flynn D, Van Wersch A, Douglass A, Cann P. Influence of illness script components and medical practice on medical decision making. *J Exp Psychol-Appl* 2005;11:187-199.
18. Bornstein BH, Emler AC. Rationality in medical decision making: A review of the literature on doctors' decision making biases. *J Eval Clin Pract* 2001;7(2):97-107.
19. Dawson NV, Arkes HR. Systematic errors in medical decision making: Judgment limitations. *J Gen Intern Med* 1987;2:183-7.
20. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *Brit Med J* 2002;324:729-32.
21. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* 2010;44:94-100.
22. Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. *Psychol Sci* 2000;11:112-117.
23. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med* 2001;76:S18-20.

24. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med* 2002;77:S67-9.
25. Mamede S, van Gog T, van den Berge K, Rikers RMJP, van Saase JLCM, van Guldener C et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;304:1198-1203.
26. Brezis M, Halpern-Reichert D, Schwaber MJ. Mass media-induced availability bias in the clinical suspicion of West Nile fever. *Ann Inter Med* 2004;140(3):234-5.
27. Skelton JA, Strohmetz, DB. Priming symptom reports with health-related cognitive activity. *Pers Soc Psychol B* 1990;16:449-464.
28. Croskerry P. Cognitive forcing strategies in clinical decision making. *Ann Emerg Med* 2003;41(1):110-120.
29. Durso FT, Drews FA. Health care, aviation, and ecosystems: A socio-natural systems perspective. *Curr Dir Psychol Sci,* 2010;19:71-5.
30. Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: A lesson from aviation. *Qual Safety Health Care* 2006;15:159-64.
31. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008;42:468-75.
32. Mamede S, Schmidt HG, Rikers RMJP, Custers EJFM, Splinter TAW, Van Saase JLCM. Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychol Res* 2010;74:586-592.
33. Ark TK, Brooks LR, Eva KW. The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Med Educ* 2007;41:281-287.
34. Brezis M, Halpern-Reichert D, Schwaber MJ. Mass media-induced availability bias in the clinical suspicion of West Nile fever. *Ann Inter Med* 2004;140(3):234-5.
35. Cohen J. Statistical power analysis for the behavioural sciences (2nd Ed.). Hillsdale, NJ: Erlbaum; 1988.
36. Maley JE, Hunt M, Parr W. Set-size and frequency-of-occurrence judgments in young and older adults: The role of the availability heuristic. *Q J Exp Psychol* 2000;53A:247-270.
37. Schwarz N, Bless H, Strack F, Klumpp G, Rittenauer-Schatka H, Simons A. Ease of retrieval as information: Another look at the availability heuristic. *J Pers Soc Psychol* 1991;61:195–202.
38. Schwarz N. When thinking feels difficult: meta-cognitive experiences in judgment and decision making. *Med Decis Making* 2005;25:105-112.
39. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol* Rev 1993;5:205-21.
40. Boshuizen HPA, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. Cognitive Sci 1992;16:153-184.
41. Hatala RM, Norman GR, Brooks LR. The impact of a clinical scenario upon electrocardiogram interpretation. *J Gen Intern Med* 1999;14:126–129.
42. Hatala R, Norman GR, Brooks LR. Influence of a single example on subsequent electrocardiogram interpretation. *Teach Learn Med* 1999;11(2):110-117.
43. Kahneman D. A perspective on judgment and choice - Mapping bounded rationality. Am Psychol 2003;58(9):697-720.

Chapter 6

Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents[1]

[1] This Chapter was published as as: Mamede S, Van Gog T, Van den Berge K, Rikers RMJP, Van Saase JLCM, Van Guldener C, Schmidt HG. Effect of Availability Bias and Refelective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents. *JAMA* 2010;304(11):1198-1203.

ABSTRACT

Context. Diagnostic errors have been associated with bias in clinical reasoning. Empirical evidence on the cognitive mechanisms underlying biases and effectiveness of educational strategies to counteract them is lacking.

Objectives. To investigate: (1) whether recent experience with clinical problems provokes availability bias (overestimation of the likelihood of a diagnosis based on the ease with which it comes to mind), resulting in diagnostic errors, and (2) whether reflection (structured re-analysis of the case findings) counteracts this bias.

Design, Setting, and Participants. Experimental study conducted in 2009 at the Erasmus Medical Centre, Rotterdam, with 18 first-year and 18 second-year internal medicine residents. Participants first evaluated the diagnoses of 6 clinical cases (Phase 1). Subsequently, they diagnosed 8 different cases through non-analytical reasoning, 4 of which had findings similar to previously evaluated cases, but different diagnoses (Phase 2). These 4 cases were subsequently diagnosed again through reflective reasoning (Phase 3).

Main Outcome Measures. Mean diagnostic accuracy scores (perfect = 4.0) on cases solved with or without previous exposure to similar problems, through non-analytical (Phase 2) or reflective (Phase 3) reasoning; and frequency that a potentially biased (ie, Phase 1) diagnosis was given.

Results. There were no main effects, but there was a significant interaction effect between "years-of-training" and "recent experiences with similar problems". Results consistent with an availability bias occurred for the second-year residents, who scored lower on the cases similar to those previously encountered (1.55; 95% confidence interval [CI], 1.15-1.96) than on the other cases (2.19; 95% CI, 1.73-2.66). They provided the Phase 1 diagnosis more frequently for Phase 2 cases they had previously encountered than for those they had not (mean frequency per resident, 1.44; 95% CI, 0.93 – 1.96; vs , 0.72; 95% CI, 0.28 – 1.17; $p = 0.04$). This pattern was not seen among the first-year residents for the similar vs other cases (2.03; 95% CI, 1.55-2.51; vs 1.42; 95% CI, 0.92-1.92). A significant main effect of "reasoning mode" was found: reflection improved the diagnoses of the similar cases compared to non-analytical reasoning for the second-year residents (2.03; 95% CI, 1.49-2.57) and the first year residents (2.31; 95% CI, 1.88-2.73).

Conclusion. When faced with cases similar to previous ones and using non-analytic reasoning, second-year residents made errors consistent with the availability bias. Subsequent application of diagnostic reflection tended to counter this bias; it improved diagnostic accuracy in both first- and second-year residents.

INTRODUCTION

A major aim of every clinical teacher is to foster the quality of students' and residents' clinical reasoning, one of the most important factors affecting individual physicians' performance.[1] Diagnostic errors constitute a substantial proportion of preventable medical mistakes,[2] and they have been attributed to a large extent to faulty clinical reasoning.[1] The development of educational strategies to minimize flaws in clinical reasoning depends on a better understanding of their underlying cognitive mechanisms.

Cognitive biases are one source of flaws in reasoning processes.[3] At least 40 types of biases that may affect clinical reasoning have been described.[4,5] A prime example is the biased use of the *availability* heuristic (the tendency to weigh likelihood of things by how easily they are recalled), which may erroneously lead a physician to consider a diagnosis more frequently and judge it as more likely if it comes to mind more easily.[4,6] Relying on availability is often helpful during reasoning, because things that come to mind easily generally do occur more frequently. However, a serious problem may arise when this first impression is wrong, because physicians often become anchored in their initial hypothesis, looking for confirming evidence to support their initial diagnosis, underestimating evidence against it, and therefore failing to adjust their initial impression in light of all available information.[4,7]

The scientific literature on the availability bias in medicine is mainly descriptive. Some correlational studies[8-11] suggest that it occurs, but these do not allow causal inferences to be made. Experimental research is required to provide direct evidence for availability bias in medical diagnosis but, to the best of our knowledge, is lacking. Moreover, if documented, it is perhaps even more important to medical education and practice to investigate ways in which availability bias can be counteracted.

Expertise might play a role in bias. Experienced physicians tend to rely more on *non-analytical* (or System 1) *reasoning* based on pattern recognition to diagnose routine problems; this is a rapid, largely unconscious diagnostic approach. Although effective (and highly efficient) in most cases, it might be more easily affected by biases.[12,13] One way to counteract biases suggested by studies in psychology[4,14] is to induce physicians to adopt more *reflective* (or analytical, also referred to as System 2) *reasoning*, which comprises careful, effortful consideration of findings in a case, or to combine non-analytical and analytical reasoning[15].

We therefore investigated whether availability bias occurs when physicians diagnose cases that have clinical manifestations similar to those of recently encountered cases, and, if so whether reflection could counteract this bias. Because non-analytical reasoning develops in association with clinical experience, we also investigated whether there would be a difference in the degree of bias between residents in the first and second year of the residency program. We hypothesized that (1) recent experiences with clinical problems would generate an availability bias when physicians non-analytically diagnose subsequent cases of similar diseases;

(2) more experienced residents would be more prone to this bias; and (3) reflective reasoning would counteract this bias and improve diagnostic accuracy.

METHODS

Overview

This experiment consisted of 3 phases conducted sequentially in a single session (Table 1). *Phase 1*, exposure, required participants to evaluate the accuracy of a diagnosis provided for

Table 1. Clinical cases used in each phase of the study.

	Phase 1 Exposure (diagnosis evaluation task)	Phase 2 Non-analytical diagnostic reasoning	Phase 3 Reflective diagnostic reasoning
Set 1	Case A: Acute viral hepatitis	Cases similar to case A[a] • Liver cirrhosis • Primary sclerosing cholangitis	Cases similar to case A: • Liver cirrhosis • Primary sclerosing cholangitis
	Case B: Inflammatory bowel disease	Cases similar to case B[a] • Celiac Disease • Pseudomembranous colitis	Cases similar to case B: • Celiac Disease • Pseudomembranous colitis
	Neutral case 1: Meningitis Neutral case 2: Pyelonephritis Neutral case 3: Pneumonia Neutral case 4: Hyperthyroidism	Cases similar to case C[b] • Acute viral pericarditis • Aortic dissection	
		Cases similar to case D[b] • Neurosyphilis • Vitamin B12 Deficiency	
Set 2	Case C: Acute myocardial infarction	Cases similar to case C[a] • Acute viral pericarditis • Aortic dissection	Cases similar to case C: • Acute viral pericarditis • Aortic dissection
	Case D: W's Encephalopathy	Cases similar to case D[a] • Neurosyphilis • Vitamin B12 Deficiency	Cases similar to case D: • Neurosyphilis • Vitamin B12 Deficiency
	Neutral case 1: Meningitis Neutral case 2: Pyelonephritis Neutral case 3: Pneumonia Neutral case 4: Hyperthyroidism	Cases similar to case A[b] • Liver cirrhosis • Primary sclerosing cholangitis	
		Cases similar to case B[b] • Celiac Disease • Pseudomembranous colitis	

[a] Cases potentially subject to bias; [b] Cases not subject to bias

6 different cases. *Phase 2*, non-analytical diagnosis, required participants to diagnose 8 new cases, 4 of which had clinical manifestations that were similar to 2 of the diseases encountered in Phase 1. This was expected to induce an availability bias for those 4 cases and reduce diagnostic accuracy. *Phase 3*, reflective diagnosis, required participants to reflect on the diagnosis of the 4 cases that could have been influenced by an availability bias in Phase 2. This was expected to overrule the bias and lead to more accurate diagnoses.

Participants

Thirty-six out of 42 eligible internal medicine residents (participation rate = 85.7%) from the Erasmus Medical Centre, Faculty of Medicine, Erasmus University Rotterdam (mean age, 29.5 y; SD, 2.1) in their first (n = 18) or second (n = 18) year of the residency program volunteered to participate in this study. It took place during an educational meeting held in September 2009; the academic year starts in January for the majority of the residents. Participants did not receive any compensation or other incentives. The nonparticipants were either doing shifts or on holidays. The ethics review committee from the Department of Psychology, Erasmus University Rotterdam, provided approval for this study. Because the nature of the study prevented prior disclosure of its objectives, oral consent was obtained after informing participants about their tasks. Debriefing was provided later.

Procedure

In total, 16 written clinical cases were used in this study (Table 1). Cases consisted of a brief description of a patient's medical history, signs and symptoms, and tests results (example case shown in Box). All cases were based on real patients with a confirmed diagnosis. They were prepared by experts in internal medicine and used in previous studies with internal medicine residents.[16,17] The cases were presented to participants in a booklet (one for each phase), in a random sequence.

In Phase 1, each case had a diagnosis listed and participants had to rate the likelihood (as percentage) that the indicated diagnosis was correct. The provided diagnosis was always correct, but participants were not aware of this, nor did they receive feedback on their likelihood ratings. This phase consisted of 6 cases: 4 neutral cases and 2 cases of diseases that have signs and symptoms also frequently encountered in 2 other diseases presented in Phase 2 (Table 1). For example, a patient with cirrhosis or primary sclerosing cholangitis (Phase 2) may present with signs and symptoms similar to acute viral hepatitis (Phase 1). To minimize potential influence of case specificity or difficulty, we used two booklets with different sets of cases in Phase 1; participants randomly received either Set 1 or Set 2. In each set, the similar cases in Phase 2 had no relationship to the Phase 1 cases in the alternate set.

In Phase 2, all participants were asked to diagnose 8 new cases (the same for all participants), doing their best to provide an accurate diagnosis as quickly as possible. This procedure aimed at inducing non-analytical reasoning based on pattern-recognition, minimizing the chances

that participants engage in elaborate analysis of case findings. The cases were presented in random order in a second booklet, and participants were reminded with each case to read the case description and then immediately write down the most likely diagnosis for the case. Four of the cases were similar to 2 cases seen in Phase 1 by participants working with Set 1, and the other 4 were similar to 2 cases seen in Phase 1 by participants working with Set 2 (Table 1). If the availability bias occurs, the diagnosis of the cases encountered in Phase 1 should more promptly and frequently come to mind when participants encounter the cases with similar signs and symptoms in Phase 2 than when they had not encountered these cases in Phase 1. For example, participants working with Set 1 in Phase 1 would be expected to erroneously give a diagnosis of acute viral hepatitis to the cases of liver cirrhosis and primary sclerosis cholangitis more frequently than participants who worked with Set 2 in Phase 1.

In Phase 3, participants were asked to again diagnose the 4 cases from Phase 2 that could have been influenced by previous exposure to similar cases (Table 1). They followed instructions aimed at inducing reflective reasoning: (1) read the case; (2) write down the diagnosis previously given for the case; (3) list the findings in the case description that support this diagnosis; (4) list the findings that speak against this diagnosis; (5) list the findings that would be expected to be present if this diagnosis were true but that were not described in the case. Participants were subsequently asked to list alternative diagnoses assuming that the initial diagnosis generated for the case had proved to be incorrect, and to follow the same procedure (steps 3-5) for each alternative diagnosis. Finally, they were asked to draw a conclusion by ranking the diagnoses in order of likelihood and selecting their final diagnosis for the case.

Data analysis

All cases had a confirmed diagnosis that was used as a standard to evaluate the accuracy of the diagnoses provided by the participants. Two experts in internal medicine (JS and CG) independently assessed the diagnoses blinded to the experimental conditions under which they were provided. The diagnoses were evaluated as fully correct, partially correct, or incorrect, scored as 1, 0.5, or 0 points, respectively. A diagnosis was considered fully correct whenever the core diagnosis was cited by the participant and partially correct when the core diagnosis was not mentioned but a constituent element of the diagnosis was cited. For example, in the case in the Box, "celiac disease" was scored as correct, and "malabsorption" as partially correct.

For each participant, we separately summed the scores obtained in Phase 2 on the 4 cases that had similarities to the cases encountered in Phase 1, and the 4 cases that did not. For Phase 3, the diagnostic scores obtained on the 4 cases were summed for each participant.

For Phase 2, an ANOVA with "years of training" as between-subjects factor (first vs second year) and "recent experiences with similar cases" as within-subjects factor (with vs without) was conducted on the mean diagnostic performance scores obtained through non-analytical reasoning on both types of cases (similar to cases seen in Phase 1 or not). This analysis tested the hypothesis that recent experiences with similar cases would generate an availability bias

and that this bias would be larger for more experienced (second-year) residents. Post-hoc paired t-tests were performed to compare the diagnostic performance of first- and second-year residents under the two experimental conditions. To assess whether the diagnoses of the cases encountered in Phase 1 were indeed provided as diagnosis of the similar cases in Phase 2, we computed the number of times the diagnoses of cases in Phase 1 were mentioned by participants in Phase 2 who had seen similar cases in Phase 1 vs those who had not, and conducted paired t-tests on these data for the first- and second-year residents.

A second ANOVA with "years of training" as a between-subjects factor (first year vs second year) and "type of reasoning" as a within-subjects factor (non-analytical vs. reflective) was conducted on the mean diagnostic performance scores in Phase 2 and Phase 3. This analysis tested the hypothesis that reflection (Phase 3) could counteract the availability bias by improving the diagnostic performance scores compared to those obtained on the same cases through non-analytical reasoning (Phase 2).

Significance was set at $p < 0.05$ for all comparisons (2-tailed). SPSS 15.0 for Windows was used for the statistical analyses.

RESULTS

Non-analytic reasoning (Phase 2)

Table 2 presents the mean diagnostic accuracy scores obtained by first-year and second-year residents when cases were solved through non-analytical reasoning (Phase 2). The ANOVA showed no significant main effects, but there was a significant interaction effect between "years

Table 2. First and second year residents' mean diagnostic accuracy scores (range from 0 – 4) in Phase 2 (non-analytical diagnostic reasoning) as a function of previous exposure to similar cases in Phase 1

	1st year of training (n=18)	2nd year of training (n=18)
	Mean (95% CI)	Mean (95% CI)
Diagnostic accuracy on the cases similar to those previously encountered	2.03 (1.55 – 2.51)	1.55 (1.14 – 1.96)
Diagnostic accuracy on the other cases	1.42 (0.92 – 1.92)	2.19 (1.73 – 2.66)
P-value	$p = 0.046$[a]	$p = 0.03$[a]

[a] Comparison of accuracy on similar vs other cases. 2-sided t-test

of training" and "recent experiences with similar cases" ($F(1, 34) = 10.35$, $MSE = 0.68$, $p = 0.003$, $\eta_p^2 = 0.23$). Mean scores for the second-year residents were consistent with an availability bias. They obtained significantly lower diagnostic scores on the cases similar to those encountered in Phase 1 than the other cases (on 0-4 scale, 1.55; 95% confidence interval [CI], 1.15-1.96 vs 2.19; 95% CI, 1.73-2.66; $p = 0.03$).

Among the 8 Phase 2 cases potentially similar to Phase 1, second-year residents more frequently gave the Phase 1 diagnosis when they had encountered the cases in Phase 1 compared with when they had not (mean frequency per resident, 1.44; 95% CI, 0.93 – 1.96; vs 0.72; 95% CI, 0.28 – 1.17; $p = 0.04$) (Table 3). Even when the participants had not encountered the similar cases in Phase 1, they sometimes incorrectly provided the Phase 1 diagnosis to the related cases, but this occurred less frequently than when they had been previously exposed to the Phase 1 cases.

In contrast, this pattern was not seen for the first-year residents, who had a higher score on the cases similar to those encountered in Phase 1 than on the other cases (Table 2). Having encountered a similar case in Phase 1 did not lead to more frequently giving this diagnosis in Phase 2 than when they had not seen a similar case (mean frequency per resident, 0.78; 95% CI, 0.34 - 1.26; vs 0.89; 95% CI, 0.47 – 1.30; $p = 0.67$) (Table 3).

Table 3. Persistence of Phase 1 Diagnoses Among 8 Potentially Similar Cases in Phase 2 and the 4 Similar Cases in Phase 3

	Training Year	
	1st year (n=18)	2nd year (n=18)
Phase 1 Diagnosis Given in Phase 2 Cases		
Frequency of Phase 1 diagnosis in Phase 2 cases, mean (95% CI) [% of all wrong diagnoses]		
Having encountered similar cases in Phase 1	0.78 (0.34 - 1.26) [42]	1.44 (0.93 – 1.96) [63]
Not having encountered similar cases in Phase 1	0.89 (0.47-1.30) [36]	0.72 (0.28 – 1.17) [42]
P-value	0.67[a]	0.04[a]
Correction of Phase 1 Diagnoses in Phase 3		
Number of Phase 1 diagnoses in Phase 2 cases	14	26
Corrected after reflection in Phase 3, No. (%)	5 (36)	8 (31)
Adhered to after reflection in Phase 3, No. (%)	7 (50)	16 (63)
Diagnosis changed incorrectly after reflection	2 (14)	2 (6)

Abbreviation: CI, confidence interval

[a] Comparison of Phase 1 diagnoses with vs without having encountered similar cases in Phase 1, 2-sided t-test.

Reflective reasoning (Phase 3 vs Phase 2)

The diagnostic scores obtained through reflective reasoning (Phase 3) on the cases similar to the diseases that had been encountered in Phase 1 (those cases subject to an availability bias in Phase 2) are presented in Table 4. A significant main effect of "type of reasoning" was found in

the ANOVA ($F(1,34) = 8.46$, $MSE = 0.30$, $p = 0.006$, $\eta_p^2 = 0.20$), indicating that reflection improved all participants' diagnoses compared to non-analytical reasoning. The percentage of Phase 1 diagnoses that were corrected or adhered to after reflection is shown in Table 3.

Table 4. First- and second-year residents' mean diagnostic accuracy scores (range from 0 – 4) in Phase 2 (non-analytical diagnostic reasoning) and Phase 3 (reflective reasoning) for the cases similar to those encountered in Phase 1

	Mean score (95% confidence interval)	
	1st year of training (n=18)	2nd year of training (n=18)
Non-analytical diagnostic reasoning (Phase 2)	2.03 (1.55 – 2.51)	1.55 (1.14 – 1.96)
Reflective diagnostic reasoning (Phase 3)	2.31 (1.89 – 2.73) [a]	2.03 (1.49 – 2.57) [a]

[a] There was a significant main effect of reflection, which improved diagnoses compared to non-analytical reasoning for both 1st and 2nd year residents (p=.006 by ANOVA)

COMMENT

This study demonstrated that an availability bias may indeed occur in response to recent experiences with similar clinical cases when a non-analytical mode of reasoning is used, yielding diagnostic errors, and that reflective reasoning may help counteract this bias. The results suggest that the occurrence and negative effects of availability bias are a function of the reasoning approach used and the expertise level.

Encountering only one case of a disease was sufficient to make the second-year residents more prone to incorrectly giving that diagnosis to subsequent cases of different, though similar, diseases. In emergency rooms and outpatient clinics, physicians are likely to see (often close in time) several patients with similar symptoms caused by different diseases. In many clinical settings, therefore, conditions propitious for the occurrence of the availability bias prevail.

Moreover, because reliance on non-analytical reasoning tends to increase with experience, it is possible that physicians with many years of clinical practice may be even more susceptible to availability bias than second-year residents, and this should be investigated. In real life situations, an initial incorrect hypothesis might be spontaneously revised before expensive or time-consuming tests are ordered. However, the effects of anchoring by an early incorrect diagnosis may still lead to inaccurate judgment and inappropriate decisions. More experienced clinicians appear to be more subject to an anchoring effect,[18] which makes it less likely that they will spontaneously overrule an incorrect initial diagnosis.

These findings contribute some insight into cognitive mechanisms underlying errors, which are the object of ongoing scientific debate.[19] Evidence of the availability bias emerged in Phase 2, when participants diagnosed the cases through a non-analytical reasoning mode, and this was in part repaired in Phase 3 by reflective reasoning. This suggests that the mistakes made in

Phase 2 did not derive from knowledge deficits. The residents who failed to correctly diagnose the cases through non-analytical reasoning may have arrived at the correct diagnoses after reflecting on the same cases by activating existing knowledge. Errors in Phase 2 were therefore more likely to have been provoked by bias in reasoning processes.

We had expected the availability bias to be larger for the more experienced residents because the tendency to diagnose cases through pattern recognition increases with clinical experience.[12-13] We had not expected, however, to find an opposite pattern for the first-year residents who had better performance on similar cases. It is possible to speculate on reasons for this finding, such as that these novice residents might have already used a more reflective mode of reasoning during the exposure phase (Phase 1), being less self-confident than their more experienced colleagues [20-21] and therefore perhaps less reliant on immediate decisions. They may not have had a sufficient amount of clinical experience to make extensive use of pattern-recognition, and had to rely on a more analytic approach that could have been activated by Phase 1 cases. However, as a post-hoc analysis yielding an unexpected finding, these are speculations that should only be interpreted as hypothesis generating.

Although reliance on non-analytical reasoning and heuristics such as availability work well in many situations, reducing the time and effort involved in decision-making and allowing physicians to make accurate diagnoses in routine situations,[19,22] it may open the door to cognitive bias. Reflection has been shown to improve diagnosis when problems are complex or nonroutine,[17,23] and this study indicates that reflection may also be a mechanism to counteract cognitive biases.

With respect to medical education, this study suggests that a relatively simple instructional procedure can be used to induce reflective reasoning and improve diagnostic accuracy. This procedure for reflective reasoning can be implemented relatively easily in educational situations. Further research should investigate the effects of this process on diagnostic reasoning in practice settings.

This study has several limitations. First, we investigated residents from two different years in the internal medicine residency program, and it is not clear whether the differences in the susceptibility to bias encountered in the study would persist in later years or occur in other specialties. Second, the test cases were presented immediately after the initial cases, and similar problems do not always come consecutively in real clinical practice. Third, there may be restrictions in generalizing these findings obtained under laboratory conditions to real life situations, which are always richer in cues that may facilitate intuitive judgments. However, we worked with cases based on real patients and with tasks that simulate medical decision making.

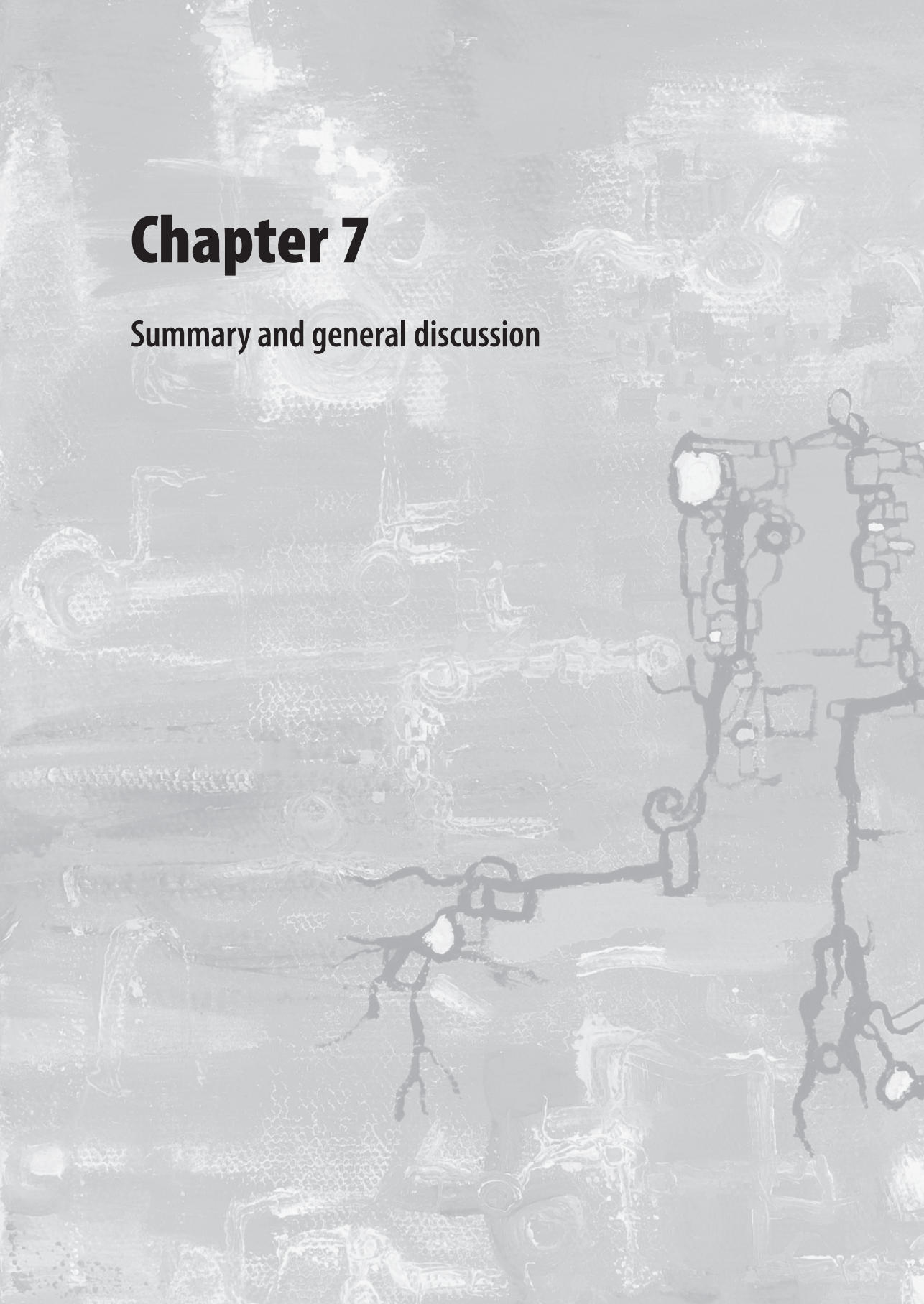
In summary, this study showed that the availability bias may occur in medical diagnosis as a consequence of recent experiences with similar cases under non-analytical reasoning conditions, and that susceptibility to this effect may be related to having more clinical experience. It provided further evidence that flaws in reasoning processes rather than knowledge gaps may underlie diagnostic errors, and showed the potential for repair by reflective reasoning.

REFERENCES

1. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med*. 2005;165:1493-1499.
2. Kohn KT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
3. Kahnemann D, Tversky A. *Judgment under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge Uni- versity Press 1982.
4. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med.* 2003;78:775-780.
5. Kahneman D. A perspective on judgment and choice: Mapping bounded rationality. *Am Psychol.* 2003;58:697-720.
6. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *BMJ* 2005;330:781–4.
7. Redelmeier DA. The cognitive psychology of missed diagnoses. *Ann Intern Med.* 2005; 142: 115-120.
8. Poses RM, Anthony M. Availability, wishful thinking, and physicians' diagnostic judgments for patients with suspected bacteremia. *Med Decis Making.* 1991;11(3):159-68.
9. Brezis M, Halpern-Reichert D, Schwaber MJ. Mass media-induced availability bias in the clinical suspicion of West Nile fever. *Ann Intern Med*. 2004;140(3): 234-5.
10. Heath L, Acklin M, Wiley K. Cognitive heuristics and AIDS risk assessment among physicians. *J Appl Soc Psychol*. 1991; 21: 1859-1867.
11. Peay MY, Peay ER. The evaluation of medical symptoms by patients and doctors. *J Behav Med.* 1988; 21(1): 57-81.
12. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev.* 1993;5:1-17.
13. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med*. 1990; 65: 611-621.
14. Evans JStBT. Dual-processing accounts of reasoning, judgment, and social cognition. *Ann Review Psychol 2008; 59:* 255-278.
15. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98–106.
16. Mamede SM, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: case ambiguity and shift to reflective approaches in clinical reasoning. *Med Educ*. 2007: 41: 1185–1192.
17. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ*. *2008;* 42: 468-475.
18. Eva KW. The aging physician: changes in cognitive processing and their impact on medical practice. *Acad Med.* 2002 Oct;77(10 Suppl):S1-6.
19. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ.* 2010;44:94-100.
20. Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? *J Gen Intern Med.* 2005;20:334–339.
21. Berner ES, Maisiak RS, Heudebert GR, Young KR Jr. Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. *AMIA Annu Symp Proc.* 2003; 2003:76 – 80.
22. Norman GR. Dual processing and diagnostic errors. *Adv Health Sci Educ Theory Pract.* 2009; 14: 37-49.
23. Mamede SM, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. *Acad Med*. 2008; 83:1210–1216.

Chapter 7

Summary and general discussion



GENERAL DISCUSSION

The preceding chapters reported empirical studies on the potential of two cognitive factors: confirmatory tendencies and availability bias, to cause errors in medical diagnosis. The background of the study of cognitive diagnostic errors, mistakes resulting from faults in physicians' thinking processes, is sketched in **Chapter 1**. First, medical errors are believed to cause preventable deaths. Although the majority of mistakes result from treatment-errors, diagnostic mistakes are sizeable contributors, especially within the domains of internal medicine and emergency medicine.[1-3] Cognitive diagnostic errors are observed in up to 75% of all cases of diagnostic mistakes,[4] however, empirical data to demonstrate causal relationships have been scarce. Second, in this chapter the theoretical frameworks that provide the backbone for conducting the studies: medical expertise research, bias, and dual-process-theory, were discussed.[5-12] Finally, a potential strategy to correct cognitive mistakes, i.e., reflective reasoning, was introduced.[13,14]

Furthermore, in **Chapter 1** the hypotheses that were to be tested in this thesis were presented. The main hypothesis was (1) cognitive factors have the potential to cause diagnostic errors in internal medicine. This hypothesis was explored using several more specific hypotheses that were tested in the two parts. In part 1, we focused on physicians' confirmatory tendencies and hypothesized that (2) suggested diagnoses on written clinical cases may evoke confirmatory tendencies in clinical reasoning, which could lead to diagnostic mistakes. Next, we hypothesized that (3) confirmatory tendencies due to diagnostic suggestions, may be mediated by a focus on case-features supportive of the suggestions. In addition, it was hypothesized that (4) encountering a number of consistent diagnostic suggestions would influence acceptation rate of subsequently presented suggestions. Besides exploring confirmatory tendencies, part 2 of the thesis was dedicated to the investigation of another cognitive factor, availability bias. We hypothesized (5) this bias to be able to induce errors in medical diagnosis. In addition, it was hypothesized that (6) reflective reasoning would lead to a correction of diagnostic errors resulting from the availability bias. Finally, we hypothesized that (7) more experienced participants would be more influenced by availability bias than less experienced participants.

The subsequent chapters delved into these issues. **Chapters 2**, **3**, and **4** reported studies aimed at testing the first four hypotheses. The studies described in **Chapters 5** and **6** focused on the first, and on the final three hypotheses. In this synopsis, I go back to the initial hypotheses and summarize the findings. This recapitulation is followed by a discussion of the practical and educational implications of the research presented in this thesis. Finally, I present some suggestions for future research.

Summary of main findings

The relationship between cognitive factors and diagnostic errors in internal medicine was explored in all the studies that are reported in this thesis. The main goals were to study confirmatory tendencies in clinical reasoning and to investigate the effect of unrelated available

information on physicians' diagnostic accuracy. The choice for these two aims is based upon several papers that mark the two related pitfalls, confirmation bias and availability bias, as the most important cognitive errors in medicine.[15-17]

Confirmation bias, the tendency to seek and interpret data to confirm, rather than refute hypotheses, is considered by Evans as "the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning."[18] Despite an increasing interest in the influence of bias on clinical reasoning, the only empirical work related to confirmatory tendencies within the medical domain was limited to studies that investigated the effect of diagnostic suggestions on physicians' diagnostic decisions using visual materials.[19-21] Besides an emphasis on visual stimuli, which might open space for noticing features that would not be described in a written case-description, the suggested diagnoses in those studies were corroborated by a supportive history that most likely influenced participants' diagnostic decisions, which showed a preference to confirm the suggestions. The outcome of these studies influenced the studies on confirmatory tendencies in this thesis, where, instead of visual materials, written cases were used, which were based on real patients with a verified diagnosis. In addition, the suggestions were not corroborated by a supportive history, and the setting for conducting the studies was the domain of internal medicine, a specialty often associated with cognitive errors.[1-3]

The study described in **Chapter 2** investigated the influence of diagnostic suggestions on physicians' diagnostic decisions using written internal medicine cases.[22] It was hypothesized that physicians would tend to go along with the suggestions and, therefore, they would have more difficulty rejecting incorrect diagnoses than accepting correct diagnoses.[23] The study was conducted in a single session with 24 internal medicine residents from an academic hospital in the Netherlands who evaluated diagnostic suggestions on six cases. The cases were the same and were presented in the same order for all participants, but the diagnostic suggestions that preceded them differed: half of the participants First evaluated a correct suggested diagnosis, the other half an incorrect suggested diagnosis, after which the correctness of the suggested diagnoses alternated. Results showed that participants' diagnostic score on the three cases with a correct suggested diagnosis ($M = 2.21$, $SD = 0.88$) was significantly higher than that on the three cases with an incorrect suggested diagnosis ($M = 1.42$, $SD = 0.97$). Therefore, physicians indeed found it easier to accept correct diagnoses than to reject incorrect diagnoses, $t(23) = 2.74$, $p < 0.05$, $d = 0.85$. This result provides support for the first and second hypothesis, and indicates that a diagnostic suggestion may indeed evoke confirmatory tendencies. Consequently, this tendency may lead to diagnostic errors, if the suggestion happens to be incorrect.

Although the study, described in **Chapter 2**, demonstrated the potential of diagnostic suggestions to evoke confirmatory tendencies during physicians' evaluation of written cases, it remained unclear *how* diagnostic suggestions might persuade clinicians to confirm the suggestions. We hypothesized, based on studies with visual materials,[20,21,24] that the tendency to confirm diagnostic suggestions might be mediated by a focus on features in the case descriptions

that support the suggestions. To test this hypothesis, the experiment described in **Chapter 3** was designed. In addition, we examined whether more features supportive of incorrect suggestions would be reported, when incorrect suggestions were accepted, than when they were rejected, because such erroneous conclusion might be caused by the tendency to gather data in line with the suggestions. In this study, 38 internal medicine residents from an academic hospital in the Netherlands evaluated suggested diagnoses on four written clinical cases. After evaluating each suggestion, participants listed the case-features they would report to their supervisor when they would discuss the patient for further management. All participants evaluated the same four cases, two with a correct suggested diagnosis, and two with an incorrect suggested diagnosis. Both the suggested diagnoses and the order of presentation of the cases differed between participants, who were prohibited from checking the case descriptions while writing down the features. Results corroborated the second hypothesis by replication of the finding from the study in **Chapter 2**, with the diagnostic score on cases with a correct suggestion (M = 1.13, SD = 0.70) being significantly higher than on cases with an incorrect suggested diagnosis (M = 0.63, SD = 0.59), $t(37) = 2.84$, $p = 0.007$, $d = 0.77$. This indicates that participants were generally inclined towards accepting the suggestions even when they were incorrect. Furthermore, a significantly higher percentage of features supporting the suggested diagnoses (M = 39.05%, SD = 12.70%) was reported than features supportive of the alternative diagnosis (M = 25.25%, SD = 13.33%), $t(37) = 4.40$, $p < 0.001$, $d = 1.06$. This reporting was irrespective of the correctness of participants' diagnostic decisions. Moreover, when participants accepted incorrect suggestions, the percentage of reported features supporting these incorrect suggestions (M = 50.10%, SD = 23.52%) was higher than when participants rejected those suggestions (M = 36.85%, SD = 26.16%), meaning their erroneous decisions may have been caused by the tendency to gather data in line with the suggestion, $t(49) = 1.88$, $p = 0.03$, $d = 0.53$. This indicates that the tendency to accept a diagnostic suggestion for written cases might indeed be mediated by a search that draws out features preferentially in favor of the suggested diagnosis - possibly leading to errors - thereby corroborating the first and third hypothesis.

The study described in **Chapter 4** took into account that clinicians in clinical practice will mainly encounter correct suggestions before they are confronted with an incorrect suggestion. This experiment investigated physicians' diagnostic performance if they would first encounter a number of correct suggestions followed by a number of incorrect suggestions, and vice versa. It was hypothesized that more incorrect suggestions would be accepted if participants had first evaluated a series of correct suggestions.[25] On the other hand, it was hypothesized that, if participants had first evaluated incorrect suggestions, they would become more critical about the suggestions, and hence will become more inclined to reject correct suggestions.[26,27] In other words, encountering a consistent number of diagnostic suggestions would influence acceptation of subsequently presented suggestions. In this study, 38 internal medicine residents evaluated suggested diagnoses on 8 written clinical cases. Four of those suggested diagnoses were correct, and four were incorrect. Half of the participants first evaluated four correct suggestions

and then evaluated four incorrect suggestions (C/I-condition). The other half started with the four incorrect suggestions followed by the correct suggestions (I/C-condition). The results showed that there was no significant main effect of case order and phase. Evaluation score in the C/I condition ($M = 2.87$, $MSE = 0.14$) equalled that in the I/C condition ($M = 2.66$, $MSE = 0.14$), $F(1,36) = 1.09$, $p = 0.30$, *ns*. Evaluation score on the first four cases ($M = 2.68$, $MSE = 0.14$) equalled that score on the last four cases ($M = 2.84$, $MSE = 0.17$), $F(1,36) < 1$, meaning that consistency in preceding suggested diagnoses did not influence the tendency to accept subsequent diagnostic suggestions. However, a significant interaction effect between case order and phase, $F(1,36) = 11.82$, $p = 0.001$, $\eta_p^2 = 0.25$, demonstrated that, in both conditions, the score on cases with correct suggestions was higher than the score on cases with incorrect suggestions. These findings indicate that consistency in preceding correct or incorrect diagnostic suggestions does not influence the tendency to accept or reject subsequent suggestions, thereby refuting the fourth hypothesis. However, overall physicians still showed a tendency to accept diagnostic suggestions, possibly leading to diagnostic errors if an incorrect suggestion is encountered.

The conclusion of the studies described in part 1 of this thesis is that suggested diagnoses on written clinical cases may indeed evoke confirmatory tendencies in physicians, which, if the suggestion happens to be incorrect, may lead to error. This tendency seems to be mediated by a focus on features that support the suggestion but does not seem to depend on consistency (i.e., correctness or incorrectness) of preceding diagnostic suggestions.

Part 2 of the thesis consists of two studies on the potential of availability bias to cause diagnostic errors. This bias arises when information from an unrelated source influences people's judgment and decisions.[28,29] In medicine a physician is subject to this bias, if information from an unlinked source influences diagnosis and causes a diagnostic mistake.[12-14] The study in **Chapter 5** described an experiment that investigated whether availability bias could impair medical diagnosis due to exposure to media-distributed information. In addition, it tested whether reflective reasoning could help to repair diagnostic errors resulting from this bias. In this three-phase experiment, 38 internal medicine residents from four teaching hospitals, were randomly assigned to read the Wikipedia entry about one of two diseases (either Legionnaires' disease or Q fever) during phase one. In phase two, which was conducted six hours later, as part of a seemingly unrelated study, they diagnosed eight clinical cases; two of which had superficial similarities to the disease in the Wikipedia entry they had read (bias expected), and two of which resembled the other disease they had not read about (bias not expected). Then, in phase 3, they again diagnosed the cases subject to bias, using structured reflection. Results showed that participants provided a higher number of wrong diagnoses of Q fever or Legionnaires' disease when they had read the Wikipedia information on this disease ($M = 0.61$, $SD = 0.55$) than when they had not ($M = 0.29$, $SD = 0.46$), $t(37) = 3.14$, $p = 0.03$, $d = 0.63$. Indicating that the exposure to a Wikipedia entry indeed led to diagnostic error, with the average number of cases mistakenly diagnosed as the 'Wikipedia disease' being significantly higher than when those cases were diagnosed without previous exposure. This result corroborates the fifth hypothesis.

Furthermore, supporting the sixth hypothesis, reflection (phase 3) significantly improved diagnostic accuracy on those cases that were affected by availability bias in phase 2 (from M = 1.08, SD = 0.49 in phase 2 to M = 1.36, SD = 0.67 in phase 3), $t(37) = 2.52$, $p = 0.016$, $d = 0.48$. Interestingly, reflection increased performance to the same level as performance on the test cases that were not subject to bias in phase 2 (M = 1.34, SD = 0.66). This improvement was due to neutralization of the bias: after reflection, the number of biased diagnoses was significantly reduced (from M = 0.61, SD = 0.55 in phase 2 to M = 0.37, SD = 0.63 in Phase 3), $t(37) = 2.16$, p = 0.037, d = 0.41.

The three-phase study described in **Chapter 6**, investigated whether availability bias might arise when physicians diagnose cases that have clinical manifestations similar to those of recently encountered cases, through non-analytical reasoning and whether a reflective reasoning approach could counteract that bias. Moreover, because non-analytical reasoning develops in association with clinical experience,[5,6] we investigated whether there would be a difference in the degree of bias between residents in the first and second year of the residency program. In phase 1, 18 first-year and 18 second-year internal medicine residents evaluated diagnoses of six clinical cases. Next, they diagnosed eight different cases through non-analytical reasoning, four of which had findings similar to previously evaluated cases but different diagnoses. In phase 3, these four cases were subsequently diagnosed again through reflective reasoning. Results showed a significant interaction between "years of training" and "recent experiences with similar problems." That is, findings consistent with an availability bias occurred for the second-year residents, who scored lower on the cases similar to those previously encountered (1.55; 95% CI, 1.15-1.96) than on the other cases (2.19; 95% CI, 1.73-2.66; p = 0.03). This pattern was not seen among the first-year residents (2.03; 95% CI, 1.55-2.51 vs 1.42; 95% CI, 0.92-1.92; p = 0.046). This finding partly supports hypotheses five and seven. That is, we expected the size of the bias to be larger in more experienced participants but we did not anticipate finding the opposite pattern for the first-year residents. Second-year residents provided the phase 1 diagnosis more frequently for phase 2 cases they had previously encountered than for those they had not (mean frequency per resident, 1.44; 95% CI, 0.93-1.96 vs 0.72; 95% CI, 0.28-1.17; p = 0.04). Furthermore, a significant main effect of reasoning mode, which corroborates the sixth hypothesis, was found: reflection improved the diagnoses of the similar cases compared with non-analytical reasoning for the second-year residents (2.03; 95% CI, 1.49-2.57) and the first-year residents (2.31; 95% CI, 1.89-2.73; p = 0.006). These findings implicate that, when faced with cases similar to previous ones, and using non-analytical reasoning, second-year residents made errors consistent with the availability bias. Subsequent reflection on the cases subject to bias tended to neutralize it, leading to improved diagnostic accuracy in both first-year and second-year residents.[30]

The conclusions of the studies described in part 2 of the thesis are that availability bias indeed could impair medical diagnosis. However, this did not apply to first-year residents, who actually benefitted from seeing a lookalike case first. This result partly supports the fifth, and

seventh hypothesis because we had expected the bias to occur in both first- and second-year residents. In addition, corroborating the sixth hypothesis, reflection again seemed to have the potential to neutralize the bias's influence on diagnostic accuracy.

Implications of the findings

Cognitive diagnostic errors are believed to offer a sizeable contribution to medical error-related mortality. The debate on the cognitive origins of diagnostic errors has, however, been quite speculative.[31,32] The findings from the studies reported in this thesis, provide empirical evidence supportive of a role for cognitive factors as causes of diagnostic mistakes. In five experimental studies, internal medicine residents' confirmatory tendencies and their susceptibility for the availability bias, leading to errors, have been explored by using written case-descriptions based on real patients, and with tasks that simulate "flesh-and-blood medical decision making." That is, the used manipulations, diagnostic suggestions, recent encounters with similar cases, and media-distributed information, are all highly relevant to everyday medical decision making. Therefore, the findings of our studies are significant for the scientific debate on cognitive diagnostic errors, for medical education and for clinical practice.

Besides demonstrating effects that confirmatory tendencies and availability bias could have on physicians' diagnostic decisions and accuracy, we showed that the size of these effects can be substantial. For instance, in the study, described in **Chapter 3**, residents' diagnostic score on cases with incorrect diagnostic suggestions was 90% lower than that score on cases with correct diagnostic suggestions, and in **Chapter 6**, reading about illnesses in Wikipedia entries increased the number of cases mistakenly diagnosed as one of those diseases by 100%. Such effects are much larger than effects in corresponding psychological research among naïve participants.[33,34]

In addition to showing physicians' susceptibility to cognitive bias, the studies in **Chapter 3**, **4**, **5**, and **6** focused on potential mechanisms underlying those mistakes, a subject of debate.[15-17,31,32] For instance, in **Chapter 3**, a focus on features, supportive of the diagnostic suggestions was found to mediate residents' inclination towards accepting them. This role for case-features is in line with findings from previous studies that used visual materials, but extends this research in several ways.[19-21] That is, since we did not use visual materials but written case descriptions, the features were explicitly mentioned in the case descriptions and, therefore, less likely to be missed by participants than features shown in a picture or an ECG. Furthermore, images are probably more powerful in opening space for detecting features that would not be expressed in a written case description. In addition, the incentive for reporting features differed from previous studies, where participants were requested to write down all relevant features.[19,20] In the study described in **Chapter 3**, participants listed features they would actually report to their supervisor while discussing the patient prior to management. Not only resembles this approach professional interaction in everyday clinical decision making, it also illustrates the potential consequences for individual patients. That is, if a supervisor

accepts a resident's erroneous diagnosis because of biased presentation of features, the mistake could have serious consequences for the individual patient. A remedy for errors related to the feature-identification mediated tendency to accept diagnostic suggestions may be instructing residents to seriously try considering alternative hypotheses as much as possible. In this perspective, reflective reasoning, based on a structured reanalysis of case findings, might be helpful. Reflection has shown to be a promising approach to deal with complex cases,[14] and, in the studies described in **Chapter 5** and **6**, an effective way to counteract diagnostic errors caused by availability bias.[30] In line with this recommendation, supervising specialists should be aware that while discussing patients with residents, there is always a risk of confirmatory tendencies. Therefore, during these discussions, both supervisors and residents should remain attentive and reflective. Moreover, supervisors should require residents to think of alternative diagnoses in order to decrease the chance of diagnostic errors to occur. In addition, such discussions could contribute to a more resourceful educational climate.

The recommendation to stimulate a reflective approach among residents in order to avoid diagnostic errors suggests another mechanism underlying cognitive errors, therefore providing input for the scientific discussion. That is, mistakes due to availability bias were found to be associated with reliance on non-analytical reasoning (System-1 approach). Non-analytical reasoning is usually efficient in routine situations,[7] but, as we have shown, it is also liable to bias.[30] The subsequent application of an analytical method, structured reflection (System-2 approach), successfully forced residents to focus on all features before making a decision. Indeed, once their attention was focussed on the relevant features, they were more likely to make the correct decision. This implicates that in medical education and clinical practice, physicians should be encouraged to use both analytical and non-analytical reasoning, since non-analytical reasoning effectiveness in routine situations. However, when, for example, cases appear similar to often-encountered ones but have another diagnosis, this approach may fail and lead to errors.[31] The problem is that people are not capable of recognizing when this happens. That is, if they would notice, the bias would not occur. In other words, it is unlikely that physicians will know when a more analytical approach should have to be adopted. If analytical reasoning, for example by means of structured reflection, would always be used, the number of errors due to bias could potentially be reduced. However, it's feasibility in clinical practice, due to the time and effort required for this procedure is questionable. Although, if this process of structured reflection itself could somehow be automated through training, it could be rapidly executed. This implies that reflection may have to be taught and the question is then *how* to teach physicians to reflect effectively. In addition, another result of the study described in **Chapter 6**, the demonstration that experience does not protect from bias, poses another important question: *if* reflection protects physicians from cognitive bias and should be taught, *who* should it be taught to? That is, the results from the study described in **Chapter 6** indicate that the influence of bias might vary with experience.[30] This is in line with findings from medical expertise research, which indicated that more experienced physicians may indeed be

more vulnerable for bias, mainly because their heavier reliance on non-analytical reasoning.[5-7] Consequently, the relationship between clinical experience, reasoning modes, and susceptibility to bias, should be further explored.

Future research

The research presented in this thesis reflects a relatively novel line of investigation in medical expertise research. Although several years ago several position papers were published on the subject of cognitive diagnostic error,[15-17,35-37] experimental evidence, and therefore the possibility to make causal inferences, was scarce. The reported studies provide evidence supportive of a role for two cognitive factors as causes of diagnostic mistakes. Nevertheless, many aspects remain to be addressed in this domain. In fact, the studies naturally bring about some ideas for future research.

One area that might be explored in future studies applies to the weight that can be assigned to experiments in clinical reasoning, since results have been obtained in laboratory conditions. Therefore, it can be argued that in real life medical decision making, the potential of cognitive factors to cause diagnostic errors would be smaller. That is, it can be pictured that physicians might become more reflective or critical when the consequences of mistakes increase. More ecologically valid scenarios, for instance, with standardized patients involved might extend the findings from the controlled experiments that we have conducted. Nevertheless, we used tasks and manipulations that are relevant for actual medical decision making: diagnostic suggestions, encountering similar cases and the absorbing of information from the media are phenomena every physician is confronted with on a daily basis. However, boundaries should be pushed forward and, as a next step studies, with scenario trainings might offer interesting insights in medical decision making in a more naturalistic environment.

Furthermore, based on the studies described in **Chapter 5** and **6**, we argue reflection might be an remedy for cognitive diagnostic errors. The limited experience with this strategy, however, warrants cautiousness regarding the generalizability of this finding.[14,30,38] For example, based on the present results, we can not predict what will happen to physicians' tendency to accept diagnostic suggestions after application of reflection. This could be explored in a future study, and given the prominent role for case features in both reflection and the tendency to confirm diagnostic hypotheses, reflection might indeed be of help neutralizing errors due physicians' tendency to accept incorrect diagnostic suggestions. Nontheless, the studies in **Chapter 5** and **6** *do* suggest that reflection has the potential to neutralize errors due to availability bias. On the other hand, certainty whether or not reflection was the actual cause of curing residents from faults due to availability bias cannot be confidently claimed. That is, not only did we instruct the participants to reflect, they also spent more time and effort on the cases. This could have contributed to improved diagnostic accuracy in phase 3 of the studies described in **Chapter 5** and **6**. Certainly, analytical reasoning is by definition slower and more effortful,[8,9] and the reverse could, but does not have to be true per se. That is, investment of more time or effort

might or might not lead to correction or prevention of mistakes. We presume that the type of activities time and effort are invested in (i.e., that increased diagnostic accuracy resulting from our instructions to reflect is not so much due to the investment of extra time, as it is to the fact that participants are not only asked to generate alternative plausible diagnoses, but are also focussed on the features supporting those diagnoses) are vital for diagnostic succes. Indeed, findings from previous studies indicated that giving clinicians more time might not prevent mistakes.[15,17] Such research suggested that physicians often become anchored in their initial hypothesis, looking for confirming evidence to support their initial diagnosis, undervalueing evidence against it, and therefore failing to alter their first impression in light of all available information. Moreover, premature closure: the failure of considering reasonable alternatives after an initial diagnosis is made, was pointed out as the most frequent cause of cognitive errors in the study by Graber and colleagues.[4] Despite the fact that we cannot claim that our instructions to reflect actually caused improved diagnostic accuracy, our findings *do* demonstrate that a more analytical reasoning approach might be an effective way to reinstate diagnostic performance. To gain further insight in the cognitive mechanism underlying reflection, a future study could, for example, ad a condition to control for the time spent on the reflective task.

Although the relationship between experience and reasoning modes has been established,[5-7] the finding described in **Chapter 6**, that more experienced physicians were subject to bias might seem counterintuitive. In this chapter, we investigated first- and second-year residents in internal medicine, and whether the differences in the susceptibility to bias encountered in the study would persist in later years or occur in other specialties is still unclear. In addition, the results suggested that first-year residents were not biased. Although we had expected a larger availability bias for the more experienced residents because the tendency to diagnose cases through pattern recognition increases with clinical experience,[5,6] we had not expected to find an opposite pattern for the first-year residents who had *better* performance on similar cases.We can only conjecture about explanations for this finding. For instance, it could be that first-year residents might have already used a more reflective mode of reasoning during the exposure phase (Phase 1), being less self-confident than second-year residents,[39,40] and therefore maybe less reliant on instant decisions. They may not have had sufficient clinical experience to make extensive use of pattern-recognition, and had to rely on a more analytic approach that could have been activated by the phase 1 cases. However, as this concerns an unpredicted result, these speculations should only be used as inspiration for future studies. In addition, the role of experience has not been addressed in **Chapters 2**, **3**, and **4**. And although some studies suggest that more experienced specialists are also influenced by diagnostic suggestions,[7,20,41] studies focusing on this theme with written cases might be helpful.

If then, experience alone does not protect physicians from cognitive bias, what should be done? An interesting option might be to explore whether or not the teaching of reflective reasoning is possible. Previous studies on the effect of reflection on diagnostic performance[14,30,38] have employed a structured procedure to guide reflection upon clinical cases, which has been

shown to improve diagnostic accuracy at least when problems are complex. Although effective as an experimental paradigm, such a highly structured procedure is unlikely to be feasible in real life. It is worthwhile to explore whether less structured and less guided forms of reflection would also lead to better diagnoses.

Finally, other cognitive biases, many of which are well described for the medical domain,[15-17] should be explored, possible mechanisms of action should be disentangled in order to strengthen physicians' defenses against cognitive errors. If we are determined to push the limits, we might one day be able to eliminate cognitive diagnostic errors and save patients' lives.

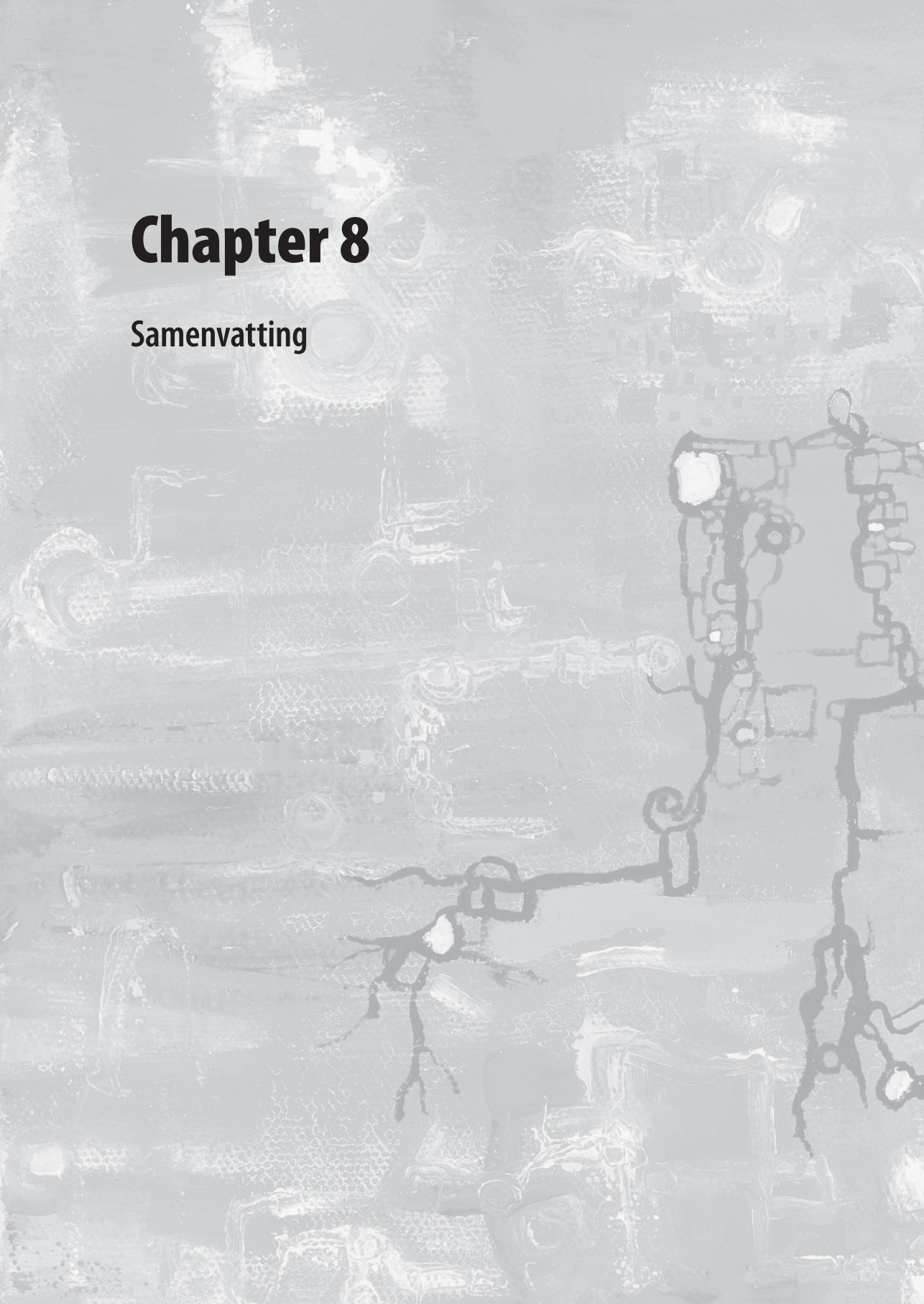
REFERENCES

1. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study 1. *New Engl J Med* 1991;324:370-6.
2. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;38:261-71.
3. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Australia* 1995;163:458-71.
4. Graber ML, Franklin N,. Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493-9.
5. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise - Theory and implications. *Acad Med* 1990;65:611-21.
6. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev* 1993;5:205-21.
7. Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ* 1997;2:173-84.
8. Evans JSBT. In two minds: Dual-process accounts of reasoning. *Trends Cogn Sci* 2003;10:454-59.
9. Evans JSBT. Dual processing accounts of reasoning, judgment and social cognition. *Ann Rev Psychol* 2008;59:255-78.
10. Evans JSBT. Bias in human reasoning: causes and consequences. NJ, Erlbaum; 1989.
11. Nickerson RS. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev Gen Psychol* 1998;2:175-220.
12. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185:1124-31.
13. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ* 2004;38:1302-8.
14. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ* 2008;42:468-75.
15. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775-80.
16. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *Brit Med J* 2005;330:781-4.
17. Redelmeier DA. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115-20.
18. Evans JSBT. Bias in human reasoning: causes and consequences. NJ, Erlbaum; 1989, p. 41.
19. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med* 2001;76:S18-20.
20. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med* 2002;77:S67-9.
21. Hatala RM, Norman GR, Brooks LR. The impact of a clinical scenario upon electrocardiogram interpretation. J Gen Intern Med. 1999;14:126–129.
22. Van den Berge K, Mamede S, Van Gog T, et al. Accepting Diagnostic Suggestions by Residents: A Potential Cause of Diagnostic Error in Medicine. *Teach Learn Med* In press, 2012.
23. Pyszczynski T, Greenberg J. Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis testing model. *Adv Exp Soc Psychol* New York: Academic Press; 1987:297-340.
24. Norman GR, Brooks LR, Coblentz CK, Babcook CJ. The correlation of feature identification and category judgments in diagnostic radiology. Mem Cognition. 1992;4:344–355.
25. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008;121:S2-S23.

26. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ* 2004;38:1302-8.
27. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: Case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ* 2007;41:1185-92.
28. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185:1124-31.
29. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychol* 1973;5:207-32.
30. Mamede S, van Gog T, van den Berge K, Rikers RMJP, van Saase JLCM, van Guldener C et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;304:1198-1203.
31. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* 2010;44:94-100.
32. Norman GR. Dual-processing and diagnostic errors. *Adv Health Sci Educ* 2009;14:37-49.
33. Schwarz N, Bless H, Strack F, Klumpp G, Rittenauer-Schatka H, Simons A. Ease of retrieval as information: Another look at the availability heuristic. *J Pers Soc Psychol* 1991;61:195–202.
34. Schwarz N. When thinking feels difficult: meta-cognitive experiences in judgment and decision making. *Med Decis Making* 2005;25:105-112
35. Bornstein BH, Emler AC. Rationality in medical decision making: A review of the literature on doctors' decision making biases. *J Eval Clin Pract* 2001;7(2):97-107.
36. Dawson NV, Arkes HR. Systematic errors in medical decision making: Judgment limitations. *J Gen Intern Med* 1987;2:183-7.
37. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *Brit Med J* 2002;324:729-32.
38. Mamede S, Schmidt HG, Rikers RMJP, Custers EJFM, Splinter TAW, Van Saase JLCM. Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychol Res* 2010;74:586-592.
39. Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? *J Gen Intern Med.* 2005;20:334–339.
40. Berner ES, Maisiak RS, Heudebert GR, Young KR Jr. Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. *AMIA Annu Symp Proc.* 2003; 2003:76 – 80.
41. Berbaum KS, Franken EA, Dorfman DD, et al. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Invest Radiol. 1986;21:532–539.

Chapter 8

Samenvatting



SAMENVATTING

Het doel van dit proefschrift is experimenteel te onderzoeken of en hoe de cognitieve factoren bevestigingsbias (confirmation bias) en beschikbaarheidbias (availability bias) tot diagnostische fouten kunnen leiden bij artsen die werkzaam zijn binnen de interne geneeskunde. Bevestigingsbias is gedefinieerd als de neiging om hypotheses, in het medisch domein diagnoses, te bevestigen en beschikbaarheidbias verwijst naar de neiging -diagnostische- beslissingen te baseren op gegevens die gemakkelijk voor de geest te halen zijn maar niet noodzakelijkerwijs betrekking hebben op de te analyseren situatie. De achtergrond van de studie van cognitieve diagnostische fouten wordt geschetst in **Hoofdstuk 1**. Medische fouten worden verantwoordelijk geacht voor vermijdbare sterfte, ongemak voor patiënten en hoge kosten. Hoewel de meerderheid van de medische fouten bestaat uit behandelfouten, leveren diagnostische fouten ook een substantiële bijdrage aan vermijdbare mortaliteit. Vooral de disciplines interne geneeskunde en acute geneeskunde zijn vatbaar voor diagnostische fouten.[1-3] Uit observationeel onderzoek blijkt dat cognitieve factoren een rol zouden kunnen spelen in ongeveer 75% van alle gevallen van diagnostische fouten.[4] Veelal zijn deze fouten geen gevolg van een gebrek aan de juiste kennis, maar van gebreken in het redeneerproces.[4] Het is gesuggereerd dat cognitieve bias zoals de bovengenoemde bevestiging- en beschikbaarheidbias, daar een belangrijke rol in spelen,[5-7] maar experimenteel onderzoek op dit gebied is schaars. In dit proefschrift worden experimentele studies gedaan naar deze bias, meer specifiek worden de volgende hypotheses getoetst. De overkoepelende hypothese luidt (1) dat cognitieve factoren kunnen leiden tot diagnostische fouten binnen de interne geneeskunde. Om deze hypothese te exploreren worden verscheidene meer specifieke hypotheses getest in de twee delen van dit proefschrift. In het eerste deel richt ik me op de neiging van artsen om een diagnostische suggestie te bevestigen. Ik onderzoek of (2) artsen de neiging hebben om een diagnostische suggestie te bevestigen, wat zou kunnen leiden tot een diagnostische fout als deze suggestie incorrect is. Vervolgens ga ik na of (3) de neiging tot bevestiging van een diagnostische suggestie gemedieerd kan worden door een focus op casuskarakteristieken die de suggestie ondersteunen. De laatste hypothese van deel 1 van dit proefschrift luidt dat (4) de neiging tot acceptatie van een diagnostische suggestie beïnvloed zou kunnen worden door consistentie in de correctheid van voorafgaande diagnostische suggesties.

Naast het exploreren van de neiging tot het bevestigen van diagnostische suggesties, toets ik in deel 2 van dit proefschrift of (5) beschikbaarheidbias kan leiden tot diagnostische fouten tijdens niet-analytisch klinisch redeneren. Bovendien ga ik na of (6) reflectief redeneren kan leiden tot correctie van diagnostische fouten veroorzaakt door beschikbaarheidbias. Tenslotte onderzoek ik of (7) meer ervaren artsen sterker beïnvloed worden door beschikbaarheidbias dan hun minder ervaren collegae omdat niet-analytisch redeneren zich parallel ontwikkelt met toename van klinische ervaring.

In **Hoofdstuk 2**, **3** en **4** rapporteer ik vervolgens de experimentele studies waarin de eerste vier hypotheses worden getoetst. In **Hoofdstuk 5** en **6** beschrijf ik de studies die gericht zijn op het toetsen van de eerste hypothese en de laatste drie hypotheses. In deze samenvatting volgen de belangrijkste bevindingen ten aanzien van elke hypothese.

Hoofdstuk 2 – 4: bevestigingsbias

Bevestigingsbias, de neiging om gegevens te zoeken en te interpreteren om hypotheses te bevestigen in plaats van ze te verwerpen, wordt gezien als een belangrijke redeneerfout.[8] Ondanks een toegenomen interesse in de invloed van bias op klinisch redeneren[5-7] zijn er slechts enkele empirische studies met betrekking tot bevestigingsbias binnen het medische domein en deze beperkten zich tot het onderzoeken van het effect van diagnostische suggesties op diagnostische beslissingen ten aanzien van *visuele* stimuli.[9-11] Een nadeel van het gebruik van visuele stimuli is echter dat ze gelegenheid kunnen bieden tot het identificeren van patiëntenkarakteristieken die niet zouden worden beschreven in geschreven casuïstiek. Daarnaast werd de correctheid van de diagnostische suggesties in deze studies aannemelijk gemaakt door ondersteunende casusbeschrijvingen voorafgaand aan de presentatie van de visuele stimulus. Met andere woorden, de presentatie van de diagnostische suggestie was in deze studies niet neutraal. Deze twee zaken hebben zeer waarschijnlijk de diagnostische beslissingen van de deelnemers in deze studies, die neigden tot het bevestigen van de diagnostische suggesties, beïnvloed. De studies in dit proefschrift bouwen voort op deze eerdere studies, met enkele belangrijke verschillen: In de studies beschreven in dit proefschrift is gebruik gemaakt maar van casusbeschrijvingen, gebaseerd op echte patiënten met een geverifieerde diagnose. Bovendien werden de diagnostische suggesties niet ondersteund door ondersteunende inleidende casusbeschrijvingen en zijn de studies uitgevoerd binnen de interne geneeskunde; een domein dat zeer vatbaar wordt geacht voor cognitieve diagnostische fouten.[1-3]

De studie die wordt beschreven in **Hoofdstuk 2,** onderzocht de invloed van een diagnostische suggestie op diagnostische beslissingen van internisten in opleiding.[12] Wanneer de hypothese correct is dat deze artsen de neiging hebben om suggesties te bevestigen, zullen ze meer moeite hebben om een incorrecte suggestie te verwerpen dan een correcte suggestie te bevestigen.[13] De studie werd uitgevoerd met 24 internisten in opleiding, die diagnostische suggesties evalueerden voor zes verschillende casus. Alle casus waren hetzelfde en werden in dezelfde volgorde gepresenteerd voor alle deelnemers, enkel de diagnostische suggesties die aan de casus voorafgingen verschilden tussen de deelnemers. De helft van de deelnemers evalueerde eerst een correcte suggestie, de andere helft eerst een incorrecte, waarna de correctheid van de suggesties afwisselde. Uit de resultaten blijkt dat de diagnostische score van de deelnemers op de drie casus met een correcte suggestie significant hoger is dan die score op de drie casus die voorafgegaan werden door incorrecte diagnostische suggesties. Hieruit blijkt dat internisten in opleiding het inderdaad lastiger vinden om een incorrecte diagnose te verwerpen dan een correcte te accepteren. Dit resultaat ondersteunt de eerste en

tweede hypothese en suggereert dat een diagnostische suggestie inderdaad een neiging tot bevestiging kan uitlokken, met als mogelijk gevolg dat, indien deze suggestie incorrect is, een diagnostische fout wordt begaan.

Hoewel de studie die wordt beschreven in **Hoofdstuk 2** laat zien dat diagnostische suggesties de neiging tot bevestiging kunnen uitlokken, was het onduidelijk *hoe* dit gebeurt.

Ik veronderstelde, gebaseerd op studies waarbij gebruik werd gemaakt van visuele materialen,[10,11,14] dat de neiging om een diagnostische suggestie voor geschreven casuïstiek te bevestigen gemedieerd zou kunnen worden door een focus casuskarakteristieken die ondersteunend zijn voor de suggestie. Om te testen of deze hypothese ook juist is ten aanzien van geschreven casus zonder de suggestie-bevestigende voorgeschiedenis, is de studie die beschreven staat in **Hoofdstuk 3** uitgevoerd. De focus op ondersteunende karakteristieken is onderzocht door na te gaan of meer karakteristieken, ondersteunend voor een incorrecte suggestie, worden gerapporteerd wanneer deze wordt bevestigd dan wanneer deze wordt verworpen. Dit zou een indicatie zijn dat een foutieve conclusie veroorzaakt wordt door de neiging om gegevens te verzamelen die in overeenstemming zijn met de suggestie.

In deze studie evalueerden 38 internisten in opleiding diagnostische suggesties voor vier klinische casus. Na evaluatie van elke suggestie noteerden deze artsen de casuskarakteristieken die zij zouden rapporteren aan hun supervisor wanneer zij de patiënt zouden bespreken voor verder beleid. Alle deelnemers evalueerden dezelfde vier casus, twee met een incorrecte suggestie en twee met een correcte suggestie. Zowel de diagnostische suggesties als de volgorde van presentatie van de casus verschilden per deelnemer. De resultaten bevestigen de tweede hypothese door replicatie van de bevinding van **Hoofdstuk 2**, met significant hogere diagnostische scores op casus met een correcte suggestie dan op casus met een incorrecte suggestie. Dit geeft aan dat de deelnemende artsen in het algemeen neigden tot het accepteren van de diagnostische suggestie. Bovendien werd een significant hoger percentage casuskarakteristieken ondersteunend voor de suggesties, dan ondersteunend voor het alternatief gerapporteerd, ongeacht de correctheid van de diagnostische beslissingen van de deelnemers. Bovendien was, wanneer een incorrecte suggestie werd geaccepteerd, het percentage gerapporteerde casuskarakteristieken ondersteunend voor deze incorrecte suggestie groter dan wanneer zij werd verworpen. Dit zou kunnen betekenen dat deze foutieve beslissingen veroorzaakt worden door de neiging data te verzamelen overeenkomstig de suggestie. Met andere woorden, de neiging tot het accepteren van diagnostische suggesties voor geschreven casus kan gemedieerd worden door een zoektocht naar casuskarakteristieken die de suggestie ondersteunen, wat kan leiden tot diagnostische fouten. Deze bevindingen bevestigen zodoende de eerste en derde hypothese.

De studie die wordt beschreven in **Hoofdstuk 4** hield rekening met het feit dat artsen in de klinische praktijk over het algemeen waarschijnlijk verschillende correcte diagnostische suggesties krijgen voor ze een incorrecte suggestie krijgen. In dit experiment onderzocht ik wat er met de diagnostische prestatie van internisten in opleiding gebeurt als zij eerst een

aantal correcte suggesties beoordelen voor zij incorrecte suggesties tegenkomen en vice versa. Ik onderzocht of er meer incorrecte suggesties worden geaccepteerd als deelnemers ervoor een reeks correcte suggesties hadden geevalueerd.[15] Tegelijkertijd ging ik na of, deelnemers wanneer zij eerst incorrecte suggesties evalueerden, kritischer zouden worden ten opzichte van daaropvolgende suggesties waardoor zij meer zouden neigen tot het verwerpen van daaropvolgende correcte suggesties.[16,17] Met andere woorden, het evalueren van een aantal consistente suggesties zou de acceptatie van daaropvolgende suggesties kunnen beïnvloeden. In deze studie evalueerden 38 internisten in opleiding diagnostische suggesties voor acht casus. Vier van deze casus hadden een correcte diagnostische suggestie en vier een incorrecte. De helft van de deelnemers evalueerde eerst de vier correcte suggesties waarna zij de vier incorrecte suggesties evalueerden (C/I-conditie). De andere helft startte met de vier incorrecte suggesties en eindigde met de vier correcte suggesties (I/C-conditie). De resultaten laten geen hoofdeffect van casusvolgorde en fase zien. De score in de C/I-conditie is gelijk aan die score in de I/C conditie. De score op de eerste vier casus is gelijk aan de score op de laatste vier casus, wat betekent dat consistentie in voorafgaande diagnostische suggesties de neiging tot het accepteren van daaropvolgende suggesties niet heeft beïnvloed. Een significant interactie effect tussen casusvolgorde en fase laat zien dat in beide condities de score op casus met een correcte suggestie hoger is dan de score op casus met een incorrecte suggestie. Deze bevindingen impliceren, in tegenspraak met de vierde hypothese, dat consistentie in voorafgaande correcte of incorrecte diagnostische suggesties de neiging om daaropvolgende suggesties te accepteren niet beïnvloedt. Desalniettemin neigden de deelnemende artsen over het algemeen naar het accepteren van de suggesties, wat kan leiden tot diagnostische fouten als de suggestie niet correct is.

De conclusie van de studies die beschreven staan in het eerste deel van dit proefschrift luidt dat een diagnostische suggestie voor geschreven casuïstiek inderdaad een neiging tot bevestiging kan uitlokken bij internisten in opleiding en dat deze neiging kan leiden tot diagnostische fouten indien de suggestie incorrect is. Deze neiging lijkt te worden gemedieerd door een focus op casuskarakteristieken die de suggestie ondersteunen maar lijkt onafhankelijk te zijn van de consistentie in correctheid van voorafgaande diagnostische suggesties.

Hoofdstuk 5 – 6: beschikbaarheidbias

Het tweede deel van dit proefschrift bestaat uit twee studies waarin onderzocht werd of beschikbaarheidbias kan leiden tot diagnostische fouten. Deze bias treedt op wanneer beschikbaarheid van informatie van een ongerelateerde bron het menselijk beslissen of oordelen beïnvloedt.[18,19] Op het vlak van de geneeskunde wordt van deze bias gesproken als informatie van een ongerelateerde bron het diagnostisch proces zo beïnvloedt dat er een diagnostische fout ontstaat.[5-7]

De studie die beschreven wordt in **Hoofdstuk 5** onderzocht of beschikbaarheidbias door blootstelling aan media informatie kan leiden tot een verkeerde diagnose tijdens niet-analytisch

redeneren. Aanvullend werd onderzocht of reflectief diagnostisch redeneren behulpzaam kan zijn om deze misdiagnoses te corrigeren. In dit driefase experiment lazen 38 internisten in opleiding in fase 1 de Wikipedia pagina over een aandoening (Legionella of Q-koorts). Tijdens de tweede fase, die zes uur later werd uitgevoerd als onderdeel van een schijnbaar ongerelateerde studie, diagnosticeerden zij acht klinische casus. Twee van deze casus hadden oppervlakkige overeenkomsten met de aandoening waarover zij hadden gelezen op de Wikipedia pagina en twee met de aandoening waarover zij niet hadden gelezen. In fase 3 diagnosticeerden de deelnemers de casus waarin de bias tot uiting zou kunnen komen opnieuw, gebruikmakend van diagnostische reflectie. Uit de resultaten blijkt dat de deelnemers een groter aantal incorrecte diagnoses van Q-koorts of Legionella stelden wanneer zij de Wikipedia informatie over die aandoening hadden gelezen in vergelijking tot wanneer zij die informatie niet hadden gelezen. Dit betekent dat blootstelling aan een Wikipedia pagina inderdaad kan leiden tot een diagnostische fout door beschikbaarheidbias. Dit resultaat ondersteunt de vijfde hypothese. Bovendien, ondersteunend voor de zesde hypothese, leidde diagnostische reflectie (fase drie) tot een significante verbetering van diagnostische accuratesse op de casus die in fase twee waren aangetast door beschikbaarheidbias. Een interessante bevinding is dat reflectie er toe leidde dat de diagnostische prestatie tot hetzelfde prestatieniveau werd gebracht als de casus die geen onderwerp van bias waren in fase 2. Deze verbetering werd veroorzaakt door neutralisatie van de bias: na reflectie was het aantal door bias beïnvloede diagnoses significant gereduceerd.

Het driefase-experiment dat wordt beschreven in **Hoofdstuk 6** onderzocht of beschikbaarheidbias kan optreden wanneer artsen casus diagnosticeren, middels niet-analytisch redeneren, die klinische manifestaties hebben gelijkend op recent gediagnosticeerde casus. Tevens werd onderzocht of reflectie diagnostische fouten die ontstaan door deze bias kan corrigeren. Bovendien, omdat niet-analytisch redeneren zich parallel ontwikkelt aan klinische ervaring,[20,21] onderzocht ik of er een verschil is in de mate van bias tussen eerste en tweedejaars internisten in opleiding.

In fase 1 evalueerden 18 eerste en 18 tweedejaars internisten in opleiding diagnoses van zes klinische casus. Vervolgens diagnosticeerden zij acht andere casus door middel van niet-analytisch redeneren. Vier van deze casus hadden vergelijkbare casuskarakteristieken met eerder geëvalueerde casus, maar met andere diagnoses. In fase 3 werden deze vier casus opnieuw gediagnosticeerd door middel van reflectief redeneren. De resultaten lieten een significante interactie tussen jaren in opleiding en recente ervaringen met vergelijkbare problemen zien. Bevindingen in overeenstemming met beschikbaarheidbias werden geobserveerd bij tweedejaars internisten in opleiding, die lager scoorden op de casus die overeenkomsten vertonen met recent geëvalueerde casus dan op de andere casus. Dit patroon werd niet gezien bij de eerstejaars internisten in opleiding. Deze bevinding bevestigt gedeeltelijk hypothese 5 en 7, waarin ik verwachtte dat de bias groter zou zijn voor de meer ervaren artsen, maar ik verwachtte geen tegenovergesteld patroon voor de eerstejaars. Tweedejaars internisten in opleiding stelden de

fase 1 diagnose frequenter in fase 2 voor casus met oppervlakkig overeenkomstige kenmerken als ze die hadden gezien in fase 1 dan wanneer ze deze niet hadden gezien. Bovendien vond ik een significant hoofdeffect van redeneermodus, wat de zesde hypothese bevestigde: reflectie verbeterde de diagnoses van vergelijkbare casus in vergelijking met niet-analytisch redeneren voor zowel de tweede- als de eerstejaars internisten in opleiding. Dit betekent dat als tweedejaars internisten in opleiding casus diagnosticeerden die overeenkomsten vertonen met eerder geëvalueerde casus tijdens niet-analytisch redeneren, zij fouten maakten door beschikbaarheidbias. Deze fouten werden geneutraliseerd door reflectief redeneren en deze reflectie leidde tot verbetering van diagnostische accuratesse bij zowel eerste- als tweedejaars internisten in opleiding.[22]De conclusie van de studies die staan beschreven in het tweede deel van dit proefschrift, luidt dat beschikbaarheidbias inderdaad kan leiden tot diagnostische fouten. Dit lijkt niet te gelden voor de eerstejaars internisten in opleiding, die zelfs voordeel lijken te hebben van het eerst zien van een gelijksoortige casus. Deze resultaten bevestigen deels de vijfde en zevende hypothese omdat ik verwachtte dat de bias ook eerstejaars internisten in opleiding zou treffen. Verder lijkt reflectie de invloed van deze bias op diagnostische accuratesse te neutraliseren, waarmee de zesde hypothese bevestigd werd.

Aangezien de studies die worden beschreven in dit proefschrift een relatief nieuw onderzoeksdomein vormen binnen het onderzoek naar de ontwikkeling van medische expertise, er een aantal vragen onbeantwoord bleef en nieuwe vragen werden opgeroepen zijn er voldoende aanknopingspunten voor verder onderzoek. Zo zullen we in de aankomende jaren trachten te achterhalen hoe reflectief redeneren precies leidt tot het herstellen van cognitieve diagnostische fouten. Tevens zullen we onderzoeken of reflectief redeneren aangeleerd kan worden. Verder zullen we nagaan of het in **Hoofdstuk 6** geobserveerde ervaringsafhankelijke verschil in vatbaarheid voor (beschikbaarheid)bias inderdaad belangrijk is en welke mechanismen aan dit fenomeen ten grondslag liggen.

REFERENCES

1. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study 1. *New Engl J Med* 1991;324:370-6.
2. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;38:261-71.
3. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Australia* 1995;163:458-71.
4. Graber ML, Franklin N,. Gordon R. Diagnosticerror in internalmedicine. *Arch Intern Med* 2005;165:1493-9.
5. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775-80.
6. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *Brit Med J* 2005;330:781-4.
7. Redelmeier DA. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115-20.
8. Evans JSBT. Bias in human reasoning: causes and consequences. NJ, Erlbaum; 1989, p. 41.
9. LeBlanc VR, Norman GR, Brooks LR. Effect of a diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med*2001;76:S18-20.
10. LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med*2002;77:S67-9.
11. Hatala RM, Norman GR, Brooks LR. The impact of a clinical scenario upon electrocardiogram interpretation. J Gen Intern Med. 1999;14:126–129.
12. Van den Berge K, Mamede S, Van Gog T, et al. Accepting Diagnostic Suggestions by Residents: A Potential Cause of Diagnostic Error in Medicine. *Teach Learn Med* In press, 2012.
13. Pyszczynski T, Greenberg J. Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis testing model. *Adv Exp Soc Psychol* New York: Academic Press; 1987:297-340.
14. Norman GR, Brooks LR, Coblentz CK, Babcook CJ. The correlation of feature identification and category judgments in diagnostic radiology. Mem Cognition. 1992;4:344–355.
15. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008;121:S2-S23.
16. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ* 2004;38:1302-8.
17. Mamede S, Schmidt HG, Rikers RMJP, Penaforte JC, Coelho-Filho JM. Breaking down automaticity: Case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ* 2007;41:1185-92.
18. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185:1124-31.
19. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychol* 1973;5:207-32.
20. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise - Theory and implications. *AcadMed* 1990;65:611-21.
21. Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *EducPsycholRev*1993;5:205-21.
22. Mamede S, van Gog T, van den Berge K, Rikers RMJP, van Saase JLCM, van Guldener C et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;304:1198-1203.

Appendices

APPENDIX 1: EXAMPLE OF A CASE (CORRECT DIAGNOSIS: LIVER CIRHHOSIS)

In the studies in Part 1: to be confounded with livermetastasis

In the studies in Part 2 (Chapter 6): to be confounded with viral hepatitis

The diagnostic suggestions (part I) were conveyed as follows:

"Read the following case quickly but carefully. The diagnosis that was made for this case is **liver cirrhosis**. When you have read the case you are asked to indicate whether you agree or disagree with this diagnosis."

History: A 45-year old lawyer complains about ongoing pain in the upper abdomen. The patient relates the pain to stress due to a decreasing number of clients and his divorce, now 2 years ago. He has sex with prostitutes on occasion but has been impotent lately. He smokes 40 cigarettes a day and drinks substantial amounts of alcohol. His medical history includes surgery for prostatic cancer 5 years ago, no intolerances for food.

Physical examination: pale man, Blood Pressure: 110/69, Heart Rate: 85 beats per minute, Temperature: 37 ºC. Thorax: heart and lungs: normal. Spider naevi present. Abdomen: mild distension of the abdomen, percussion normal. Palpable liver with irregular surface. No splenomegaly. Extremities: ankle edema. Testicles: very small

Laboratory testing: Hemoglobin 5.0 mmol/L (8,6-10,5); ESR 44 mm/h (<20); Sodium 138 mmol/L (135-145); Potassium 3.6 mmol/L (3,5-5,0); ALAT 120 U/L (<41); ASAT 84 U/L (<37); LDH 800 U/L (<450); y-GT 250 U/L (<50); Alkaline Phosphatase. 200 U/L (<120); Bilirubin 42.7µ mol/L (<17).

Please indicate whether you agree or disagree with the diagnosis:

I agree with the diagnosis ☐

I disagree with the diagnosis ☐

APPENDIX 2: EXPERIMENTAL PARADIGM USED TO INDUCE REFLECTIVE REASONING

Read the case again (the case is presented).

A) Please provide in the first column the diagnosis you made for this case in the previous session.

B) List the findings that support and oppose this hypothesis in the next two columns and list findings that would be expected if this hypothesis were true but which were not encountered in the case in column 4.

C) List alternative diagnoses if the initial hypothesis proved to be incorrect in column 1 and repeat step B for each alternative.

D) Based on this analysis, rank your hypotheses according to diagnostic likelihood and present a final diagnosis.

Diagnostic hypothesis	Findings, supporting this hypothesis	Findings, opposing this hypothesis	Findings that were expected but not present

Dankwoord

DANKWOORD

12. Geen boekje zonder hulp…

Na deze belangrijke laatste, zij het wat verdekt opgestelde, stelling te onderbouwen leg ik mijn pen neer. Een proefschrift schrijf en bedenk je niet alleen. Ik heb tijdens de totstandkoming van dit boekje het geluk gehad dat ik ondersteund ben door een heleboel mensen. Als deze hulp er niet was geweest, had dit ook gegolden voor mijn proefschrift.

Allereerst wil ik alle collega-AIOS en internisten hartelijk danken voor hun medewerking aan de studies. In het bijzonder veel dank voor het geduld en de inzet van de AIOS uit de OOR Zuidwest Nederland.

Op de voet gevolgd door mijn promotoren, Prof.dr. J.L.C.M. van Saase en Prof.dr R.M.J.P. Rikers. Beste Jan, mede door jouw toedoen ging ik "down the rabbit hole". Jij hebt me de mogelijkheden en het vertrouwen gegeven om dit traject goed te doorlopen. Hiernaast gaf je me de vrijheid het promoveren te combineren met mijn bestuurlijke interesse en het werken aan een klinische studie. Ik kijk uit naar onze verdere samenwerking en je supervisie tijdens mijn laatste twee opleidingsjaren tot internist.

Beste Remy, je ontving een vreemde, "nogal eigenwijze" eend in de bijt met Limburgse gastvrijheid, regelde een werkplek en een geweldig begeleidingsteam, waaraan je zelf leiding gaf. Ik heb veel geleerd van je originele kijk op de dingen, je kritische commentaren en alle metaforen. Je deur stond altijd open, heel veel dank hiervoor.

Prezada Sílvia, você sabe como ninguém o que significa para um físico se aventurar em discuções e pesquisas sobre cognição. Durante esse tempo todo seus conhecimentos sobre este domínio mostraram-se ser sem igual. Suas opiniões foram sempre extremamente inteligentes e detalhadas. Se não fosse você, os resultados não teriam sido tão bons. Eu espero que possamos vir a unir forças em muitos estudos futuros! -Espero que seja bom Português-

Mijn andere steun en toeverlaat, Dr. T.A.J.M. van Gog. Beste Tamara, ik ben er trots op dat ik een tijdje met zo'n briljant wetenschapper, maar bovenal onbaatzuchtig en ontzettend leuk mens, heb mogen samenwerken. Jij bent degene die mij, veelal met behulp van "worked examples", hebt leren schrijven.

Prof.dr. H.G. Schmidt. Beste Henk, ik weet nog goed dat ik je voor het eerst ontmoette en aan je probeerde uit te leggen waarom je met me in zee moest gaan. Binnen vijf minuten was het

geregeld en je bent me gedurende het gehele traject, ook tijdens je nieuwe baan als rector, blijven vertrouwen, steunen en inspireren.

Prof.dr. G.R. Norman, dear Geoff, the other "mastodon of medical expertise". Your comments on my first paper were sharp, confronting en somewhat disappointing (i.e., they helped a lot). I am honored by your regular: "What do you think, Kees?" during last year's meetings with Henk.

Prof.dr R.P. Koopmans, Prof.dr. F.G.W.C. Paas en Prof.dr. H.T. van der Molen. Beste Richard, Fred en Henk: dank voor het beoordelen van het manuscript en jullie bereidheid tot plaatsnemen in de grote commissie.

Prof.dr.ir A.P.N. Themmen. Beste Axel, tijdens een gemeenschappelijk congresbezoek in de VS kwamen we erachter dat we twee passies, opleiden en endocrinologie, delen. Hopelijk kunnen we na mijn remigratie gaan samenwerken op deze gebieden. Dank voor het plaatsnemen in de grote commissie.

Prof.dr. J. de Graaf. Beste Jacqueline, dank voor je hulp bij het uitvoeren van de features-studie, het kritische commentaar, het plaatsnemen in de grote commissie en de samenwerking binnen het concilium.

Prof.dr. J.A. Romijn. Beste Hans, dank voor je hulp bij het uitvoeren van de eerste studie en voor het kritisch beoordelen van het paper.

Dr. A. Berghout, Dr. B. Bravenboer, Dr. P.M. Netten en Dr. P.L. Rensma. Beste Arie, Bert, Paetrick en Wiek: veel dank voor het onbaatzuchtig mobiliseren en motiveren van jullie AIOS!

Dr. C. van Guldener, beste Coen, jij begeleidde me tijdens mijn eerste stappen in de interne geneeskunde. We hebben samen veel ondernomen: case reports geschreven, studies gedaan, lang vergaderd in het concilium en tijdens visitaties en nog veel langer gepraat tijdens de gezamelijke reizen van en naar deze besprekingen. Hopelijk kunnen we onze samenwerking verder uitbreiden en nog lang voortzetten!

Prof.dr. W.W. de Herder. Beste Wouter, we hebben meermaals mijn endo-opleidingsschema tegen het licht gehouden en je dacht altijd in oplossingen. Ik ben blij dat ik mijn internistenopleiding mag vervolgen op jouw afdeling.

Dr. P.L.A. van Daele en Prof.dr. P.M. van Hagen. Beste Paul en Martin, dank voor jullie vertrouwen in en hulp met de Estetrol-studie. Laten we snel afronden! Hopelijk is er straks nog tijd en inspiratie voor een aantal vervolgstudies.

Mijn kamergenoten van T-13.01, Esther, Marlies en Annemarie. Het waren twee leuke jaren met voor ieder van ons mijlpalen op persoonlijk en professioneel vlak. Dank voor de gezelligheid en de klankborden.

Simone, waar moet ik beginnen? "Is de JNIV niets voor jou..?", "Ik heb nog een leuke studie maar niet zo veel tijd…", "Promotieonderzoek? Ja, doen!" Zonder jouw voorzetten op maat, een stuk minder inkoppers.

Beste Peter, grote steun tijdens mijn (e.v.a.) eerste internistische stappen, goede vriend en fantastische dokter. Ik hoop dat we elkaar veel blijven zien!

JNIV bestuursleden, in het bijzonder mijn (oud) conciliumcollega's Douwe en Monique. Wat hebben we de afgelopen jaren intensief samengewerkt aan en gebrainstormd over nieuwe opleidingsplannen (en perikelen), onderwijsinnovaties, opleidingsvisitaties, expertisecentra et cetera. Dank voor alle afleiding en inspirerende gesprekken. Het (is en) komt goed met de interne!

Arjan & Suuz, Peter, Eric, Erik-Jan, Niels et Marie, Sven & Marlieke, Evert-Jan (sharkman) & Marjolijn, Jubi, Robbert & Karin, Werner & Sandra, Lukas, Arnold en Rein, dank voor jullie belangstelling en afleidingsmanouvres.

Mijn paranimfen Jouke en Anthony. Sommige mensen zijn zo duidelijk bepalend in een mensenleven. Zouden ze dat zelf doorhebben? Jouke, je bent er, op 377 dagen van mijn leven na, altijd voor me geweest. Van spelen via muziek en (vaak!) helpen verhuizen naar "gewoon" broer zijn. Anthony, vanaf "ik ben Anthony en ik heb hooikoorts" (tijdens een voorstelrondje zittend op een grasveld aan het begin van de introweek in Antwerpen) is het lachen niet opgehouden. Mentale sparringspartner (ongelijke strijd), beste vriend. Zonder jou een andere kijk op de wereld, geen artsenbul, niets. Ik prijs me gelukkig met jou en Jouke achter me tijdens mijn verdediging.

Mijn schoonfamilie: Ad, Ine, Jordi en Kristel: jullie zijn goud waard. Dank voor de belangstelling, het oppassen, de hulp in en om het huis, kortom het altijd klaarstaan voor mij, Mies en de jongens.

Pa en ma, dank voor jullie liefde, steun èn kritische noten bij alles wat ik doe. Ik had niet zonder gekund! Rest van het nest, Jouke & Sanne (en ?), Marjolein & Itay, Kristiaan & Sonja, Jasper en Margreet, ik prijs me gelukkig met zulke fijne broers, zussen, schoonzussen en אח.

Mijn liefste lieve Mies, lucky charm. De tijd vliegt! Op de plaats, rust, of toch maar niet? ;-) Laten we samen jong blijven en oud worden. Waar we ook belanden, met jou erbij komt het goed. Lieve Sam en Job, jullie zijn allebei geboren toen papa met zijn "boekie" bezig was en maken mijn leven samen met mama tot een groot feest. Er is niets belangrijker dan samen vechten en jullie geschater. De toekomst ligt open, ik houd van jullie!

Kees

PhD Portfolio

PHD PORTFOLIO

Name PhD student:	C.K.A. (Kees) van den Berge
Erasmus MC Department:	Internal Medicine
PhD period:	May 2009 – October 2011
Promotoren:	Prof.dr. J.L.C.M. van Saase
	Prof.dr. R.M.J.P. Rikers
Copromotor:	Dr. S. Mamede

PhD training:

Applied multivariate data analysis, research methodology, scientific English communication.

Publications:

Van den Berge K, Mamede S, Van Gog T, Romijn JA, Van Guldener C, Van Saase JLCM, Rikers RMJP. Accepting diagnostic suggestions by residents: a potential cause of diagnostic error in medicine. *Teach Learn Med* in press, 2012.

Van den Berge K, Van Gog T, Mamede S, Schmidt H, Van Saase J, Rikers R. Acquisition of visual perceptual skills from worked examples: learning to interpret electrocardiograms. *Interact Learn Environ* 2011; Epub ahead of print.

Van den Berge K. A postoperative puzzle. *Neth J Med* 2011;4:184.

Mamede S, Van Gog T, **Van den Berge K**, Rikers RMJP, Van Saase JLCM, Van Guldener C, Schmidt HG. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010; 304(11):1198-1203.

Van den Berge K, Van Guldener C, Verburg GP. Recurrent pericarditis and eosinophilia in a young woman. *Tijdschr Infect* 2010;5:180-3.

Van den Berge K, Diderich K, Poddighe P, Berghout A. Symptomatic hypoparathyroidism based on a 22q11 deletion first diagnosed in a 43-year-old woman. *Neth J Med* 2009; 67(3):102-4.

Van den Berge K, Van Guldener C, Van Wijngaarden P. Erectile dysfunction as the presenting symptom of a pheochromocytoma. *Eur J Intern Med* 2007;18(8):593-4.

Schmidt HG, Mamede S, **Van den Berge K**, Van Gog T, Van Saase JLCM, Rikers RMJP. Exposure to media information causes availability bias in medical diagnosis. *Submitted*

Van den Berge K, Mamede S, Van Gog T, De Graaf J, Van Saase JLCM, Rikers RMJP. Supportive features mediate the tendency to accept diagnostic suggestions. *Submitted*

Van den Berge K, Mamede S, Van Gog T, Van Saase JLCM, Rikers RMJP. Can diagnostic suggestions lead physicians astray? *Submitted*

Ennecker-Jans SAM, **Van den Berge K**, Van Hagen PM, Van Laar JM, Van de Merwe JP, Van Daele PLA. Estetrol therapy in Sjögren's syndrome: An Open Proof of Concept Study. *Submitted*

Non-scientific publications:

JNIV bestuur. *Een nieuw opleidingsplan: Vernieuwing, verandering en Golden Rules*, Interne geneeskunde, maart 2011

JNIV bestuur. *Opleidingsvernieuwing interne geneeskunde,* Informatiefolder voor AIOS interne geneeskunde

Speelman P, Slaets JPJ, De Graaf J, Van Saase JLCM, Smulders YM, Van der Helm HMJ, Cloesmeijer WE (red). *Opleidingsplan interne geneeskunde: Eén team, één taak. 2009*

Van den Berge K, Nieuwenhuijzen L. *Digitaal portfolio,* http://www.internisten.nl/jniv/opleiding/digitaal-portfolio2, maart 2010

Presentations at international conferences:

Van den Berge K, Mamede S, Van Gog T, Van Saase JLCM, Rikers RMJP. *Consistency in diagnostic suggestions does not influence the tendency to accept them*. Paper accepted for presentation at the Annual Meeting of the AERA, Vancouver, BC, Canada.

Van den Berge K, Mamede S, Van Gog T, De Graaf J, Van Saase JLCM, Rikers RMJP. (2011, October). *A focus on supportive features mediates the tendency to accept diagnostic suggestions*. Paper presented at the Annual Diagnostic Error in Medicine Conference, Chicago, Illinois, USA.

Van den Berge K, Mamede S, Van Gog T, De Graaf J, Van Saase JLCM, Rikers RMJP. (2011, April). *A focus on supportive features leads residents to accept diagnostic suggestions*. Paper presented at the Annual Meeting of the AERA, New Orleans, Louisiana, USA.

Van den Berge K, Van Gog T, Mamede S, Schmidt HG, Van Saase JLCM, Rikers RMJP. (2010, April). *The effect of worked examples on ECG interpretation*. Paper presented at the Annual Meeting of the AERA, Denver, Colorado, USA.

Occasional Reviewing:

Medical Education, Advances in Health Sciences Education & BMC Medical Education

Award:

AERA Division I - outstanding publication award for paper: Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA, 2010; 304(11):*1198-1203

Scientific boards / committees: Dutch (Junior) Society for Internal Medicine, (J) NIV

2011 – current	Lid plenaire projectgroep opleidingsetalage
2010 – current	Lid Commissie Onderwijs Interne Geneeskunde (COIG)
2008 – current	Lid Concilium Medicinae Internae
2010 – current	Lid implementatiecommissie E-portfolio Concilium Medicinae Internae
2009 – 2011	Lid landelijke taakgroep implementatie opleidingsplan NIV
2009 – 2010	Lid adviescommissie E-portfolio Concilium Medicinae Internae
2008 – 2009	Lid commissie herziening raamplan (COHRA) interne geneeskunde
2008 – current	Bestuurslid, vertegenwoordiging OOR Zuidwest Nederland
2009	Organisator "De toekomst van de interne geneeskunde: De complexe patiënt", jaarlijkse JNIV conferentie, Zeist
2008	Vertegenwoordiger JNIV bestuur tijdens European Federation of Internal Medicine (EFIM) Conference, Rome, Italy

Didactic skills/ Teaching:

2011 – current	Supervising and teaching MSc students, Erasmus MC
2010	Teach the Teacher
2007 – 2009	Supervising and teaching MSc students, Amphia hospital
2006	Supervising and teaching MSc students, MCRZ hospital
2004 – 2005	Supervising and teaching MSc students, Amphia hospital

Society member

Dutch Society for Internal medicine (NIV)
Dutch Society for Medical Education (NVMO)
Dutch Endocrine Society (NVE)
European Federation for Internal Medicine (EFIM)

Curriculum Vitae

CURRICULUM VITAE

Kees van den Berge was born in Willemstad, the Netherlands, on May 4th, 1978. He completed his secondary education in 1996 at the C.S.G. Willem van Oranje in Oud-Beijerland. In June 2004, after clinical clerkships in Belgium and Sri Lanka, he received his medical degree at Antwerp University (with distinction). Following his work as an internal medicine resident at the Amphia hospital, Breda, he started his residency training at the Maasstad hospital, Rotterdam (supervisor: Dr. A. Berghout), followed by the Amphia hospital, Breda (supervisor: Dr. C. van Guldener) and the Erasmus Medical Center, Rotterdam (supervisor: Prof.dr. J.L.C.M. van Saase). In May 2009, he commenced work on the present dissertation. During these activities, he has been a board member of the Dutch Junior Society for Internal Medicine (JNIV) and a member of the Concilium of the Dutch Society for Internal Medicine (NIV), where he developed a special interest in medical education and management. Concurrently, he studied immuno-endocrine interactions in patients with Primary Sjögren's Syndrome with Dr. P.L.A. van Daele and Prof.dr. P.M. van Hagen. In November 2011, Kees started his fellowship in endocrinology in the Erasmus Medical Center, Rotterdam (supervisors: Prof.dr. W.W. de Herder and Prof.dr. J.L.C.M. van Saase). He lives in Breda with his wife Micha and their sons, Sam and Job.