



Trinity College Dublin
The University of Dublin

**Dual Process Interventions for Diagnostic Reasoning Among Medical
Students and Junior Doctors**

A thesis submitted to
The University of Dublin, Trinity College
for the degree of
Doctor of Philosophy

Kathryn Ann Lambe
School of Psychology
The University of Dublin, Trinity College
March 2018

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed: _____

Kathryn Ann Lambe

Date: ____/____/____

Summary

Diagnostic error is a significant challenge for the patient safety movement. A growing body of literature examines interventions to reduce diagnostic error, targeting both healthcare systems and cognitive processes of individual doctors. The dual process model of reasoning, wherein analytical and non-analytical modes of thinking contribute to judgement and decision-making, has emerged as the dominant model by which diagnostic reasoning is understood. This thesis examined the range of interventions for medical and psychiatric diagnostic reasoning using the dual process model of reasoning as an organising framework, with particular focus on guided reflection interventions, exploring the psychological processes underpinning the intervention's effect.

Chapter 2 presented a narrative review of the most relevant literature on medical error, diagnostic reasoning and the dual process model of reasoning as it has been applied to diagnosis. This review concluded that the dual process model represents a reasonably useful, if not perfect, method for analysing error, and that its application to diagnostic reasoning has not yet been subject to rigorous evaluation and validation. Research into cognitive interventions for diagnostic reasoning is gathering pace, though questions remain about the extent to which the interventions being proposed and examined reflect the psychological mechanisms being invoked to explain diagnostic behaviour in the broader literature.

Chapter 3 presented a systematic review of interventions, based on the dual process model of reasoning, to enhance analytical and non-analytical reasoning among medical trainees and doctors. The review concluded that while many of the studies found some effect of interventions, guided reflection interventions emerged as the most consistently successful in improving diagnostic accuracy, compared with educational interventions, checklists, cognitive forcing strategies and diagnostic instructions. In

addition, the review indicated that the presence of theory as a foundational rationale for intervention design was relatively shallow in most cases, and the need for theory to be incorporated into intervention design in this context was discussed.

Chapter 4 described a think-aloud protocol, in which thematic analysis was used to explore how trainee doctors in psychiatry and general practice made voluntary use of a guided reflection tool in diagnosing complex psychiatric clinical scenarios. A majority of participants chose to use the guided reflection tool and responded positively to it. The findings highlighted the impact of reflection on the process of diagnosis and on diagnostic confidence, and participants broadly indicated that the tool could be successfully implemented in real-world diagnosis, particularly for ambiguous or complex cases.

Chapter 5 presented an experimental study comparing a short and long version of the guided reflection intervention, undertaken with senior medical students. The study found little impact of the reflective intervention on diagnostic accuracy, and suggested that a short version of the guided reflection process elicits similar performance as a long version. The results also revealed that participants seldom changed their minds, highlighting the importance of the first impression in diagnosis and the relative lack of utility of the guided reflection process as a tool to ‘catch’ errors, at least for inexperienced diagnosticians.

Chapter 6 presented a discussion of the main findings of the thesis and their contribution to the literature on the dual process model of diagnostic reasoning and interventions based upon it. The theoretical, practical and educational implications of the findings were discussed, along with suggestions for future research in the field.

Acknowledgements

This PhD project was made possible through funding by the Irish Research Council, for which I am enormously grateful. I would also like to sincerely thank all of the participants who gave their time to take part in this research.

I have had the great fortune to work with three wonderful supervisors over the course of this project, and I am truly grateful to each of them.

To Prof Brendan Kelly, who instigated the project, encouraged me to be ambitious, and inspired me from the outset with his skilled wrangling of university bureaucracy and extraordinary work ethic. Thank you for giving so generously of your time, for your reliable support at all hours of the day, and for making this project possible.

To Prof Gary O'Reilly, who supported me through the first two years of the project, especially with the ethics process and the systematic review. Thank you for your patient advice and for always encouraging me to strive for scientific rigour and excellence; I learned so much working with you.

To Prof David Hevey, who has been a generous and kind mentor to me since my time as an undergraduate. Thank you for always challenging me to high standards of scholarship, for giving me confidence in my own ideas, for being a brilliant editor of ideas and words, and for your steady guidance throughout this process.

I extend my thanks to the Schools of Psychology at University College Dublin and Trinity College Dublin, and to all of the academic, executive and IT staff who assisted me in so many ways. Particular thanks go to Prof Mac MacLachlan, Prof Clare Kelly, and Prof Siobhán Corrigan for their appraisal of this research and for their encouragement and helpful feedback. Thanks to Prof Sam Cromie, who was a wonderful mentor as I began my teaching career and gave excellent advice whenever I

needed it. My thanks also to Prof Charlotte Wilson and Dr Ken McKenzie for their advice on analysis, and to Prof Jason Last at University College Dublin for his support. Warm thanks also to the students of Room E113 at University College Dublin, for your great camaraderie and friendship.

My thanks to Angela Rice, who kindly gave her time and expertise to teach me about systematic review techniques, and to Odhrán McCarthy, who helped to develop this project in the early stages.

I extend sincere thanks to Prof Charles Friedman and colleagues at the University of Michigan, who kindly gave me permission to use their clinical vignettes in the experimental study.

Dr Shane Rooney wrote and edited the clinical vignettes for the think-aloud study; Dr Rooney also assisted with recruitment, along with Dr Richard Farrelly. I owe an enormous debt of gratitude to both for their support, without which the study would not have been possible. My thanks also to Dr Richard Duffy for reviewing the vignettes and offering helpful comments and to Siobhán O’Leary for help with scheduling.

Sarah Curristan provided enormous assistance on the systematic review and think-aloud study – thank you so much for your patience and great attention to detail.

I owe a very special thanks to the past and present residents of Room 0.20 in Áras an Phiarsaigh. I’m so proud to be counted among your ranks, and I will always be grateful for your generous advice, great humour, and abiding friendship on this mad journey. Your companionship has made all the difference for me. Particular thanks to Maria, for giving me my first opportunity in research and teaching me so much; to Alicia and Laura, for your great company and many kindnesses; and to Sinéad and Emily, for your wisdom, encouragement and inspiring example.

Sincere thanks to Andy and Trisha Nugent, for your prayers and for cheering me on throughout the years.

To dear friends, Gemma and Lorraine, for helping with the life side of work-life balance, for making me laugh until my face hurts, and for always being there for me.

To Emmy (my *aibou!*), who never saw an Oxford comma she didn't love! Thank you not only for generously proofreading the work and providing hilarious commentary, but for being a mighty and inspiring companion in postgraduate research, for your great kindness and comfort in times of stress, and for your unfailing friendship.

Enormous thanks to Brendan for encouragement, pizza, and commiserations on late nights spent toiling over our laptops! Finally, to my Mum and Dad, Margaret and Paul Lambe, for your unwavering support and belief. I can never express my thanks for all you've done for me, and I hope that I've made you proud.

Dedication

I dedicate this work to the memory of my Nana, Annie Lambe, who loved education and whom I love very much.

Publications And Conference Proceedings Arising From This Work

Publications:

Lambe, K.A., O'Reilly, G., Kelly, B.D., Curristan, S. (2016). Dual process cognitive interventions to enhance diagnostic reasoning: A systematic review. *BMJ Quality and Safety*, 25:808-820.

Lambe, K.A., Hevey, D. & Kelly, B.D. (under review). Short and long guided reflection interventions for diagnostic error show no effectiveness for senior medical students. *Frontiers in Psychology*.

Conference Proceedings:

Lambe, K.A., O'Reilly, G., Kelly, B.D., Curristan, S. (2016, June). *Guided reflection interventions for diagnostic accuracy: A systematic review and meta-analysis*. Poster presented at Diagnostic Error in Medicine 1st European Conference, Rotterdam, Netherlands.

Lambe, K.A., O'Reilly, G., Kelly, B.D., Curristan, S. (2017, March). *Guided reflection interventions for diagnostic accuracy: A systematic review and meta-analysis*. Poster presented at the Psychology, Health and Medicine Conference, Dublin, Ireland.

Lambe, K.A., Hevey, D., Kelly, B.D. (2018, August). *Voluntary use of guided reflection for diagnostic reasoning in psychiatry: A think-aloud protocol with junior doctors.*

Poster presented at the Diagnostic Error in Medicine 2nd European Conference, Bern, Switzerland.

Table of Contents

Declaration	li
Summary	iii
Acknowledgements	v
Dedication	viii
Publications and Conference Proceedings Arising From This Work	ix
Table of Contents	xi
List of Tables	xiv
List of Figures	xv
Chapter 1: Introduction	1
Chapter 2: Diagnostic Error and Applications of the Dual Process Model	4
2.1 Introduction	4
2.2 Medical error	4
2.3 Diagnostic error	10
2.4 Theories of reasoning	16
2.5 Dual process theories	19
2.6 Interventions for diagnostic error	31
2.7 Conclusion	36
Chapter 3: Systematic Review of Dual Process Interventions for Diagnostic Reasoning	38
3.1 Introduction	38
3.2 Method	40
3.3 Results	48
3.4 Discussion	68
Chapter 4: Voluntary Use of Guided Reflection for Diagnostic Reasoning: A Think-Aloud Protocol Study	78
4.1 Introduction	78
4.2 Method	87

4.3 Pilot study	93
4.4 Results	98
4.5 Discussion	115
Chapter 5: Impact of Short and Long Guided Reflection Interventions for Diagnostic Reasoning: An Experimental Study	125
5.1 Introduction	125
5.2 Method	129
5.3 Pilot study	133
5.4 Results	137
5.5 Discussion	143
Chapter 6: General Discussion	151
6.1 Aims and contributions	151
6.2 Summary of research	152
6.3 Implications	153
6.4 Limitations	158
6.5 Future research	163
6.6 Summary and conclusions	167
References	168
Appendix A: Additional search strategies (Systematic review, Chapter 3)	180
Appendix B: Coding frameworks: Use of theory in intervention design (Systematic review, Chapter 3)	184
Appendix C: Ethical approval (Think-aloud study, Chapter 4)	186
Appendix D: Recruitment message (Think-aloud study, Chapter 4)	187
Appendix E: Information sheet (Think-aloud study, Chapter 4)	188
Appendix F: Consent form (Think-aloud study, Chapter 4)	191
Appendix G: Debriefing sheet (Think-aloud study, Chapter 4)	192
Appendix H: Vignettes (Think-aloud study, Chapter 4)	194
Appendix I: Correct / acceptable diagnoses (Think-aloud study, Chapter 4)	204

Appendix J: Semi-structured interview schedule (Think-aloud study, Chapter 4)	205
Appendix K: Experimenter script (Think-aloud study, Chapter 4)	206
Appendix L: Guided reflection written instructions (Think-aloud study, Chapter 4)	207
Appendix M: Coding framework (Think-aloud study, Chapter 4)	208
Appendix N: Ethical approval (Experimental study, Chapter 5)	215
Appendix O: Recruitment message (Experimental study, Chapter 5)	216
Appendix P: Information sheet (Experimental study, Chapter 5)	217
Appendix Q: Debriefing sheet (Experimental study, Chapter 5)	220
Appendix R: Instructions and diagnostic tables (Experimental study, Chapter 5)	222
Appendix S: Correct / acceptable diagnoses (Experimental study, Chapter 5)	232

List of Tables

Table 3.1 Search strategy for Medline (Ovid, 1990 to present)	46
Table 3.2 Characteristics of included studies	51
Table 3.3 Use of the dual process model as theoretical basis for included studies	59
Table 3.4. Risk of bias assessment of included studies	62
Table 4.1 Frequency analysis of participants and cases using guided reflection	101
Table 4.2 Contingency table for changed diagnoses with and without use of guided reflection	102
Table 5.1 Means and standard deviations obtained for proportions of Cases 1, 2, 3 and 4 diagnosed correctly	142
Table 5.2 Means and standard deviations obtained for proportions of all cases diagnosed correctly	143
Table 5.3 Contingency table for changed diagnoses in Cases 1, 2, 3 and 4 in short and long conditions	145
Table 5.4 Contingency table for changed diagnoses in all cases in short and long conditions	145

List of Figures

Figure 3.1 Kirkpatrick's adapted hierarchy of education outcomes	44
Figure 3.2. Literature search results and study flow	48
Figure 3.3. Forest plot with effect sizes (control vs intervention)	69
Figure 4.1. Diagrammatic illustration of the thematic analysis, showing major themes and sub-themes	103

Chapter 1: Introduction

Medical error is a serious concern for patients and healthcare practitioners alike. Since the Institute of Medicine's seminal report *To Err is Human* (Kohn, Corrigan, & Donaldson, 1999) claimed that up to 98,000 patients may die annually in American hospitals due to preventable medical errors, large-scale international studies have confirmed the scale of the problem of medical error, and the patient safety movement has gathered pace in responding to it (Leape & Berwick, 2005).

Diagnostic error presents unique challenges for researchers attempting to determine its scope and nature: it is more difficult to detect and solutions are less obvious than for other types of medical error, such as medication error and wrong-site surgery (Graber, Gordon, & Franklin, 2002; Kuhn, 2002). Although this type of error was initially somewhat neglected in the surge of research on medical error, it has received increasing attention in the last number of years and was highlighted in *Improving Diagnosis in Healthcare*, the Institute of Medicine's 2015 follow-up report to *To Err is Human* (Balogh, Miller, Ball, & Press, 2015).

A growing body of literature has emerged on interventions to reduce diagnostic error, targeting both healthcare systems and the cognitive processes of individual doctors (Graber et al., 2012; Singh et al., 2011). An attendant, though smaller, literature has explored how cognitive theories from psychological science may be applied to our understanding of diagnostic reasoning in general and diagnostic error in particular. The dual process model of analytical and non-analytical reasoning has been widely adopted as the dominant framework by which diagnostic reasoning may be understood (Balogh et al., 2015; Croskerry, 2009a). This model posits that two systems, or modes, of thinking contribute to reasoning. One system (analytical reasoning) is conscious,

deliberate, explicit, rational, and controlled, contrasting with the other system (non-analytical reasoning), which is unconscious, associative, implicit, intuitive, and automatic (Kahneman, 2011; Shafir & LeBouef, 2002; Stanovich & West, 2000).

The field of implementation science argues for the value of incorporating theory as a foundation for intervention design (Michie, van Stralen, & West, 2011). However, the extent to which interventions for diagnostic reasoning relate to the theoretical processes underpinning diagnosis remains largely unexplored, and the particular lens of psychological science is largely absent from the literature, which is almost exclusively the purview of medical researchers. Some research has already identified successful interventions to improve diagnostic error, as discussed in Chapter 3; methods and insights from psychological science have the potential to add nuance to this literature and to answer outstanding questions, including the illumination of the nature of the processes by which doctors make use of these techniques in real time and whether attempts to alter natural styles of diagnostic reasoning may have unintended consequences.

The aim of the thesis, therefore, is to examine a range of interventions for medical and psychiatric diagnostic reasoning using the dual process model of reasoning as an organising framework, to examine the extent to which the model has been utilised in intervention design, and to add nuance to our understanding of how interventions impact on diagnostic processes. By way of background to this programme of research, Chapter 2 presents a narrative review of the most relevant literature on medical error, diagnostic reasoning, and the dual process model of reasoning as it has been applied to diagnosis. Building on this, Chapter 3 presents a systematic review of interventions for diagnostic reasoning based on the dual process model, evaluating their effectiveness in improving diagnostic accuracy and critically considering the role of theory in their

design. Chapter 4 focuses on the most successful of these interventions – guided reflection – and presents findings from a think-aloud protocol study, in which the processes by which junior doctors diagnose psychiatric cases using guided reflection are elicited and participant opinions of the reflection process are explored. Chapter 5 further explores the guided reflection intervention, comparing the relative effectiveness of two versions of the task in improving diagnostic accuracy among medical students, and examining the potential for the intervention to have unintended negative consequences for diagnostic reasoning. Finally, Chapter 6 presents a discussion of the results and findings from the systematic review and empirical studies, and considers their contribution to the literature on dual process interventions for diagnostic reasoning and potential future directions for continued work in this area.

Chapter 2: Diagnostic Error and Applications of the Dual Process Model

2.1 Introduction

The dual process model of diagnostic reasoning has emerged as the dominant model by which diagnostic reasoning is conceptualised and understood within the medical literature (Balogh et al., 2015), and has received similarly widespread acceptance in other applied literatures, including human factors (e.g., use of heuristics and biases in understanding risk perception and planning behaviours in emergency incident management and military operations (Cook, Noyes, & Masakowski, 2007)) and behavioural economics (e.g., the development of effective prescriptive programs employing ‘nudges’ to harness natural heuristics and encourage individuals to make beneficial long-term economic choices, such as automatic enrolment into future savings schemes (Thaler & Benartzi, 2004; Thaler & Sunstein, 2003)). This chapter will critically evaluate the evidence underpinning the application of the dual process model to diagnostic reasoning and consider the research on interventions to improve diagnosis that has arisen from the model.

2.2 Medical error

In 1999, the Institute of Medicine’s (IOM) seminal report *To Err is Human* (Kohn et al., 1999) claimed that up to 98,000 patients may die annually in American hospitals due to preventable medical errors. Several large-scale international studies have supported this claim; a major 2007 UK review reported that 8%–10% of hospital admissions in the National Health Service result in an adverse event, between 30% and 55% of which are to some extent preventable (Sari et al., 2007). Studies between 1995 and 2009 in Australia (R. M. Wilson et al., 1995), the UK (Vincent, Neale, &

Woloshynowych, 2001), New Zealand (Davis et al., 2002), Sweden (Soop, Fryksmark, Koster, & Haglund, 2009), and Canada (Baker et al., 2004; Forster et al., 2004) report adverse event rates between 7% and 17% and preventability rates between 30% and 70%.

The IOM report, as well as illuminating the scale of the problem, highlighted the human factors in medical error, and signalled the beginning of the modern era in the patient safety movement. In the years that followed, similar international studies of error incidence gathered pace, national agencies for patient safety were founded in Canada and the UK, and the role of the larger health system in which the individual clinician operates was drawn into sharper focus in the international discourse around medical error (Croskerry, 2010). Importantly, clinicians and researchers broadly accepted the IOM's assertion that preventable error is a significant and heretofore underappreciated problem (Leape & Berwick, 2005).

While no standardised nomenclature exists within the literature, the IOM report defines medical error as a “failure of a planned action to be completed as intended (i.e., error of execution) or the use of a wrong plan to achieve an aim (i.e., error of planning)” (Kohn et al., 1999, p28). The report describes four types of error:

- (1) Diagnostic error (e.g., delay or error in diagnosis, failure to carry out indicated tests, failure to act on results of monitoring or tests)
- (2) Treatment error (e.g., error in performance of a procedure or test, error in treatment administration, dosage or method, delay in treatment, inappropriate care, delay in responding to test results)
- (3) Preventative error (e.g., failure to provide prophylactic treatment, inadequate monitoring or follow-up)

- (4) Other error (e.g., communication failure, equipment failure, other system failure)

The broader category of “adverse events” describes injuries caused by medical management rather than the underlying condition of the patient. An adverse event attributable to error is a “preventable adverse event” (Kohn et al., 1999, p28).

Medical error is a major concern for both patients and healthcare practitioners. Importantly, in a 2002 study of views on medical error in general, conducted with 1,207 members of the public and a random sample of 831 physicians, more than 70% of both groups assigned “a lot” of responsibility to the physicians involved, and more than 50% of both groups believed that mistakes made by individual professionals were more important than those made by health care institutions (Blendon et al., 2002). A majority of physicians cite understaffing of nurses and overwork, stress, and fatigue of hospital staff as root causes of error (Blendon et al., 2002). Patients cite carelessness and negligence of practitioners who are overworked or stressed as the main cause of medical mistakes (Kohn et al., 1999).

The economic cost of medical error is significant; errors incur both direct (additional health expenditure) and indirect (lost productivity, disability care) costs for patients and society at large. In 2012, the Health Service Executive of Ireland reported that more than 85,000 adverse events in Irish hospitals that year led to taxpayer costs of €81m. More than 15% of these were medical or treatment errors (O’Cionnaith, 2012). Diagnostic error, in particular, is the leading cause of paid malpractice suits in the USA; these claims are among the most costly, averaging more than \$380,000 per claim (Saber Tehrani et al., 2013).

2.2.1 Why errors happen

Errors occur in all settings and technological industries, and health care is a

particularly complex technological and cognitive environment that is accident-prone. Failures in large systems occur due to a convergence of events; that is, multiple faults occur together to create an event in which errors are not caught, and thereby can be compounded. Within the human factors literature, James Reason's seminal work *Human Error* (1990) presents a number of fundamental concepts, derived from study of industrial accidents, that have been used to analyse systems in healthcare and to understand adverse events.

2.2.1.1 Active errors and latent errors

At the broadest level, two types of error in complex systems may be identified: active errors and latent errors. The effects of active errors are immediately apparent, while latent errors may lie 'dormant' within a system until they combine with other factors in the system to produce accidents (Reason, 1990). Active errors may be perceived as being more readily addressed; as they are more usually associated with frontline operators in the system, the specific actions of an individual can be identified as the main cause of an event and corrective or punitive action can be taken. However, this focus allows latent errors, such as poor management, organisational structure, and faulty design, to accumulate undetected in the system and to contribute to future errors. In this way, latent errors may be described as 'accidents waiting to happen'.

For example, a nurse (a frontline operator) may, at the end of a long shift, make an active error in programming an intravenous pump, leading to an adverse event for a patient. However, the institutional use of multiple different types of pump may have rendered such programming errors more likely; this is a latent error, lying dormant within the system until it combined with the nurse's fatigue to create an adverse event. This conceptualisation of the root causes of error is particularly important in light of the

findings outlined above that practitioners and the public alike assign much responsibility for errors to individual healthcare workers; a general belief in the primacy of the individual human actor in medical error overlooks the crucial role of systems approaches in overcoming error.

2.2.1.2 The 'Swiss cheese' model

Therefore, one of the key insights in Reason's 1990 volume is that safety failures tend to result not from isolated active errors by individual workers, but from their interaction with latent errors in the system. This interaction was described by Reason in the 'Swiss cheese' model. Reason conceptualised systems as containing multiple components, represented by slices of cheese in a stack. Each component offers a layer of protection against error, but also contains vulnerabilities, represented by the holes in a slice of Swiss cheese. A failure of one component of the system alone does not necessarily create an accident, as other components can compensate and 'catch' the error. However, an accident occurs when failures occur in all of the components in such a way that vulnerabilities align, like holes in the slices of Swiss cheese (Reason, 1990).

2.2.1.3 The Human Factors Analysis and Classification System

Based on Reason's conceptualisations of active and latent errors and the Swiss cheese model, the Human Factors Analysis and Classification System (HFACS) was developed as a system for the analysis of human factors in aviation accidents (Shappell & Wiegmann, 2000). The HFACS describes four levels of potential failures (mapping onto the 'layers' of the Swiss cheese model): unsafe acts of operators / frontline staff, preconditions for unsafe acts, unsafe supervision, and organisational influences. As with the Swiss cheese model, an accident is triggered when failures occur in all levels of the

HFACS. Within the four levels are nested a total of 19 categories of causal factors that can be used to describe both active and latent errors. In this way, the HFACS offers a more detailed version of the Swiss cheese model that specifies the nature of potential failures (i.e., identifies the potential holes in the cheese), so that these may be highlighted in accident investigation or detected before an accident occurs.

2.2.1.4 Perrow's 'Normal accidents'

The 'normal accident' model proposed by Charles Perrow (Perrow, 1984) shares a number of features with other models of systems error. Perrow, like Reason, conceptualises the frontline operator as an internal part of a system, rather than an external agent who acts upon it. Operators interact with the system in complex ways, and constitute just one point at which failure can trigger complex, far-reaching consequences. It is this interaction of components within systems that Perrow places at the centre of the 'normal accidents' model; high-risk technological systems have potential for failure built into their connective tissue, rendering accidents inevitable and expected, or 'normal'.

Perrow describes two features of high-risk technological systems. The first is complexity; in complex systems, where the pathways between components are numerous and not always well understood, there are many opportunities for failure due to the myriad potential interactions of system components. The second is coupling; a system that is tightly coupled is one in which the potential for corrective action in the wake of a failure, before adverse consequences arise, is non-existent or very narrow.

At least some aspects of healthcare may be understood as a highly complex and tightly-coupled system. This is particularly the case in emergency medicine, where common presentations (e.g., weakness, chest or abdominal pain) carry very wide

differential diagnoses, the inherent stress of medical crises can hinder communication and impair physician judgement (Crookery & Sinclair, 2001), rapid decisions are made and enacted, and critically or urgently ill patients may suffer harm when their caregivers cannot recover from errors made in initial diagnosis or treatment (Kovacs & Crookery, 1999).

The complexity of converging events that create accidents, therefore, means that such conditions are difficult to foresee. Investigators reviewing accidents are vulnerable to hindsight bias; that is, warning signs and multiple contributing factors that were not considered at the time can seem obvious in retrospect (Bates, 2003; Soop et al., 2009). As described above, a single active error may be highlighted as the root cause of an adverse event, overlooking latent factors, and even significant reforms around these areas are generally insufficient to prevent future errors. The individual actor must be understood in context, and while the necessity of a combined systems approach to addressing medical error is generally acknowledged, the empirical research on interventions of this sort arguably lags behind the enthusiasm for them (Singh et al., 2011). This point will be developed further in this chapter in Section 2.6.

2.3 Diagnostic error

Fifteen years after the publication of *To Err is Human*, the Institute of Medicine published a follow-up report focusing specifically on diagnostic error. In *Improving Diagnosis In Health Care* (Balogh et al., 2015), the Institute highlighted the particular challenges in studying and addressing diagnostic error and commented that, compared to other aspects of medical error included in the original report, diagnostic error has received relatively little attention in the patient safety literature. There are many reasons for this; errors of this kind are difficult to detect and measure, their causes more subtle

and solutions less obvious than for other types of medical error, such as medication error and wrong-site surgery (Graber et al., 2002; Kuhn, 2002). However, diagnosis remains a crucial challenge for the field and a key concern for the public, given that meaningful diagnostic error is likely to affect all of us within our lifetime (Balogh et al., 2015), and diagnostic errors are associated with a proportionately higher rate of morbidity than the other types of error (Croskerry, 2003b). An Australian study indicated that permanent disability arose from diagnostic error in 32% of adverse events, compared with 17% for operative errors, 29% for therapy-related errors, and 17% for drug-related errors (R. M. Wilson et al., 1995).

The difficulty in studying diagnostic error comes in part from the lack of consensus around a definition for error. Operational definitions of diagnostic error are required for empirical studies attempting to measure the incidence of diagnostic error. One such operational definition is provided by Graber and colleagues: a diagnostic error occurs when a diagnosis was “unintentionally delayed (sufficient information was available earlier), wrong (another diagnosis was made before the correct one), or missed (no diagnosis was ever made), as judged from the eventual appreciation of more definitive information” (Graber, Franklin, & Gordon, 2005, p1493). Singh and colleagues provided another operational definition as part of their case review of diagnostic errors in primary care, emphasising missed or delayed diagnoses: an error was judged to have occurred when sufficient data was present at the primary care index visit that the final, correct diagnosis could have been made at an earlier stage. “Thus, errors occurred only when missed opportunities to make an earlier diagnosis occurred based on retrospective review” (Singh et al., 2013, p419). This definition is somewhat narrower than Graber and colleagues’ definition, but is highly practical for use as part of the retrospective review method. Schiff and colleagues, in eliciting reports of

diagnostic errors from physicians, provide a rather broader definition: “any mistake or failure in the diagnostic process leading to a misdiagnosis, a missed diagnosis, or a delayed diagnosis” (Schiff et al., 2009, p1881). While arguably less tightly specified than Graber’s or Singh’s definitions, this broader wording is suitable as a trigger for eliciting recall of a wide range of incidents by survey participants.

Improving Diagnosis in Healthcare offers an alternative definition, describing diagnostic error as “the failure to (a) establish an accurate and timely explanation of the patient’s health problem(s) or (b) communicate that explanation to the patient” (Balogh et al., 2015, p4). This definition uniquely emphasises the requirement for successful communication of the diagnosis to the patient; “From a patient’s perspective, an accurate and timely explanation of the health problem is meaningless unless this information reaches the patient so that a patient and health care professionals can act on the explanation” (Balogh et al., 2015, pp4-6). This definition arguably does not lend itself as readily to the empirical task of measuring diagnostic error as the alternatives listed above; however, it is significantly patient-centred and highlights the central role of communication in providing effective care.

2.3.1 Incidence of diagnostic error

Estimates from a diverse range of methodologies and medical specialisms indicate a diagnostic error rate of around 10%–15%. This estimate appears to be stable across hospitals, countries and decades (Kuhn, 2002). Each research method is likely to underestimate the actual error rate.

Autopsy studies may be regarded as a ‘gold standard’ for providing definitive data on diagnostic discrepancies; however, autopsies are only carried out on a small, non-representative number of cases. Autopsy studies reveal only that an error occurred,

offering little information as to why; standardised patient studies can offer more insight into the causes of error (Graber, 2013). Autopsy studies identify major diagnostic discrepancies in 10-20% of cases (Shojania, Burton, McDonald, & Goldman, 2002) and include cases of infectious, cardiovascular, pulmonary, renal and gastrointestinal disease (Sonderegger-Iseli, Burger, Muntwyler, & Salomon, 2000).

Malpractice claims provide a convenient source of data on large numbers of confirmed cases of diagnostic error, although such data privileges information on the consequences of such errors over their aetiology. Like autopsy studies, only a small number of errors lead to claims, and these cases are predominantly younger patients with cancer or cardiovascular disease (Saber Tehrani et al., 2013). Second reviews, retrospective case reviews, clinician surveys and diagnostic testing audits also contribute useful data on diagnostic error rates (Graber, 2013). However, none are particularly effective for studying human factors aspects of diagnostic error.

The visibility of errors in the medical setting presents challenges in studying and addressing them. When errors occur in healthcare, harm is done to a third party, not to the system itself. The majority of errors affect only one patient at a time, not large groups of people, making trends harder to detect (Kohn et al., 1999).

In particular, feedback and error visibility are important when considering diagnostic error. Feedback is most effective when it is communicated quickly, and when the interval between action and outcome is short. In the medical field, if diagnostic feedback is received at all, it often comes after a significant delay (Croskerry, 2000b). In the absence of specific, timely feedback, a diagnosis is easily assumed to be correct. However, if the patient receives an incorrect diagnosis or incorrect treatment, this may never be communicated back to their doctor; the patient may leave the practice to seek a second opinion, spontaneously improve, or return with more pronounced symptoms and

eventually receive a correct diagnosis (Berner & Graber, 2008).

2.3.2 Categorising diagnostic error

Graber, Gordon and Franklin (2002) propose a system of three broad aetiological categories of diagnostic error. This taxonomy is largely consistent with older taxonomies in the literature (Bordage, 1999; Chimowitz, Logigian, & Caplan, 1990; Kassirer & Kopelman, 1989) and was included in the IOM's 2015 report as a useful organisational tool for understanding the root causes of errors and, crucially for this area, places cognitive errors in the context of the system within which they arise.

No-fault errors include cases where the illness presents silently or is masked, or where the presentation is so atypical that successful diagnosis cannot be expected, given the current limits of medical knowledge. The category can also include presentation of very rare conditions and cases where patients are noncompliant or fail to represent their complaints clearly.

System errors comprise latent failures in the healthcare system, including technical (e.g., faulty equipment, lack of appropriate equipment, faulty tests) and organisational (e.g., inefficient processes, patient neglect, policy failures, poor coordination, inadequate training) failures. Organisational factors that lead to suboptimal working conditions, such as physician fatigue, stress, environmental distraction and excessive workload, are also included in this category. A related and important concept is that of safety culture, defined as “the product of individual and group values, attitudes, competencies and patterns of behaviour that determine the commitment to, and the style and proficiency of, an organisation's health and safety programmes” (Health and Safety Commission, 1993, cited in Halligan & Zecevic, 2011, p339). Organisations with positive safety culture are characterised by open

communication and teamwork, mutual trust, leadership commitment to organisational learning and safety, and shared belief in the importance of safety (Halligan & Zecevic, 2011).

Cognitive errors arise at the level of the individual physician; this category comprises cases of inadequate knowledge, faulty data gathering, faulty information processing or clinical reasoning, and errors in metacognition (e.g., faulty verification and monitoring of diagnostic processes).

In a 2005 retrospective review study of 100 cases of diagnostic error, both system-related and cognitive factors were found to have contributed to diagnostic error in approximately half of all cases. While the retrospective case review is by no means a perfect method, this study represents an important and relatively uncommon attempt to estimate the relative contribution of a variety of factors to diagnostic error. Cognitive errors (e.g., errors in knowledge, data gathering, or information synthesis) were implicated, alone or in combination with system-related errors, in three out of four cases (Graber et al., 2005). The study's findings confirm that, as in other industries, multiple factors contribute to diagnostic error; an average of six factors were implicated in each case. Faulty information processing or clinical reasoning was found to be the most common cognitive error; that is, the physician had adequate knowledge and information, but failed to synthesise it in the correct way to reach an accurate diagnosis. Examples of this sort of error include misidentification of symptoms, misinterpretation of test results, failure to account for contextual information, over- or underestimation of the salience of a finding, failed heuristics or susceptibility to cognitive bias, and failure to act on available information.

The above indicates that understanding clinical reasoning at the level of the individual diagnostician is critical in the effort to understand diagnostic error and, by

extension, to design interventions to reduce it. Established theories from psychological science around reasoning and decision-making, particularly under conditions of uncertainty, have therefore been given a prominent role in the discussion.

2.4 Theories of reasoning

The study of reasoning and decision-making has been bolstered in the last century by contributions from many fields, including psychology, economics, statistics, philosophy, and military and political science. Two broad categories of decision theories are commonly identified; normative decision theories are concerned with how optimal decisions may be made under optimal conditions, while descriptive decision theories are concerned with how decisions are actually made. Decisions prescribed by normative theories may be regarded as a standard against which real-world decisions may be judged, and thereby explained, by descriptive theories. A third category of decision theory is also sometimes invoked, namely prescriptive theories, which are concerned with how decision-making processes and outcomes may be improved within the limitations of real-world decision contexts (Baron, 2007). It is worth considering these models, as medical decision-making of all kinds may be considered a special case of decision-making under uncertainty (Graber et al., 2012).

2.4.1 Normative models

One of the key normative models of decision-making under uncertainty is expected utility theory, which describes a process whereby decision-makers mathematically calculate the expected utility of a set of outcomes (the product of the value of a given outcome and its probability) and thereby identify and select the option with the greatest expected benefit (Baron, 2007).

Bayes' theorem, another major normative model of decision-making under uncertainty, represents a procedure used to estimate probability in the light of iteratively updated information derived from hypothesis testing; it therefore has clear theoretical applications in diagnosis, which requires updating beliefs regarding a hypothesis in the light of new evidence (Baron, 2007).

Both of these normative theories have clear utility in diagnosis; however, their implementation in deriving 'optimal' decisions for clinical scenarios is subject to much debate (Payne, 2011) and they do not constitute feasible methods by which diagnosticians can reach conclusions in everyday practice. Both methods require specific estimates of probabilities at each stage of the process; in a complex clinical environment with multiple patient, practitioner and system factors interacting, such probabilities are often impossible to reasonably calculate, with or without the additional pressures of real-world decision-making on a ward (Baron, 2007). This is compounded by concerns that doctors, like the general population, may struggle significantly with estimating probabilities and with statistical literacy more generally (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007).

2.4.2 Descriptive models

Considered to be the most widely studied and accepted model of clinical reasoning until around the turn of the century (Kovacs & Croskerry, 1999), hypothetico-deductive reasoning proceeds in a number of distinct stages. Generation of diagnostic hypotheses begins early in the process, perhaps even before the clinical encounter with the patient. In the next stage, the alternative hypotheses are evaluated to either eliminate or, more commonly, to confirm them. Subsequently, or sometimes concurrently, some hypotheses are iteratively refined and others are dropped. Finally, the chosen hypothesis

may be verified, on the basis of its adequacy, coherency and parsimony, before a working diagnosis is accepted and action is taken (Kovacs & Croskerry, 1999).

The hypothetico-deductive model derives from both psychological theory and from the scientific method, and is concerned heavily with logical validity; if the premises in a deductive argument may be shown to be valid, the conclusions logically derived from these premises must also be valid or true. However, in clinical reasoning, “there is never any guarantee, only an inductive inference of varying degrees of certainty” (Croskerry, 2000a, p1226). Broadly speaking, deductive reasoning progresses from the general to the specific; inductive reasoning, in contrast, progresses from the specific to the general, or from the clinical data to a diagnosis.

Scheme-inductive reasoning is based upon the concept of schemes from the psychological literature, referring to the organisational structure of knowledge that experts develop over repeated exposure to information. In scheme-inductive reasoning, such structures provide specified pathways for diagnosis, conceptualised as decision ‘trees’. Diagnosticians gather data to distinguish between choices at branching points in the decision tree; it is at these points that decisions explicitly occur (Coderre, Mandlin, Harasaym, & Fick, 2003).

A related strategy is pattern-recognition; repeated exposure to a range of illness presentations allows a novice diagnostician to develop a ‘repertoire’ of clinical problems in their domain of expertise. Increasingly, new problems can be recognised as similar or identical to old problems, which the clinician has already solved, and their solutions can be recalled (Schmidt, Norman, & Boshuizen, 1990).

The mental model theory of logical reasoning posits that, given a set of premises or declarations about a particular situation, individuals construct mental representations, or models, of the situation and use these mental models to draw conclusions (Baron,

2007; Johnson-Laird, Byrne, & Tabossi, 1989). An individual may generate multiple mental models of a given situation, each representing a possible version of reality. Deductive reasoning proceeds as the reasoner examines whether a given conclusion holds in all models that fit the given premises; reasoning is most effective when the reasoner deliberately seeks models that fit the given premises but in which their conclusion does not hold. Errors may arise when reasoners do not consider all possible mental models, particularly models in which the conclusion does not hold true (Baron, 2007). Mental models can be influenced by prior general knowledge about the world, and reasoners may develop strategies to deal with particular types of problems that they regularly encounter. The mental model theory has been used to describe how disease information may be organised in mental representations and how these knowledge structures undergo change with the accumulation of experience: transitioning from elaborated causal networks (causes and consequences of pathophysiological processes) to abridged networks (simplified causal models, incorporating signs and symptoms observed by the doctor in real patients) to illness scripts (list-like structures or prototypes) to instance scripts (real patient encounters that do not conform to the prototype) (for a summary, see Payne, 2011) .

2.5 Dual process theories

The family of dual process theories of decision-making, associated most famously with the large body of work of Kahneman and Tversky (Kahneman, 2011; Tversky & Kahneman, 1974), posits that two systems, or modes, of thinking contribute to reasoning. One system (non-analytical reasoning, or System 1) may be described as unconscious, associative, implicit, intuitive, and automatic, contrasting with the other

(analytical reasoning, or System 2), which is conscious, deliberate, explicit, rational, and controlled (Kahneman, 2011; Shafir & LeBouef, 2002; Stanovich & West, 2000). Deductive and inductive (or pattern recognition) strategies are incorporated into the analytical and non-analytical modes respectively, though they alone cannot be said to entirely constitute these modes of reasoning. While a large number of dual process models of reasoning exist (for an overview, see Evans, 2008; Osman, 2004), all broadly describe the two systems in similar ways and acknowledge that the relationship between the two is interactive.

Evans and Stanovich (2013) write that three types of evidence exist for dual processes. First, experimental evidence demonstrates that each process may be experimentally manipulated independently, leaving the other unaffected. Researchers have shown that logical accuracy in responding decreases under conditions of time pressure (Evans & Curtis-Holmes, 2005), which facilitate primarily non-analytical or System 1 responses. Increasing working memory load, thus restricting resources available for analytical or System 2 processing, also leads to increased logical errors (De Neys, 2006b, 2006a), and mistakes due to the conjunction fallacy on the famous Linda problem (Kahneman, Slovic, & Tversky, 1982) have been associated with faster response times (De Neys, 2006a). It is important to note that if these effects were due simply to increased task difficulty or guessing, one would expect to see a pattern of increased random errors; instead, in many tasks a modal response is observed, indicating a systematic intuitive bias distinct from random error (Kahneman, 2011).

Second, evidence from neural imaging supports the existence of distinct brain activation patterns for each of the two processes. For example, a number of studies demonstrate that logic-based and belief-based responses are associated with activation in different brain areas (e.g., Tsujii & Watanabe, 2009). Similarly, decisions to

immediately accept (based on emotional processing) elicit activation in the brain's limbic system, while decisions to defer rewards (requiring mental simulation of the future) elicit activation in the prefrontal and frontal cortices (McClure, Laibson, Loewenstein, & Cohen, 2004).

Third, evidence from the literature on individual differences supports the dual process model of reasoning. Greater working memory capacity, dispositional motivation and intelligence are associated with more normative responses on decision-making tasks that are cognitively demanding; the effect is not present on tasks where pragmatic cues can lead to success (Stanovich & West, 1998).

The model is not without its critics (Keren & Schul, 2009; Kruglanski & Gigerenzer, 2011; Osman, 2004), and Kahneman himself acknowledges that the model is set up primarily as a metaphor (Kahneman, 2011). Common criticisms include arguments that the proposed two-process model does not adequately explain the full range of reasoning processes, that inferences reflecting normative principles can be made automatically, that heuristics can be invoked in the process of explicit reasoning, and that both modes of reasoning are based on a common rule-based principle. (For an overview of criticisms, see Keren & Schul, 2009; for a response to common criticisms, see Evans & Stanovich, 2013.) While alternative models that posit analytical and non-analytical reasoning as opposing extremes of a spectrum (Custers, 2013), or as a two-step process with common underlying mechanisms (Kruglanski & Gigerenzer, 2011), have gained some traction, the distinction is mostly academic in the diagnostic reasoning literature (Norman, Monteiro, & Sherbino, 2013).

2.5.1 Advantages of the dual process model

The dual process model has been widely adopted in the medical error literature

as the dominant paradigm by which clinical reasoning may be understood (Balogh et al., 2015). The model offers a number of advantages that may explain its widespread adoption. First, it constitutes a relatively simple model with sound theoretical structure (see Section 2.5.2). The model is more readily applied to real-world clinical scenarios than quantitative or actuarial models, which are frequently unable to account for the range of unknown factors, large numbers of variables, ethical implications, and resource restrictions that characterise complex clinical scenarios (Croskerry, 2009a). This simplicity means that the model is relatively straightforward to teach to novice diagnosticians, who may use it as a framework within which to examine and correct their own decision-making and metacognitive processes (Croskerry, 2009b).

Second, the dual process model is situated within a ‘common ground’ between a number of schools of thought, and incorporates aspects of diverse approaches to decision-making that have historically been invoked in the medical field. In particular, the model represents a ‘bridge’ from normative to descriptive and prescriptive models of reasoning, as outlined above. Where normative models provide statistically- and mathematically-derived standards for thinking and judgement, the dual process model, as a descriptive model, accounts for the ways in which reasoning deviates from these ‘optimal’ processes and, critically, provides a pathway towards prescriptive processes by which thinking and decision-making may be improved (Baron, 2007).

Third, the model provides potentially fertile ground for hypothesis generation, in outlining distinct processes that can be tested for their appropriateness, robustness, and vulnerability in a variety of clinical contexts. However, much of the work to date in this field appears to be primarily problem-driven, rather than theory-driven, and so the model’s potential as a source of experimental hypotheses remains to be fully explored.

2.5.2 Application of the dual process model to diagnosis

One common criticism of dual process theories is that the models are numerous and poorly differentiated, with disparate proposed clusters of attributes for each of the two systems (Evans & Stanovich, 2013). The diagnostic reasoning literature regarding ‘the dual process model’ seldom makes reference to any particular dual process model, invoking instead a rather broad, generic model derived from this class of theories and frequently adopting the language of ‘Systems 1 and 2’.

The most highly-specified conceptualisation of the dual process model's application to the diagnostic reasoning process was proposed by Croskerry in 2009, and is a representation of the model as it has been broadly understood and endorsed within the clinical reasoning research. The central assertion is that diagnosis can follow a number of paths. First, the patient's symptoms and signs are presented to the diagnostician. If salient features of the presentation are immediately recognised, non-analytical pattern-recognition processes are engaged, reflexively and unconsciously. Emotional responses to the patient's presentation, as well as cognitive heuristics and intuitions, can also be triggered at the same time as the pattern-recognition response.

If salient features of the presentation are not recognised as part of a meaningful pattern by the diagnostician, the analytical mode of reasoning may be triggered. A slower, more linear process of data examination proceeds, characterised by reliance on education and training, conscious critical thinking, logic, and rationality.

Importantly, the modes of reasoning may both be activated and interact with one another, leading to a “blended” (Croskerry, 2009a, p1025) output. Analytical reasoning may trump non-analytical reasoning; the monitoring and metacognitive function of analytical reasoning can apply a “rational override” (Croskerry, 2009a, p1024) to force reassessment of an initial, intuitive response on the part of the diagnostician. Errors may

arise when this monitoring function is compromised through fatigue, distraction, or inattentiveness. Alternatively, non-analytical reasoning can trump analytical reasoning. A doctor may consciously consider international guidance, statistical knowledge or decision rules, and choose to ignore these in favour of their own intuition when faced with a patient in real practice, away from the academic setting in which they were educated. Overconfidence, which has been explored in the literature as a major source of diagnostic error (Berner & Graber, 2008), may be understood as a basic mechanism by which such a “dysrationalia override” (Croskerry, 2009a, p1024) functions; some individuals hold an irrational belief that they can do better for the patient than clinical guidelines formulated on solid international research – guidelines of which they are consciously aware at the time of diagnosis. Research indicates that doctors’ judgement of their own diagnostic accuracy is poorly calibrated and that it is difficult to know when one is wrong (Berner & Graber, 2008; Podbregar et al., 2001). This is the problem of error-blindness: “it [feels] like something to be wrong; it feels like being right” (Schulz, 2011, p18).

As mentioned above, specific dual process models from the psychological literature are seldom invoked in the clinical reasoning literature; however, the widely-endorsed model described above arguably most closely resembles Stanovich and West’s two-systems theory (Stanovich & West, 2000). This model was popularised in the early 2000s, around the time when diagnostic error became a specific focus of the patient safety movement. The model’s authors posit that individuals respond differently to tasks depending on their conceptualisation of the task, prior knowledge, cognitive ability, and disposition. This is clearly seen in Croskerry’s popular model of dual processes in diagnosis, wherein familiar or unfamiliar features and patterns in the patient’s presentation are theorised to cue different cognitive responses.

This conceptualisation of the model's application to the diagnostic process has gained significant currency in the medical literature; it was prominently included in the IOM's 2015 report as a central framework by which the process may be understood (Balogh et al., 2015). In the literature on diagnostic reasoning, evidence from psychology and behavioural economics research in support of the dual process model is often cited, but there seems to be very little, if any, evidence from within the medical field regarding its application to diagnosis specifically. While there is much discussion of the model and its implications, and a growing programme of related interventions being tested (see Section 2.6.3 below), the evidence for the validity of the model's unique processes, as well as conceptual consideration of ontological concerns implicit therein, has not yet caught up with the model's general endorsement in practice, education and policy.

One notable exception is Moulton and colleagues' work on 'slowing down when you should', which illuminates the process by which doctors and surgeons transition from routine practice to effortful, deliberate practice (Moulton, Regehr, Lingard, Merritt, & MacRae, 2010b; Moulton, Regehr, Mylopoulos, & MacRae, 2007). A study of this phenomenon among surgeons (Moulton et al., 2010b) revealed that while some instances of slowing down are planned proactively in response to procedure- or patient-specific factors, others are initiated spontaneously in response to situational factors. Surgeons described the process as one of "regrouping" and "reassessing" (p1022), and, importantly, commented that this shift can happen without conscious awareness, only becoming clear on reflection after the fact. This has clear parallels with the rational override function proposed in Croskerry's conceptualisation of the model and provides an important example of how other aspects of the model may be validated going forward. (This process is considered further in Chapter 4.)

2.5.3 *Heuristics and biases*

The dual process model has been used as a framework to account for errors in diagnostic reasoning, and attention has overwhelmingly been paid to errors arising from the non-analytical mode of reasoning. Of the non-analytical psychological processes and mechanisms, heuristics and biases have received most attention in the extensive commentary, though the attendant empirical work is more limited (Croskerry, 2003b; Redelmeier, 2005).

A significant research tradition in psychology (Baron, 2007; Kahneman, 2011) demonstrates that heuristics, or mental rules of thumb, may be used in the non-analytical mode to reach fast decisions with approximate accuracy. When a heuristic fails, however, it induces a cognitive bias, or a predisposition to make judgements that lead to incorrect conclusions. Traditionally, the use of such non-analytical reasoning has been regarded as a trade-off between accuracy and expediency, and a number of reviews have compiled taxonomies of heuristics leading to biases that have been implicated in faulty diagnostic reasoning (Croskerry, 2003b; Crumlish & Kelly, 2009; Saposnik, Redelmeier, Ruff, & Tobler, 2016). While some commentators query the utility of such taxonomies and whether each heuristic genuinely represents a distinct phenomenon, it is worth describing some of the major heuristics and biases and their relevance to diagnostic error.

The *availability* heuristic is observed when a doctor's assessment of the probability of an event occurring is determined by the ease with which examples come to mind. An availability error occurs where the probability of a diagnosis is overestimated because it comes easily to mind, or underestimated because it does not. Diagnoses may come to mind easily because they are very common, well publicised, or very unusual and memorable (Elstein, 1999). For example, a doctor reviewing a patient

with a headache may overestimate the likelihood of a subarachnoid haemorrhage if they have recently seen such a case (Crumlish & Kelly, 2009). This heuristic is powerful and can be very useful, as both common diagnoses and diagnoses that are uncommon but very serious are likely to come to mind with ease. However, it can lead to a missed diagnosis due to non-availability of a less common condition (Waddington & Morley, 2000).

The *representativeness* heuristic is observed when a doctor's thinking is guided by a prototype, and so a given diagnosis is not considered unless the presentation is prototypical. A representativeness error occurs where the absence of prototypical features of a disease leads to an atypical variant of the disease being missed (Croskerry, 2003b). For example, a doctor can easily diagnose pulmonary embolism where a patient presents with pleuritic chest pain of acute onset with dyspnoea following a deep venous thrombosis, but may not consider the diagnosis in the absence of severe chest pain. Similarly, a diagnosis may be missed when the patient is atypical, for example, a young woman presenting with myocardial infarct. This heuristic is helpful in that it allows for quick diagnosis through pattern recognition, an approach that is increasingly being adopted by medical educators; however, an atypical presentation of a disease, particularly where comorbidity is an issue for a patient, can lead to a missed diagnosis (Croskerry, 2002).

Anchoring and adjustment describes a process whereby individuals make conscious judgements of probability by taking an initial starting point (the anchor), which may be arbitrary, and making revisions up or down around that starting point (adjustment). An *anchoring bias* occurs where the anchor is arbitrary or misleading or where the adjustment is insufficient (Elstein, 1999). In diagnosis, errors owing to this heuristic can occur when a doctor fixates on specific features of a presentation too early

in the diagnostic process, or fails to adjust an early hypothesis in light of new information. For example, a patient may receive an incorrect early diagnosis that ‘sticks’ (diagnostic momentum), as a working hypothesis easily becomes a *de facto* diagnosis even if no new information is acquired to confirm it, or if new information is acquired that contradicts it (Croskerry, 2002). While first impressions often contain important information, particularly for experienced clinicians with highly developed pattern recognition skills, it is difficult for people to move from incorrect first impressions and to disregard information that has been proven false or irrelevant.

A related error that compounds other biases is *base rate neglect*, which occurs where the underlying statistical prevalence of a disease is ignored or not adequately taken into account by a doctor. In some cases, this may be a deliberate choice, for example to rule out a worst-case scenario (Croskerry, 2003a); in other cases, it may be due to statistical illiteracy or misunderstanding of health statistics (Gigerenzer et al., 2007). For example, a doctor may assign all possible diagnoses equal likelihood in an attempt to review their options comprehensively, or they may not have a solid understanding of the sensitivity, specificity and false positive rates for a given diagnostic test. While many doctors routinely pursue rare diagnoses as part of a strategy to rule out the worst-case scenario for a patient, this can lead to over-utilisation and waste of resources, and the intermittent reinforcement of this behaviour can maintain it in some doctors (Croskerry, 2002).

An important concept arising out of this literature is *premature closure*, which occurs when cognitive heuristics lead a doctor to adopt the first solution that seems to fit the presentation and stop the search for reasonable alternatives before all of the relevant information has been considered. Premature closure is the single most common cognitive bias implicated in diagnostic error (Graber et al., 2005) and has arguably

received the most extensive empirical study. Experienced and junior doctors are equally likely to display premature closure (Kuhn, 2002), and elderly doctors may be particularly susceptible (Eva, 2002) as patterns become well-entrenched over many years.

Confirmation bias refers to the tendency to seek out only information that supports an initial hypothesis, and to neglect to seek information that may refute it or information that allows the testing of alternative hypotheses. Similarly, confirmation bias may lead a clinician to selectively interpret information acquired after the formation of the hypothesis in a way that supports it (Crumlish & Kelly, 2009). Besides the potential for missed or incorrect diagnoses, confirmation bias also carries implications for resource utilisation; ordering additional tests with the intention of gathering additional evidence in favour of the current hypothesis may bolster diagnostic confidence, but the additional data does not necessarily offer increased informational value or allow for efficient testing of alternative hypotheses (Elstein, 1999). Crucially, as clinicians incrementally gather additional evidence in support of their diagnostic hypothesis without considering alternative diagnoses, they may gain confidence without achieving greater accuracy (Oskamp, 1965).

Hindsight bias refers to the influence of current knowledge of outcomes on the perception of past events (particularly the overestimation of the probability that the outcome would have occurred), which has the potential to prevent an accurate appraisal of what has happened (Croskerry, 2003b). It is of particular importance in the study of diagnostic error using retrospective techniques, as mentioned above, and also in the ability of doctors and teams to learn from errors. When the ultimate correct diagnosis is known, cues to the correct diagnosis are judged as being more evident than they appeared at the time, and causal explanations for error are generated after the fact, often

based on fragmented information, creating a “coherent causal framework” that makes narrative sense of the process (Wears & Nemeth, 2007, p207). Hindsight bias can lead to overestimation (illusion of control) or underestimation (illusion of failure) of a diagnostician’s abilities, compromising calibration of diagnostic confidence and the potential to learn from experience of error and success (Croskerry, 2003b).

While the attribution of diagnostic errors to such mental shortcuts has intuitive appeal, the truth is likely to be more complex. Non-analytical reasoning has arguably been unfairly vilified: “To call a cognitive process a bias is ... tantamount to calling for its elimination or, at a minimum, minimising the role it plays in our reasoning” (Elstein, 1999, p793). In recent years, the discourse has begun to rightly acknowledge that the successful use of pattern recognition and heuristics is essential to the delivery of effective and timely medical care, and that successful non-analytical reasoning is recognised in other industries as a hallmark of mature decision-making, distinguishing experts from experienced non-experts (Moulton et al., 2007; Quirk, 2006). Indeed, excessive reliance on analytical reasoning can lead to over-utilisation of tests and resources, constant second-guessing, paralysis-by-analysis, patient anxiety due to doctors’ expressions of uncertainty, and new opportunities for error as additional investigative and treatment options are explored (Scott, 2009).

In summary, cognitive error is implicated in up to three-quarters of cases of diagnostic error, which may negatively impact up to 15% of patients. The dominance of the dual process model in the literature around diagnostic reasoning emerged against a background of increased attention on medical error in general and diagnostic error in particular. The model conceptualises the diagnostic process as utilising analytical or non-analytical modes of reasoning, and heuristics and biases arising from the non-analytical mode have drawn particular attention as causes of diagnostic error and,

therefore, potential targets for corrective interventions at the level of the individual diagnostician. Given the lack of empirical work devoted to the validation of the model itself, it is now important to critically review the evidence for the interventions to which it has given rise.

2.6 Interventions for diagnostic error

A broad range of strategies has been suggested to reduce diagnostic error, and the empirical research examining these interventions is rapidly gathering pace (Croskerry, Singhal, & Mamede, 2013a, 2013b; Graber et al., 2012; McDonald et al., 2013; Singh et al., 2011). They may be broadly categorised as systems-related interventions and cognitive interventions.

2.6.1 System-related interventions

The number of suggestions in the literature for systems-level interventions far outstrips the number of empirical studies actually carried out to examine their effectiveness; interventions at this level come with additional costs and logistical challenges, compared to efforts at the level of the individual doctor. However, systems approaches to error reduction have the potential to decrease error for all healthcare providers within the system over an extended period of time, and expansion of research in this area will be vital in the continued effort to reduce error (Kohn et al., 1999; Singh et al., 2011).

Some research studies have examined efforts to improve the interaction between doctor and patient, to implement electronic systems for results delivery, and to improve follow-up and tracking (e.g., around communication of abnormal results, planned follow-up, test-tracking), though the number of studies is small and they are largely

non-experimental or quasi-experimental in nature (Singh et al., 2011).

System-related interventions come with some disadvantages. As systems are constantly changing in response to new technology and advances in medical knowledge, improvements aimed at reducing error necessarily degrade over time and system repairs play ‘catch-up’ to system advances (Graber et al., 2002). They can also introduce new opportunities for error, and incur inherent trade-offs; improvements in one area of a system can have consequences in other areas, through redirected resources of work hours, personnel, or finances, or additional complexity. For example, physicians suffer less fatigue when shift lengths are limited, but new problems may arise from an increased number of hand-overs during a 24-hour period and more pressure on coordination of care.

While cognitive errors can be described at the level of the individual physician and their faulty cognitive processes, many such errors are induced by system factors and could be ameliorated by improvements at that level (Graber et al., 2002). Unclear or misleading presentation of test results and patient information, lack of availability of expertise, and poor statistical literacy among physicians (Gigerenzer et al., 2007) are all cognitive problems for which system-level solutions can be adopted.

2.6.2 Cognitive interventions

Cognitive interventions to improve diagnostic error at the level of individual clinician have also been studied to some extent. A large number of studies have examined interventions to encourage doctors to seek external assistance during the diagnostic process, with generally promising results. Sources of external assistance may include second opinions from other clinicians, support from non-clinical staff, such as technicians and librarians, or a range of computerised decision support tools and

differential diagnosis generators (Graber et al., 2012).

Interventions to increase knowledge and experience in particular areas have also shown promise; training in specific content areas, simulation to expose doctors to a greater number and variety of illness presentations, provision of intense, detailed, rapid, and specific feedback, and training and testing competency in diagnostic accuracy have all been examined in the literature (Graber et al., 2012; McDonald et al., 2013).

2.6.3 Dual process interventions

Interventions based on the dual process model are of particular interest, and are considered in greater depth in Chapter 3. By way of introduction, researchers and theorists have recommended a range of options for medical students and doctors to enhance analytical and non-analytical reasoning. While some lip service has been paid to the notion that errors can arise from either mode of reasoning, efforts are overwhelmingly targeted at reducing errors of non-analytical reasoning, primarily by prompting doctors to switch to a more reflective, analytical style of reasoning.

Educational curricula to enhance awareness of reasoning styles and understanding of psychological theories are an important category of interventions in this regard (Hershberger, Markert, Part, Cohen, & Finger, 1997; Reilly, Ogdie, von Feldt, & Myers, 2013); such programmes carry the assumption that with greater understanding and awareness of reasoning styles, and particularly of cognitive biases, diagnosticians will be better able to engage in metacognition, reflect on their own diagnostic behaviours, and ‘catch’ errors arising from faulty non-analytical reasoning (Trowbridge, Dhaliwal, & Cosby, 2013). However, the research to date does not appear to support this assumption (see Section 3.3.2.1).

Checklist interventions have shown success in other high-risk, high-reliability

fields (e.g., aeronautical engineering, nuclear power) and in non-diagnostic areas of medicine (e.g., procedural skills, surgery) (Ely, Graber, & Croskerry, 2011). Their use in diagnosis encourages a switch from fast, intuitive reasoning to more rigorous and thorough examination of the features of a clinical scenario, characteristic of slower, analytical processing. Similarly, ‘slowing down when you should’ (Moulton et al., 2007) and diagnostic timeouts provide an opportunity for doctors to adopt a more reflective reasoning style (Trowbridge, 2008).

Cognitive forcing strategies have been proposed in the literature with the aim of counteracting specific heuristics; these involve a diagnostician consciously applying a metacognitive step during diagnosis in order to avoid pitfalls and examine their own diagnostic behaviour. Cognitive forcing strategies deliberately broaden the diagnostician’s thinking and attempt to counteract important cognitive biases; for example, a forcing strategy whereby a clinician asks themselves “what else could this be?” ameliorates the risk of premature closure, or querying “what’s missing?” or “what doesn’t fit?” could counteract a representativeness bias (Croskerry, 2003a; Rajkomar & Dhaliwal, 2011).

These studies reflect the general acceptance of the dual process model in understanding the mechanisms underlying diagnostic reasoning. However, questions remain about how firmly rooted these studies are in psychological theory. First, as discussed earlier, the version of the dual process model that is frequently invoked in the diagnostic reasoning literature is often quite general, reflecting broad conceptualisations of System 1 and 2 reasoning more than any specific model. Croskerry's dual process model of diagnostic reasoning is highly specified (Croskerry, 2009a) and invoked by major publications in the field (e.g., Balogh et al., 2015), but is relatively recent and is not always cited as the particular model under examination in empirical studies, and its

specific features (e.g., rational override) are seldom described by researchers as the targets of their interventions.

Second, a majority of studies position increased reliance on analytical reasoning as a solution to errors arising from non-analytical reasoning. However, as outlined above, non-analytical reasoning provides essential benefits in diagnosis, and a combination of the two strategies is most likely to occur and to be successful (Norman & Eva, 2003). From this perspective, a simple ‘debiasing’ approach that solely encourages increased use of analytical reasoning through insight into unconscious processes is unlikely to be theoretically suitable or practically successful; rather, encouraging metacognitive skills and judicious switching between the two modes of reasoning (i.e., ‘slowing down when you should’ (Moulton et al., 2010b)) may be more realistic (see Chapter 4). The advantages of non-analytical reasoning have been widely acknowledged in psychology for some time (Gigerenzer, Czerlinski, & Martignon, 2002); it is possible that the relative lack of discussion of this point in the diagnostic reasoning literature reflects a somewhat shallow grounding of research in theory, and an approach that is more problem-driven than theory driven.

This is perhaps unsurprising; research in medical fields, including health services research, has been characterised as frequently being atheoretical, focusing primarily on evaluating outcomes rather than mechanisms of change (Brazil, Ozer, Cloutier, Levine, & Stryer, 2005). The extent to which interventions for diagnostic reasoning are rooted in dual process theory has not been systematically examined. This is worth considering, however, as grounding research in theory offers profound benefits. An intervention may yield improvement in the target outcomes, but in the absence of theory, mediational analysis becomes impossible; the researcher cannot identify how or why the effect occurred, or how to enhance it. In this way, the application of theory in

health services research offers pragmatic advantages. Theory helps researchers to identify core variables and key processes, clarify measurement issues, and propose testable predictions, and it offers clear rationale for studies and frameworks to guide analysis and interpretation of results (Michie & Prestwich, 2010; Michie et al., 2011).

2.7 Conclusion

In summary, diagnostic error is a significant concern for medical practice, and the dual process model has been adopted as a foundational model for understanding diagnostic error and how it may be countered. Importantly, the conclusion should not be that the majority of clinical decisions, whether reached by analytical or non-analytical reasoning, are seriously flawed (Elstein, 1999). Rather, by understanding the relative contributions of each mode of reasoning, system-related and cognitive interventions may be developed, studied, and implemented, and the body of research on interventions is growing rapidly. This is an exciting time for the field and there are many fertile areas for future research.

The dual process model represents a reasonably useful, if not perfect, method for analysing error, and has been overwhelmingly adopted by the medical community, though arguably without being subjected to rigorous evaluation and validation. Cognitive interventions for diagnostic reasoning are gathering pace, but it appears that the empirical research lags considerably in comparison to the enthusiasm for such work. In addition, the empirical research may not be as firmly rooted in the psychological theory underpinning the dominant dual process model as it may appear; the general approach in the literature is not fundamentally theory-driven, but problem-driven. It is unclear to what extent the interventions being proposed and examined actually reflect the psychological mechanisms being invoked to explain diagnostic behaviour.

On this basis, the first contribution of this thesis will be to undertake a systematic review of the existing interventions based on the dual process model, in order to identify the range of interventions that have been examined, synthesise the evidence for their effectiveness, and understand the underlying psychological mechanisms being targeted, using the specific lens of psychological science and dual process theory as an organising principle.

Chapter 3: Systematic Review of Dual Process Interventions for Diagnostic

Reasoning¹

3.1 Introduction

Researchers and theorists have recommended a range of options for medical students and doctors to enhance analytical and non-analytical reasoning. In a 2013 review (Crookery et al., 2013b) of theoretical and empirical work on debiasing strategies, two broad approaches are described: (1) educational strategies (techniques aimed at enhancing future decision-making through increased knowledge and awareness of reasoning styles; for example, educational curricula (Hershberger, Part, Markert, Cohen, & Finger, 1995; Reilly et al., 2013), simulation training (Bond et al., 2004), and instruction in Bayesian reasoning (Trowbridge, 2008)), and (2) workplace strategies (techniques aimed at enhancing decision-making in the moment, for example, checklists (Ely et al., 2011), cognitive forcing strategies (Crookery, 2003a), slowing down (Moulton et al., 2007), and diagnostic time-outs (Trowbridge, 2008)).

Interventions for diagnostic reasoning based on the dual process model of reasoning may fit conceptually under both categories. The model has achieved widespread acceptance as the dominant theoretical framework for conceptualising the diagnostic process and is relatively simple, making it easy to teach to novice diagnosticians and a likely candidate for inclusion in educational strategies to combat diagnostic error (Crookery, 2009b). Similarly, the model features processes (e.g., rational override) that may be consciously induced or implemented during diagnosis, opening up possibilities for experimental inquiry.

¹ A version of this chapter was published in BMJ Quality and Safety in 2016. The chapter expands on the paper by the inclusion of the meta-analysis and analysis of the use of theory in the included studies.

Interventions based on the dual process model were included in McDonald and colleagues' broad 2013 review (McDonald et al., 2013) of patient safety strategies for diagnostic errors. In its discussion of dual process interventions, the review primarily highlighted the value of educational interventions, which focused on improving content knowledge rather than reasoning *per se*, and interventions to improve diagnostic processes through increased structure, such as use of checklists and diagnostic algorithms.

Interventions based on this model were explored in greater depth in Graber and colleagues' comprehensive 2012 review of cognitive interventions to reduce diagnostic error (Graber et al., 2012). Graber et al. highlighted the debate in the literature around whether it is, in fact, possible to improve or change intuitive processes, which are by definition unconscious and instantaneous. They reviewed experimental studies of interventions for both analytical and non-analytical reasoning, finding that, while there is much enthusiasm in the literature for interventions of this sort, relatively few empirical studies have been carried out, and these have been largely limited to trainee populations (e.g., Mamede, van Gog, et al., 2010; Sherbino, Dore, Siu, & Norman, 2011). However, neither review focused in great detail on the theoretical foundations for these interventions or explored the extent to which theory informed their design; this is important, as theory-driven intervention design has profound benefits and offers fertile ground for hypothesis generation (Michie & Prestwich, 2010).

Given the above, a new systematic review of the literature on dual process interventions for diagnostic reasoning, using the dual process model as an organising framework and examining its place in the rationale for these interventions, is now required to advance the field.

3.1.1 Aims and objectives

This review adds to the literature by presenting an up-to-date synthesis (including some studies that have not appeared in previous reviews) and, most importantly, focusing in greater depth than any previous review on analytical and non-analytical thinking as an organising principle in considering this area, and offering a more granular analysis of this specific subset of studies. The review particularly focuses on the use of dual process theory within the included studies to provide clear conceptual rationale, identify core variables for examination, and guide analysis of outcomes. As mentioned in Section 2.6.3, the incorporation of theory as a conceptual basis for research design offers significant benefits; this review therefore aims to assess the extent to which the included studies have adopted this approach.

This review therefore aims to identify, describe and appraise studies that have empirically investigated interventions to enhance analytical and non-analytical reasoning among medical students and doctors at all stages of their careers, to assess their effectiveness in improving diagnostic accuracy, and to critically consider the use of dual process theory to provide clear conceptual rationale.

3.2 Method

3.2.1 Search parameters and inclusion criteria

The search process was guided by the PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009) for systematic reviews and documented in advance in a protocol. The following inclusion criteria were applied:

Participants: The search focused on efforts to improve analytical and non-analytical decision-making among medical trainees and doctors. The search included

only interventions aimed at the individual clinician and excluded system-level strategies, new or improved diagnostic tests or technologies, shared decision-making, patient decision-making, diagnostic decision-making by non-medical healthcare professionals (e.g., nurses, dentists, radiographers), or clinical-like decision-making by laypersons, in order to keep the range of studies relatively homogenous and to increase the relevance of conclusions for medical education in particular.

Interventions: The search targeted interventions to reduce cognitive-related diagnostic error, rather than studies of the origins or prevalence of such error. As the dual process model was the focus of the review, only interventions to enhance analytical and non-analytical reasoning were included; other styles of cognitive interventions (e.g., to increase content knowledge or experience, to provide general feedback on diagnostic accuracy, or to provide external assistance from other professionals, non-cognitive decision support tools or diagnosis generators) were excluded. All intervention formats were considered, including educational strategies and workplace strategies, as described above (Croskerry et al., 2013b), in order to capture the full breadth of existing interventions.

Comparisons: A control group, control condition, or baseline measure for comparison was required for inclusion, so that the specific effect of the intervention may be identified.

Outcomes: Educational intervention studies examining outcomes at any level of Kirkpatrick's adapted hierarchy (Barr, Freeth, Hammick, Koppel, & Reeves, 2000) were included, in order to capture the full breadth of existing interventions. This model categorises outcomes of educational interventions into six levels: Level 1 outcomes describe the learner's reaction to the intervention (e.g., satisfaction, acceptability); Level 2a describes changes in attitudes or perceptions; Level 2b describes the acquisition of

knowledge or skills; Level 3 describes transfer of behavioural changes from the learning environment to the workplace; Level 4a describes wider changes in the delivery of care; and Level 4b describes direct benefits to the patient (see Figure 3.1). Outcomes around diagnostic accuracy, resource utilisation, and testing behaviour were considered.

Study design: In order that the specific effect of the intervention may be identified, randomised controlled trials, quasi-randomised studies, within-subject studies (at least one control condition data point and one intervention condition data point, using a single participant group), between-subject studies (using treatment and control groups), and pretest-posttest (at least one data point before and one data point after the intervention phase, using intervention and control groups) studies were all considered, with no minimum length of follow-up. Commentaries, reviews, surveys, and audits were not included.

Other parameters: The search included only studies presented, published, or indexed after 1990, which takes into account the work that led to the publication of the Institute of Medicine's *To Err Is Human* (Kohn et al., 1999) and the work that emerged in its wake.

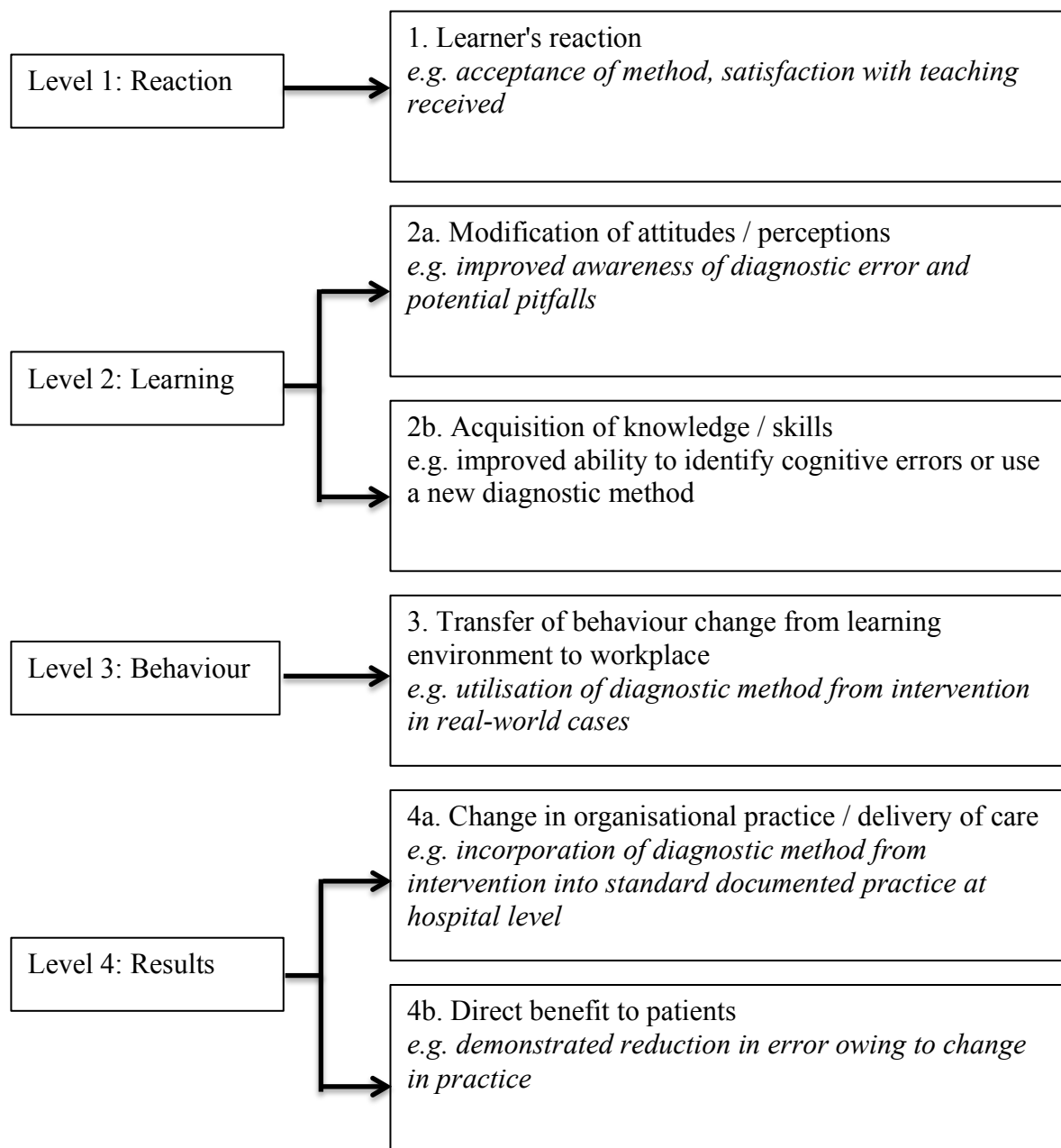


Figure 3.1 Kirkpatrick's adapted hierarchy of education outcomes

3.2.2 Search strategy

Five databases were searched: Medline (Ovid, 1990-present), PsycInfo (Ovid, 1990-present), Embase (Elsevier, 1990-present), Education Resource Information Centre (ERIC) (ProQuest, 1990-present), and the Cochrane Database of Controlled Trials (Wiley, 1990-present). A sample set of six papers on cognitive interventions for diagnostic reasoning (Coderre, Wright, & McLaughlin, 2010; Eva, Hatala, LeBlanc, & Brooks, 2007; Mamede, van Gog, et al., 2010; Mamede, Schmidt, & Penaforte, 2008;

Sherbino et al., 2011; Wolpaw, Papp, & Bordage, 2009) was compiled to facilitate the database search. Two of these were included in the final analysis (Mamede, van Gog, et al., 2010; Mamede et al., 2008); the remainder were excluded due to the intervention not being based on the dual process model (Coderre et al., 2010; Wolpaw et al., 2009), an inappropriate population (Eva et al., 2007), or the lack of a control group (Sherbino et al., 2011). The search combined the most relevant thesaurus / controlled vocabulary terms (provided by the indexed papers in the sample set) and natural language text words (extracted from the abstracts and titles in the sample set). Search terms were selected for four domains: cognition (e.g., “intuition”, “metacognition”), medical decision-making (e.g., “diagnostic reasoning”, “medical error”), the populations of interest (e.g., “physicians”, “medical students”, “medical education”), and empirical studies (e.g., “interventions”, “experiments”). The main searches were conducted between April and September 2014, with a supplementary follow-up search in February 2015. The full search strategy for Medline is displayed in Table 3.1; full search strategies for the remaining databases are included in Appendix A. This database search was supplemented in four ways. First, the bibliographies of review articles, book chapters, the sample set of articles, and articles identified for inclusion were manually reviewed for additional papers. Second, the contents of two relevant journals (*Medical Education* and *Medical Decision Making*), chosen for their scope and frequency in the preliminary searches, were manually reviewed. Third, thirteen key researchers in the area were contacted to request their recommendations and any relevant unpublished work. Suggested citations were also received from two peer reviewers. Fourth, the ‘Similar Articles’ feature of PubMed was used to identify papers similar to those included.

Table 3.1

Search strategy for Medline (Ovid, 1990 to present)

Diagnosis	Cognition	Population	Studies
physician's practice patterns/ OR decision support systems, clinical/ OR decision support techniques/ OR judgment/ OR diagnosis/ OR clinical competence/ OR exp diagnostic errors/ OR exp medical errors/ OR problem solving/ OR diagnostic reasoning.mp OR clinical reasoning.mp OR reflective practice.mp OR reflective reasoning.mp OR diagnostic accuracy.mp OR cognitive adj2 strateg\$.mp OR reasoning adj strateg\$.mp. OR metacogniti\$.mp OR diagnostic error\$.mp OR medical error\$.mp	cognition/ OR intuition/ OR thinking/ OR cogniti\$.mp OR cognitive bias\$.mp OR cognitive error\$.mp	exp students, medical/ OR exp medical staff/ OR internship and residency/ OR educational measurement/ OR exp education, medical/ OR teaching/ OR physician\$.mp OR doctor\$.mp OR clinician\$.mp OR medical student\$.mp OR consultant\$.mp OR intern.mp OR interns.mp OR resident.mp OR residents.mp	Evaluation studies/ OR longitudinal studies/ OR follow-up studies/ OR prospective studies/ OR cross-sectional studies/ OR comparative studies/ OR experiment\$.mp OR control\$.mp OR trial\$.mp OR intervention\$.mp OR effect.mp OR effects.mp OR effectiveness.mp OR efficacy.mp

Note:

/ indicates MeSH descriptor

.mp searches for term in Title, Original Title, Abstract, Subject Heading, Name of Substance, and Registry Word fields

\$ indicates truncation

adj searches for both terms next to each other in the order presented

adj 2 searches for both terms within two words of each other, in any order

exp indicates exploded search

3.2.3 Selection of studies and data synthesis

Titles were reviewed independently by the researcher and a research assistant, with manual searches of journals and reference lists performed by the researcher. The abstracts of articles flagged as potentially relevant were reviewed, and full texts of articles appearing to meet eligibility criteria were obtained for full review. The inclusion criteria were applied by the researcher and two doctoral supervisors; disagreements about relevance were resolved by discussion. Studies excluded due to the nature of the intervention were reviewed a second time, following feedback from peer review. The flow diagram (Figure 3.2) shows the results of the literature search. Twenty-eight studies meeting inclusion criteria were identified.

Information on the following variables was extracted from each included study using a form: setting, country, year of data collection/publication, study design, participant characteristics, sample size, intervention components, duration of follow-up, outcome measures, findings, and authors' recommendations. An additional form was used to extract information on each study's theoretical rationale (described narratively), mention of dual process models and elements of models, and specific dual process mechanisms targeted by the intervention (see coding frameworks in Appendix B).

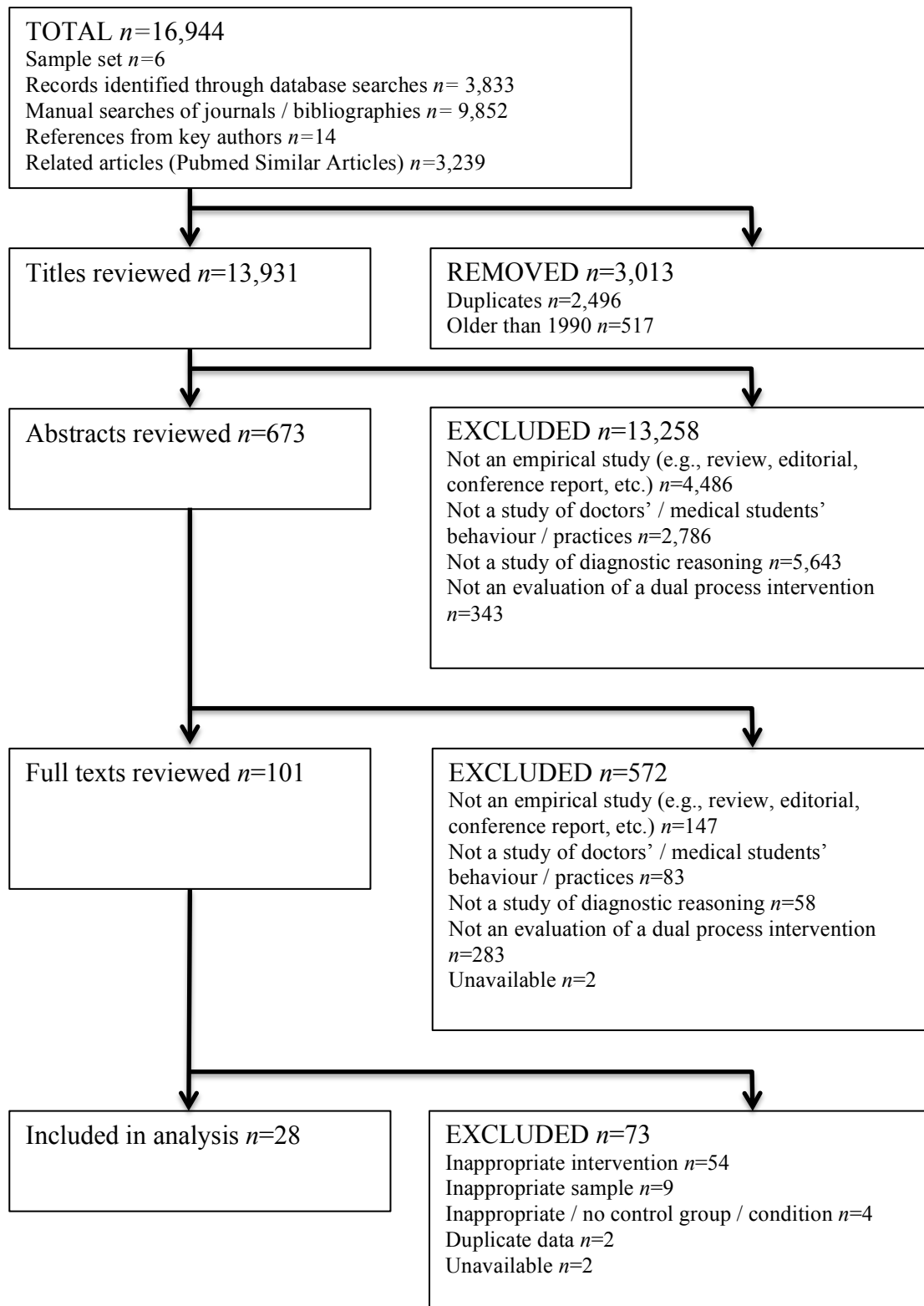


Figure 3.2. Literature search results and study flow

3.2.4 Risk of bias assessment

The Cochrane Collaboration's Tool For Assessing Risk of Bias (Higgins & Green, 2011) was used to assess the quality of the studies included. Five of the tool's six criteria were used: sequence generation, allocation concealment, blinding of outcome assessment, incomplete outcome data, and other source of bias. The remaining criterion, blinding of participants and personnel, was deemed inapplicable, as participants and personnel will by default be aware of whether they are receiving or delivering an intervention, respectively. The sequence generation and allocation concealment criteria were not relevant in every case, as some studies utilised within-subject designs. The researcher and a doctoral supervisor independently rated each criterion for each study as "not applicable", "high risk", "low risk", or "unclear", with 100% agreement.

3.2.5 Publication bias

With a view to minimising publication bias, multiple databases were searched and thirteen authors were contacted to seek unpublished work; five authors responded and three supplied references for review. Suggested citations were also received from two peer reviewers. No language restriction was imposed.

3.3 Results

3.3.1 Study characteristics

An overview of the twenty-eight included studies appears in Table 3.2. Two thousand, seven hundred and thirty-two participants took part across the twenty-eight studies.

Participants and setting: Nine studies included medical students only, ten included residents only, three included practicing doctors only, four included both medical students and residents, and two included medical students, residents and doctors. Seven studies were based in the USA, ten in Canada, four in the Netherlands, two in Brazil, and one each in the UK, Japan, Korea, Israel, and Switzerland.

Table 3.2

Characteristics of included studies

Author	Design	Participants	Experimental conditions	Outcome measure	Sample size	Findings
<i>Educational</i>						
(Hershberger et al., 1997)	Between-subjects	Medical students	(I/V) Seminar teaching on major biases / (C) No seminar	Evidence of heuristic use from performance on Inventory of Cognitive Bias in Medicine, pre- and post-intervention	<i>n</i> =118 intervention, <i>n</i> =112 control	Intervention significantly improved performance on ICBM
(Nendaz, Gut, Louis-Simonet, Perrier, & Vu, 2011)	Between-subjects	Sixth-year medical students	(I/V) Case-based reasoning seminars with emphasis on explicit reasoning steps, self-reflection and cognitive psychology concepts (C) Standard case-based reasoning seminars	Accuracy of final and differential diagnoses on two assessment cases	<i>n</i> =15 intervention, <i>n</i> =14 control	Intervention group performed better on differential diagnosis but no better on final accuracy of diagnosis
(Reilly et al., 2013)	Pretest-posttest	Residents	(I/V) Three-part, one-year curriculum on cognitive bias and diagnostic error, including lectures, group-based discussion, online interactive modules, and video vignettes (C) No curriculum	Performance on the Diagnostic error Knowledge Assessment Test (D-KAT) (13 items)	<i>n</i> =38 intervention, <i>n</i> =42 control	Knowledge of cognitive biases and strategies significantly improved compared to baseline and control group
(Round, 1999)	Between-subjects	Fourth-year medical students	(I/V) Seminar consisting of small group discussion followed by teaching on cognitive biases and Bayes Theorem (C) No seminar	Performance on Diagnostic Thinking Inventory	<i>n</i> =84 intervention, <i>n</i> =102 control	Intervention group significantly outperformed control group on DTI
(Schnapp & Shah, 2014)	Within-subjects	Residents	(I/V) 30-minute lecture on 10 key diagnostic errors	Error identification on 10 clinical vignettes at three-month follow-up	<i>n</i> =37 completed pre- and post-intervention	Significant improvement on only 3/10 error types

Note:

C: Control condition; I/V: Intervention condition; DTI: Diagnostic Thinking Inventory; ER: Emergency room; OSCE: Objective Structured Clinical Examination

Table 3.2 *cnt'd*

Author	Design	Participants	Experimental conditions	Outcome measure	Sample size	Findings
(Sherbino, Kulasegaram, Howey, & Norman, 2014)	Between-subjects	Senior medical students	(I/V) 90-minute interactive seminar on cognitive forcing strategies (C) No seminar	Evidence of heuristic use from instigation of search for secondary diagnosis on 6 clinical scenarios	<i>n</i> =145 intervention, <i>n</i> =46 control	Immediately after testing, <50% success showed debiasing, vulnerable to false positives. No difference between intervention and control groups.
<i>Checklist</i> (Graber et al., 2014)	Within-subjects	Staff-level ER physicians	(IV) Use of general and symptom-specific checklists on shift	Extensiveness of differential diagnosis on shifts, resource utilisation	<i>n</i> =15 completed pre- and post-intervention	Trends towards more items in differential ordering of more tests and consults with checklist use compared to baseline, though not statistically significant
(Shimizu, Matsumoto, & Tokuda, 2013)	Between-subjects	Senior medical students	(1) Instructions to interpret (C) intuitively, then (I/V) using general debiasing checklist (2) Instructions to interpret (C) intuitively, then (I/V) using differential diagnosis checklist	Diagnostic accuracy on 5 clinical scenarios	<i>n</i> =91 debiasing checklist, <i>n</i> =91 differential diagnosis checklist	Differential diagnosis checklist significantly improved diagnostic accuracy over intuitive reasoning; no effect of general debiasing checklist
(Sibbald, de Bruin, & van Merriënboer, 2013)	Within-subjects	Cardiology fellows	(1) Verification with checklist (2) interpretation and verification with checklist (3) Undirected interpretation (4) Interpretation and verification without checklist	Diagnostic accuracy on 18 ECGs	<i>n</i> =15 completed all conditions	Checklist use resulted in more correct interpretations, fewer errors in verification conditions and fewer errors overall
(Sibbald, de Bruin, Cavalcanti, & van Merriënboer, 2013)	Pretest posttest	Residents	(I/V) (1) Verification of diagnosis with checklist and ability to re-examine information (2) Verification of diagnosis with checklist and without ability to re-examine information (C) Pre-checklist diagnosis	Diagnostic accuracy on one of six clinical cases delivered by cardiopulmonary simulator	<i>n</i> =96 able to re-examine simulator, <i>n</i> =95 unable to re-examine simulator	Verification with a checklist improved diagnostic accuracy only for participants who could re-examine the simulator

Note:

C: Control condition; I/V: Intervention condition; DTI: Diagnostic Thinking Inventory; ER: Emergency room; OSCE: Objective Structured Clinical Examination

Table 3.2 *cnt'd*

Author	Design	Participants	Experimental conditions	Outcome measure	Sample size	Findings
<i>Cognitive forcing strategies</i>						
(Arzy, Brezis, Khoury, Simon, & Ben-Hur, 2009)	Pretest posttest	Attending doctors	(I/V) All groups instructed to diagnose cases, identifying one leading detail and re-diagnose after removing it (1) No warning of misleading detail (2) Prior warning of misleading detail (3) Trivial detail in place of misleading detail (C) Baseline diagnosis	Diagnostic accuracy on 10 clinical scenarios	<i>n</i> =17 warning, <i>n</i> =17 no warning, <i>n</i> =17 trivial detail	Warning doctors to be cautious of misleading details had no effect on diagnostic accuracy; significant reduction in error resulted from examination following removal of leading detail
(Feyzi-Behnagh et al., 2014)	Pretest posttest	Residents	(I/V) (1) Instructions to diagnose cases using Consider Alternatives scaffold; prompting participants to review their decision-making diagrams showing steps taken and correct steps (2) Instructions to diagnose cases using Playback scaffold; showing participants the steps they took in reaching their diagnosis, followed by the correct steps (C) Baseline diagnostic accuracy	Diagnostic accuracy on 8 clinical scenarios, accuracy of confidence judgements	<i>n</i> =16 control, <i>n</i> =15 intervention	Both groups demonstrated learning gain from pre-test to post-test; intervention group outperformed control group post-test in accuracy of confidence judgements
(Regehr, Cline, Norman, & Brooks, 1994)	Within-subjects	First-year residents	(I/V) Instructions to argue for alternative diagnoses (C) Instructions to diagnose cases based on first impression	Diagnostic accuracy on 12 slides of common dermatological diseases	<i>n</i> =32 completed control and intervention conditions	No statistically significant differences between intervention and control groups

Note:

C: Control condition; I/V: Intervention condition; DTI: Diagnostic Thinking Inventory; ER: Emergency room; OSCE: Objective Structured Clinical Examination

Table 3.2 *cnt'd*

Author	Design	Participants	Experimental conditions	Outcome measure	Sample size	Findings
<i>Guided reflection</i>						
(Mamede et al., 2008) (from sample set)	Within-subjects	Residents	(I/V) Instructions to interpret cases through guided reflective reasoning (C) Instructions to interpret cases intuitively ("the first diagnosis that comes to mind")	Diagnostic accuracy on 16 clinical scenarios	<i>n</i> =42 completed control and intervention conditions	No main effect of reflective practice condition, but some improvement for complex cases.
(Mamede, van Gog, et al., 2010) (from sample set)	Within-subjects	Residents	(I/V) Instructions to interpret cases through guided reflective reasoning (C) Instructions to interpret cases intuitively ("the first diagnosis that comes to mind")	Diagnostic accuracy on 8 clinical scenarios with bias manipulation	<i>n</i> =36 completed control and intervention conditions	Reflective reasoning improved diagnostic accuracy over automatic reasoning in cases where bias manipulation is present
(Mamede, Splinter, van Gog, Rikers, & Schmidt, 2012)	Within-subjects	Medical students, residents	(1) Conscious thought: Instructions to diagnose case following elaborate analysis of information (2) Unconscious thought: Instructions to diagnose cases after spending time solving anagrams (3) Instructions to diagnose cases based on first impression	Diagnostic accuracy on 12 clinical scenarios	<i>n</i> =84 completed all conditions	Reflective reasoning improved diagnostic accuracy over non-analytical reasoning residents only, and was associated with errors resulting from salient distracting clinical features
(Myung, Kang, Phyoo, Shin, & Park, 2013)	Between-subjects	Fourth-year medical students	(I/V) Instructions to interpret cases through guided reflective reasoning (C) No instructions	Diagnostic accuracy on 4 OSCE case vignettes	<i>n</i> =65 intervention, <i>n</i> =80 control	Intervention group significantly outperformed control group
(Schmidt et al., 2014)	Within-subjects	Residents	(I/V) Instructions to interpret cases through guided reflective reasoning (C) Instructions to diagnose cases quickly and accurately after exposure to relevant information earlier in the day	Diagnostic accuracy on 2 clinical scenarios primed to trigger cognitive bias and 2 comparison cases	<i>n</i> =38	Errors arising from cognitive bias triggered by prior exposure were reversed by guided reflection

Note:

C: Control condition; I/V: Intervention condition; DTI: Diagnostic Thinking Inventory; ER: Emergency room; OSCE: Objective Structured Clinical Examination

Table 3.2 *cnt'd*

Author	Design	Participants	Experimental conditions	Outcome measure	Sample size	Findings
<i>Instructions at test</i>						
(Ilgen et al., 2011)	Within-subjects	Senior medical students, junior and senior residents	(I/V) Instructions to diagnose cases using 'directed search' approach prompting structured analytical responses (C) Instructions to diagnose cases based on first impression	Diagnostic accuracy on 6 simple and 6 complex clinical scenarios	<i>n</i> =115 completed control and intervention conditions	Across total sample, intervention group significantly outperformed control group
(Ilgen et al., 2013)	Between-subjects	Senior medical students, residents, faculty clinicians	(I/V) Instructions to diagnose cases using 'directed search' approach prompting structured analytical responses (C) Instructions to diagnose cases based on first impression	Diagnostic accuracy on 12 clinical scenarios	<i>n</i> =201 control, <i>n</i> =192 intervention	No statistically significant differences between intervention and control groups
(Kulatunga-Moruzi, Brooks, & Norman, 2001)	Pretest-posttest	Medical students	(I/V) Instructions to diagnose cases using clinical features list (C) Instructions to diagnose cases based on first impression	Diagnostic accuracy on 12 slides of common dermatological diseases	<i>n</i> =16; randomised into two conditions	Control group outperformed intervention group when diagnosing similar cases at timepoints
(Kulatunga-Moruzi, Brooks, & Norman, 2011)	Between-subjects	Medical students	(I/V) Instructions to diagnose cases (1) based on first impressions, and then by listing all clinical features (2) based on first impressions only (3) by listing all clinical features only (C) Instructions to diagnose cases with no instructions	Diagnostic accuracy on 22 slides of common dermatological diseases	<i>n</i> =18 using clinical features, <i>n</i> =21 combined, <i>n</i> =12 no instruction	No significant differences between clinical features only and first impressions only. Instructions improved accuracy over no instructions, though did not reach statistical significance
(Norman et al., 2014)	Between-subjects	Residents	(I/V) Instructions to diagnose reflectively and thoughtfully (C) Instructions to diagnose quickly, without errors	Diagnostic accuracy on 20 case vignettes	<i>n</i> =108 intervention, <i>n</i> =96 control	No differences in accuracy; intervention condition took longer to reach diagnosis incorrect answers took longer than correct answers

Note:

C: Control condition; I/V: Intervention condition; DTI: Diagnostic Thinking Inventory; ER: Emergency room; OSCE: Objective Structured Clinical Examination

Table 3.2 *cnt'd*

Author	Design	Participants	Experimental conditions	Outcome measure	Sample size	Findings
(Sibbald et al., 2012)	Between-subjects	Fellows, residents, medical students	(I/V) Instructions to use dual processing reasoning (C) No instructions to use any specific reasoning approach	Diagnostic accuracy on 8 cardiopulmonary simulator cases	<i>n</i> =26 experienced fellows (<i>n</i> =12 control), <i>n</i> =13 residents (<i>n</i> =6 control), <i>n</i> =23 medical students (<i>n</i> =12 control)	No overall effect of instruction
(Sibbald & de Bruin, 2012)	Within-subjects	Residents	(I/V) Cases reinterpreted with instructions to approach diagnosis with (1) the same reasoning style as reported in control phase (potentiating) or (2) an alternative reasoning style to one reported in control phase (balancing) (C) Baseline free choice of reasoning style	Diagnostic accuracy on 8 ECGs (each interpreted twice)	<i>n</i> =24 completed control and intervention conditions	No differences observed between potent and balancing instructions; improvement in initial diagnosis only with instructions to analytical reasoning.

Note:

C: Control condition; I/V: Intervention condition; DTI: Diagnostic Thinking Inventory; ER: Emergency room; OSCE: Objective Structured Clinical Examination

Table 3.2 *cnt'd*

Author	Design	Participants	Experimental conditions	Outcome measure	Sample size	Findings
<i>Other interventions</i>						
(Ibiapina, Mamede, Moura, Elói-Santos, & van Gog, 2014)	Within-subjects	Senior medical students	(I/V) Instructions to diagnose cases using (1) modelled reflection or (2) cued reflection (C) Instructions to diagnose case quickly and accurately	Diagnostic accuracy on eight clinical scenarios	<i>n</i> =39 modelled reflection group, <i>n</i> =39 cued reflection group, <i>n</i> =39 free reflection group	Modelled and cued reflection groups outperformed free reflection group, with significant differences between modelled cued groups.
(Mamede, Schmidt, et al., 2010)	Within-subjects	Fourth-year medical students, residents	(I/V) (1) Conscious thought: Instructions to diagnose case following elaborate analysis of information; or (2) Unconscious thought: Instructions to diagnose case after spending time solving anagrams (C) Instructions to diagnose cases based on first impression	Diagnostic accuracy on 12 clinical scenarios	<i>n</i> =84 completed all conditions	Conscious deliberation improved accuracy for complex cases for experts; no effect experimental condition for simple cases Novices benefited from deliberation-without attention for simple cases
(Payne, 2011)	Pretest-posttest	Medical students, residents	(I/V) Participants provided with feedback on diagnostic accuracy after entering final diagnosis for sample clinical scenarios in learning phase (C) Participants provided with general disease information after entering final diagnosis for sample clinical scenarios in learning phase	Evidence of heuristic use from think-aloud protocol, eye-tracking; diagnostic accuracy on 15 clinical scenarios	<i>n</i> =20 intervention, <i>n</i> =20 control	No impact of metacognitive feedback intervention on use of heuristics or diagnostic accuracy

Note:

C: Control condition; I/V: Intervention condition; DTI: Diagnostic Thinking Inventory; ER: Emergency room; OSCE: Objective Structured Clinical Examination

Use of theory: Table 3.3 displays an overview of the elements of dual process theory incorporated into the studies and the specific mechanisms targeted by the interventions.

Educational interventions did not appear to target specific analytical or non-analytical processes; instead, they were based on the rationale that awareness alone of cognitive bias and reasoning strategies more generally may help to reduce error.

Checklist studies generally showed clear foundations in dual process theory and explicitly evoked the monitoring function of System 2 thinking; authors describe cognitive checklists as tools to encourage diagnosticians to examine specific aspects of their reasoning process (Graber et al., 2014; Shimizu et al., 2013) and to reduce cognitive load, leaving more resources available for more complete System 2 reasoning (Sibbald, de Bruin, Cavalcanti, et al., 2013).

The guided reflection studies showed a clear grounding in dual process theory in most cases, as did most of the other studies that involve instructing diagnosticians to proceed through elaborate reflective processes. The components of these reflective instructions generally include interrogation of an initial hypothesis, which would appear to invoke the rational / executive override function of System 2, along with evidence-gathering, case summarising, and instructions to consider important diagnoses that should not be missed. These components have some face value as inducing a deeper analysis of the case features, and Mamede and Schmidt (2004) describe the generation of alternative hypotheses and predictions based on these components as being aspects of inductive and deductive reasoning, key aspects of System 2 thought. However, the rationale behind the particular instructions given is not clearly explored in most of the papers that describe them.

Other studies generally attempted to induce System 2 reasoning, often framed in terms of metacognition, but often without specificity as to the processes being targeted. The monitoring function of System 2 was mentioned in a number of cases, but generally without invoking the language of dual process theory (e.g., executive / rational override).

Table 3.3
Use of the dual process model as theoretical basis for included studies

Lead author, year	Elements of dual process model mentioned	Specific dual process mechanisms targeted	General rationale
<i>Educational</i>			
(Hershberger et al., 1997)	Heuristics / biases	None	Awareness of bias may improve decision quality
(Nendaz et al., 2011)	Analytical / non-analytical reasoning	None	Explicit insight into processes and self-reflection may improve diagnostic accuracy
(Reilly et al., 2013)	Dual processes; Heuristics / biases	None	Education may improve knowledge of biases
(Round, 1999)	Heuristics / biases	None	Education on biases may improve clinical reasoning
(Schnapp & Shah, 2014)	Heuristics / biases	None	Education may improve knowledge of biases
(Sherbino et al., 2014)	Heuristics / biases	General metacognition	Metacognition may reduce errors due to cognitive bias
<i>Checklist</i>			
(Graber et al., 2014)	Heuristics / biases	System 2 oversight / monitoring; Interrogation of heuristic thinking	Opportunities to reflect and recognise situations prone to error may reduce error
(Shimizu et al., 2013)	Dual processes; Heuristics / biases; Systems 1 and 2	System 2 oversight / monitoring; Interrogation of heuristic thinking	Metacognition may allow errors to be caught and corrected
(Sibbald, de Bruin, & van Merriënboer, 2013)	Systems 1 and 2	System 2 oversight / monitoring; Cognitive load	Checklists induce reflection, which may reduce effect of biases

Table 3.3 *cnt'd*

Lead author, year	Elements of dual process model mentioned	Specific dual process mechanisms targeted	General rationale
(Sibbald, de Bruin, Cavalcanti, et al., 2013)	Systems 1 and 2	System 2 oversight / monitoring; Cognitive load; Repair / verification	Checklists create extra demands on System 2 reasoning, which may or may not be beneficial
<i>Cognitive forcing strategies</i>			
(Arzy et al., 2009)	Heuristics / biases	None	Misleading details can induce biases and advance warning may help avoid errors
(Feyzi-Behnagh et al., 2014)	Heuristics / biases; Metacognition	General metacognition; System 2 reasoning	Providing metacognitive feedback about the diagnostic process may help improve learning and accuracy judgements
(Regehr et al., 1994)	Analytical / non- analytical reasoning	System 2 reasoning	Analytical and non-analytical reasoning both have benefits
<i>Guided reflection</i>			
(Mamede et al., 2008)	Analytical / non- analytical reasoning	System 2 reasoning	Reflection (comprising hypothesis generation, evidence gathering, and ranking tasks) induces System 2 thinking and thereby reduces error
(Mamede, van Gog, et al., 2010)	Heuristics / biases; Systems 1 and 2	System 2 oversight / monitoring	Reflection (comprising hypothesis generation, evidence gathering, and ranking tasks) induces System 2 thinking and thereby reduces error arising due to cognitive bias
(Mamede, Splinter, et al., 2012)	Analytical / non- analytical reasoning	System 2 oversight / monitoring	Reflection (comprising hypothesis generation, evidence gathering, and ranking tasks) induces System 2 thinking and thereby reduces error arising from initial pattern recognition
(Myung et al., 2013)	Heuristics / biases; Analytical reasoning	System 2 oversight / monitoring	Reflection (comprising hypothesis generation, evidence gathering, and ranking tasks) induces System 2 thinking and thereby reduces error arising due to cognitive bias
(Schmidt et al., 2014)	Heuristics / biases	System 2 reasoning	Reflection (comprising hypothesis generation, evidence gathering, and ranking tasks) induces System 2 thinking and thereby reduces error

Table 3.3 *cnt'd*

Lead author, year	Elements of dual process model mentioned	Specific dual process mechanisms targeted	General rationale
<i>Instructions at test</i>			
(Ilgen et al., 2011)	Analytical / non-analytical reasoning	System 2 reasoning	Reflection (comprising case summarising, hypothesis generation, evidence gathering, ranking, and consideration of must-not-miss diagnosis) induces System 2 thinking and thereby can reduce error, depending on experience and case features
(Ilgen et al., 2013)	Dual processes; Systems 1 and 2; Heuristics / biases; Metacognition	System 2 reasoning	Reflection (comprising case summarising, hypothesis generation, evidence gathering, ranking, and consideration of must-not-miss diagnosis) induces System 2 thinking; Reasoning mode can be manipulated
(Kulatunga-Moruzi et al., 2001)	Analytical / non-analytical reasoning	System 2 reasoning	Consideration of clinical features induces analytical thinking; Diagnosis requires coordination of analytical and non-analytical processes
(Kulatunga-Moruzi et al., 2011)	Analytical / non-analytical reasoning	System 2 reasoning	Consideration of clinical features induces analytical thinking; Diagnosis requires coordination of analytical and non-analytical processes
(Norman et al., 2014)	Dual processes; Systems 1 and 2	System 2 reasoning	Analytical and non-analytical reasoning are both effective
(Sibbald et al., 2012)	Dual processes; Systems 1 and 2	Interrogation of initial hypothesis; Switching	Balance between analytical and non-analytical reasoning strategies is beneficial
(Sibbald & de Bruin, 2012)	Dual processes; Systems 1 and 2; Heuristics / biases	Interrogation of initial hypothesis; Switching	Balance between analytical and non-analytical reasoning strategies is beneficial
<i>Other interventions</i>			
(Ibiapina et al., 2014)	None	System 2 reasoning	Reflection (comprising hypothesis generation, evidence gathering, and ranking tasks) induces System 2 thinking and thereby reduces error
(Mamede, Schmidt, et al., 2010)	Dual processes; Analytical / non-analytical reasoning; Pattern recognition	System 2 oversight / monitoring	Reflection (comprising hypothesis generation, evidence gathering, and ranking tasks) induces System 2 thinking, allowing errors to be caught and corrected
(Payne, 2011)	Dual processes; Systems 1 and 2; Heuristics / biases	General metacognition	Metacognitive feedback reduces errors due to cognitive bias

Risk of bias assessment: Risk of bias assessments for all included studies are shown in Table 3.4. Details of randomised sequence generation, concealment of allocation and blinding of outcome assessment were scarce. Outcome data was complete or adequate in all cases. The heterogeneity of measurement techniques used is a significant issue; while eighteen of the studies used clinical vignettes in some form, these were not standardised. The reliability and validity of such measures is therefore difficult to ascertain.

Table 3.4

Risk of bias assessment of included studies

Lead author, year	Sequence generation	Concealment of allocation	Blinding of outcome assessment	Incomplete outcome data	Other source of bias
<i>Educational</i>					
(Hershberger et al., 1997)	unclear	unclear	unclear	low	high
(Nendaz et al., 2011)	unclear	high	unclear	low	low
(Reilly et al., 2013)	high	high	unclear	low	high
(Round, 1999)	high	high	unclear	unclear	unclear
(Schnapp & Shah, 2014)	n/a	n/a	unclear	low	high
(Sherbino et al., 2014)	high	high	unclear	low	unclear
<i>Checklist</i>					
(Graber et al., 2014)	n/a	n/a	unclear	low	low
(Shimizu et al., 2013)	n/a	n/a	unclear	low	low
(Sibbald, de Bruin, & van Merrienboer, 2013)	n/a	n/a	low	low	low
(Sibbald, de Bruin, Cavalcanti, et al., 2013)	low	high	unclear	low	high

Table 3.4 *cnt'd*

Lead author, year	Sequence generation	Concealment of allocation	Blinding of outcome assessment	Incomplete outcome data	Other source of bias
<i>Cognitive forcing strategies</i>					
(Arzy et al., 2009)	unclear	high	unclear	low	low
(Feyzi-Behnagh et al., 2014)	unclear	high	unclear	low	low
(Regehr et al., 1994)	n/a	n/a	unclear	low	low
<i>Guided reflection</i>					
(Mamede et al., 2008)	n/a	n/a	unclear	low	low
(Mamede, van Gog, et al., 2010)	n/a	n/a	low	low	low
(Mamede, Splinter, et al., 2012)	n/a	n/a	low	low	low
(Myung et al., 2013)	low	unclear	unclear	low	unclear
(Schmidt et al., 2014)	n/a	n/a	low	low	low
<i>Instructions at test</i>					
(Ilgen et al., 2011)	n/a	n/a	unclear	low	low
(Ilgen et al., 2013)	low	high	unclear	low	low
(Kulatunga-Moruzi et al., 2001)	unclear	high	unclear	low	low
(Kulatunga-Moruzi et al., 2011)	unclear	high	unclear	low	low
(Norman et al., 2014)	high	high	high	low	unclear
(Sibbald et al., 2012)	unclear	unclear	unclear	low	unclear
(Sibbald & de Bruin, 2012)	n/a	n/a	unclear	low	unclear
<i>Other interventions</i>					
(Ibiapina et al., 2014)	n/a	n/a	unclear	low	low
(Mamede, Schmidt, et al., 2010)	n/a	n/a	low	low	low
(Payne, 2011)	unclear	unclear	unclear	low	low

Publication bias: While many of included studies find some impact of interventions, not all achieve statistical significance; this suggests that publication bias based on statistical significance was not an issue. No formal analyses were conducted to estimate the risk of publication bias.

3.3.2 Qualitative synthesis of results

Studies were grouped by intervention style. There were two major categories of intervention: educational and workplace. Workplace interventions were further divided into four categories: checklists, cognitive forcing strategies, guided reflection, and instructions at test to use analytical or non-analytical reasoning. Three studies did not fit these categories; each is considered separately below.

3.3.2.1 Educational interventions

Six studies examined the impact of educational interventions. All examined outcomes at Level 2b of Kirkpatrick's adapted hierarchy (acquisition of knowledge or skills) (Barr et al., 2000). One study (Reilly et al., 2013) employed a longitudinal intervention consisting of a one-year curriculum delivered to residents, while the remainder were one-off seminars. One study included a three-month follow-up timepoint (Schnapp & Shah, 2014).

The measured outcomes varied considerably, with one study each examining diagnostic accuracy (Nendaz et al., 2011), evidence of heuristic use (Sherbino et al., 2014), and error identification on test cases (Schnapp & Shah, 2014). No significant impact of the interventions was found in any of these studies. Three studies each used an idiosyncratic measure of thinking styles: these were the Diagnostic Thinking Inventory (Round, 1999); the Diagnostic error Knowledge Assessment Test (D-KAT) (Reilly et al., 2013), a 13-item multiple-choice measure of knowledge of cognitive biases and strategies; and the Inventory of Cognitive Biases in Medicine (Hershberger et al., 1997). The latter two measures were composed and validated by the authors of the respective papers. Significant post-intervention gains were observed in each of these studies.

3.3.2.2 Workplace interventions

3.3.2.2.1 Checklists

Four studies examined the impact of checklist interventions. Two studies used a general diagnostic checklist (Sibbald, de Bruin, Cavalcanti, et al., 2013; Sibbald, de Bruin, & van Merrienboer, 2013), one used a general checklist and a symptom-specific checklist (Graber et al., 2014), and one used a debiasing checklist and a differential diagnosis checklist (Shimizu et al., 2013). Three studies (Shimizu et al., 2013; Sibbald, de Bruin, Cavalcanti, et al., 2013; Sibbald, de Bruin, & van Merrienboer, 2013) examined the impact of checklist use on diagnostic accuracy for clinical scenarios, while the fourth (Graber et al., 2014) examined its impact on extensiveness of differential diagnosis and resource utilisation during physicians' hospital shifts.

Findings were mixed. One study indicated that checklist use led to fewer errors overall and more correction of errors on verification (Sibbald, de Bruin, & van Merrienboer, 2013); one reported trends towards more extensive differential diagnosis, though this effect was not statistically significant (Graber et al., 2014); one reported that checklists were beneficial only where participants could review the content of the case (Sibbald, de Bruin, Cavalcanti, et al., 2013); and the last reported that a differential diagnosis checklist improved diagnostic accuracy over intuitive reasoning, though the effect was not observed for a debiasing checklist (Shimizu et al., 2013).

3.3.2.2.2 Cognitive forcing strategies

Three studies examined the impact of an intervention using cognitive forcing strategies on diagnostic accuracy for test cases.

Two studies (Feyzi-Behnagh et al., 2014; Regehr et al., 1994) instructed participants to consider alternative diagnoses. Only one of these found improvements in diagnostic accuracy compared to instructions to diagnose based on first impressions or without specific instruction; the same study also found that considering alternatives improved the accuracy of participants' confidence judgements (Feyzi-Behnagh et al., 2014).

One study (Arzy et al., 2009) instructed participants to reconsider their diagnosis after removing a misleading detail from the case outline, resulting in a significant improvement in diagnostic accuracy; however, merely warning participants to be aware of misleading details had no impact on accuracy.

3.3.2.2.3 Guided reflection

Five studies examined the impact of interventions instructing participants to diagnose cases through a guided, structured reflective process, compared to instructions to diagnose cases quickly, based on their first impressions, or in the absence of instructions. The outcome measure in all five studies was diagnostic accuracy on test cases; four with clinical vignettes (Mamede, Schmidt, et al., 2010; Mamede et al., 2008; Mamede, Splinter, et al., 2012; Schmidt et al., 2014) and one with Objective Structured Clinical Examination (OSCE) cases (Myung et al., 2013).

All five studies revealed some impact of guided reflection on diagnostic accuracy. Two highlighted its utility in overcoming experimental manipulations to induce cognitive biases (Mamede, van Gog, et al., 2010; Schmidt et al., 2014). One study reported an effect for complex cases only (Mamede et al., 2008). Findings for medical student samples were mixed, with one study finding an effect (Myung et al., 2013) and another study finding none (Mamede, Splinter, et al., 2012).

It is worth noting that of these five studies, four (Mamede, Schmidt, et al., 2010; Mamede et al., 2008; Mamede, Splinter, et al., 2012; Schmidt et al., 2014) were conducted by significantly overlapping research teams, with two authors contributing to all four studies and two other authors contributing to three.

3.3.2.2.4 Instructions at test

Seven studies examined the impact of interventions instructing participants to use a particular reasoning approach. The outcome measure was diagnostic accuracy on clinical scenarios in five cases and on dermatology slides in two. Experimental manipulations varied significantly. Two studies compared the impact of instructions to diagnose through a ‘directed search’ protocol to instructions to diagnose based on first impressions (Ilgen et al., 2011, 2013); one study compared instructions to use dual process reasoning to no instructions (Sibbald et al., 2012); one study compared instructions to use lists of clinical features to first impressions (Kulatunga-Moruzi et al., 2001); one study compared instructions to use dual processing, clinical feature-listing, first impressions and no instructions (Kulatunga-Moruzi et al., 2011); one study compared instructions to diagnose thoughtfully to instructions to diagnose quickly (Norman et al., 2014); and the final study compared free-choice reasoning to instructions to approach the case in a style that either matched or differed from the participant’s initial chosen approach (Sibbald & de Bruin, 2012).

Findings were mixed. Five studies found no statistically significant difference between instruction groups. One study reported that instructions to use an analytical approach improved accuracy, irrespective of whether this aligned with participants’ initial diagnostic approach (Sibbald & de Bruin, 2012). The two studies by Ilgen and colleagues are of particular note, as the ‘directed search’ protocol features instructions

similar to the guided reflection protocol outlined above, with additional instructions to briefly summarise the case and to identify a diagnosis that must not be missed in the case. One of these studies found that the ‘directed search’ protocol improved diagnostic accuracy (Ilgen et al., 2011); the other did not (Ilgen et al., 2013).

3.3.2.2.5 Other interventions

One study (Payne, 2011) examined the impact of a metacognitive feedback intervention on diagnostic accuracy and use of heuristics when diagnosing clinical scenarios; no effect was found.

One study (Ibiapina et al., 2014) examined the effects of free, cued and modelled reflection on diagnostic accuracy when diagnosing clinical scenarios. Diagnostic accuracy was higher under both the cued and modelled reflection conditions compared to free reflection, with no differences between these two conditions.

One study (Mamede, Schmidt, et al., 2010) compared the impact of conscious versus unconscious deliberation on diagnostic accuracy. Experts who interpreted a clinical case following an elaborate analysis outperformed those who diagnosed the case following a distractor task or based on first impressions; however, this effect held only for complex cases. By contrast, novices benefited from unconscious deliberation for simple cases only.

3.3.3 Quantitative synthesis of results

There was a sufficient number of studies in the guided reflection category to conduct a meta-analysis; as the sample was relatively small, Hedge’s *g* was chosen as the statistic. Data were pooled using Comprehensive Meta-Analysis (CMA; 2005). Heterogeneity among studies was high; therefore, a random effects meta-analysis model

was deemed most appropriate. Meta-analysis indicated a small ($g \approx 0.2$ (J. Cohen, 1988)) positive effect on diagnostic accuracy for participants using guided reflection compared to intuitive reasoning (Hedge's $g = 0.365$, 95% CI, 0.175 - 0.554) (Figure 3.3).

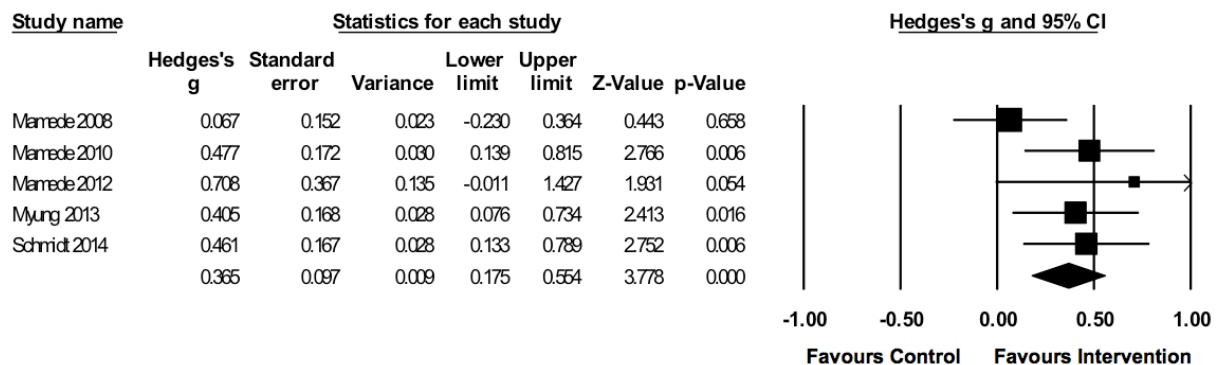


Figure 3.3. Forest plot with effect sizes (control vs intervention)

3.4 Discussion

3.4.1 Findings

The present review reveals a varied picture of the literature on dual process interventions for diagnostic reasoning. The guided reflection approach emerged as the most promising style of intervention, while interventions that provided less structured instructions or education on diagnostic reasoning appear less consistently successful. A number of authors contributed two or more studies to this review and, as mentioned, the research teams on the majority of the guided reflection studies overlapped significantly. Confidence in the effects observed would be greatly increased by replication by new research teams in other settings as the field expands.

The findings from the studies of instructions at test suggest that instructions to use some style of reasoning at the point of diagnosis may amount to little more than

exhortations to ‘think harder’, insufficient to alter cognitive behaviour. However, the sample of studies is small, and the patterns should therefore be interpreted with caution.

Studies of cognitive forcing strategies reveal an interesting split; an educational intervention on these strategies did not produce an effect, while studies examining instructions at test to use a specific cognitive forcing strategy offered mixed evidence for their efficacy. However, the small sample of studies limits the strength of any conclusions that may be drawn.

3.4.1.1 Use of theory

The positioning of theory in these papers is of particular interest and represents an area where the strengths of psychological science may be brought to bear on the practical work of quality improvement. As discussed in Chapter 2, while a relatively detailed conceptual model of dual diagnostic processes exists and has received broad acceptance in the medical community (Croskerry, 2009a), empirical verification of its unique features and the general validity of the model is lacking.

While most of the papers in this review offer some theoretical background and rationale for the interventions they investigate and make reference to the dual process model, cognitive biases, and related concepts, the discussion of this is often somewhat shallow. The vast majority attempt to induce System 2 reasoning, and few explicate the precise cognitive mechanisms the interventions are designed to target. Studies of checklists arguably make the most compelling case, positing that checklists elicit a System 2 override, encourage consideration of whether specific biases could have influenced a decision (Graber et al., 2014; Shimizu et al., 2013), and reduce information overload for diagnosticians (Sibbald, de Bruin, & van Merriënboer, 2013). However, although the language of System 2 monitoring and executive override is used,

Croskerry's 2009 conceptualisation of the dual process model is itself seldom directly cited.

Studies of educational and reflective interventions argue that awareness and metacognition broadly improve diagnosis; however, a recent review of reflective interventions considers how reflection at different temporal points within the diagnostic process may produce different benefits (Mamede & Schmidt, 2017). Generally, the interventions included in this review are rooted in dual process theory in broad terms, but the specific rationale behind the expected effects is not always clear and they offer little insight into the mechanisms by which they work.

Regarding the dual process mechanisms targeted by the interventions, System 2 reasoning and oversight / monitoring were overwhelmingly dominant in this sample of studies. For this reason, it is not possible to identify particular mechanisms that have successfully been targeted by interventions. Additionally, no clear association between firm theoretical rationale and successful interventions can be identified at this time (although checklist studies arguably stand out as being both broadly successful and firmly rooted in dual process theory, the number of studies is too small to identify the pattern definitively).

No studies were found that attempted to improve non-analytical reasoning. This is perhaps unsurprising; non-analytical processes are generally held to be unconscious, not under conscious control, and therefore difficult to access or manipulate. However, perhaps more significantly, while a majority of studies examined the potential impact of switching from non-analytical to analytical reasoning, only two (Sibbald et al., 2012; Sibbald & de Bruin, 2012) examined the potential impact of encouraging switching from analytical to non-analytical reasoning. Despite a number of papers acknowledging

the value of both modes of reasoning, the common-sense hypothesis that slowing down and ‘thinking harder’ will lead to better decisions appears to hold sway in this literature.

3.4.2 Methodological trends

3.4.2.1 Risk of bias

The risk of bias assessment revealed a mixed picture of the methodological robustness of studies, and many important details were missing from study reports. Outcome measurement approaches varied widely, with most studies using bespoke clinical vignettes, and the field lacks high-quality and broadly accepted tools to measure reasoning styles and error rates. Scales composed for this purpose have not been widely validated and the clinical vignette approach favoured by most researchers has limitations (see Section 3.4.4). Sample sizes differed widely across studies with little consideration in the reports of statistical power, and researchers still lack clear expectations around effect sizes. While many of these methodological concerns are understandable in the early phases of research in a field that is only now coming to prominence, they are worth highlighting as avenues for potential improvement in future studies.

3.4.2.2 Timeline

The literature contains many recommendations for various interventions (Graber et al., 2012), but relatively few empirical trials examining their effectiveness, and only one study in this review examined the lasting impact of interventions over a period longer than four weeks (Schnapp & Shah, 2014). The vast majority of the studies in this

review were presented or published in the last six years, suggesting that this is still a relatively immature field, but one that is currently accelerating.

3.4.2.3 Level of outcomes

The ultimate goal of researchers' efforts to enhance reasoning is arguably the improvement of patient outcomes. However, on Kirkpatrick's adapted hierarchy (Barr et al., 2000), none of the educational interventions measured outcomes at a level higher than Level 2b (improvement in knowledge or skills); that is, outcomes were overwhelmingly measured using clinical scenarios and vignettes, not real-world decision-making in the workplace. While clinical vignettes can accurately and reliably detect differences in physicians' performance in real settings (Peabody, Luck, Glassman, Dresselhaus, & Lee, 2000), the transfer of any effects to the real-world workplace, and ultimately to the level of patient outcomes, remains unexplored. Notably, only one study (Graber et al., 2014) examined the impact of an intervention in a hospital setting. This focus on lab-based efficacy studies is likely a product of the fact that this research is still in a relatively early, developmental phase; the challenge of identifying appropriate outcomes in implementation and patient safety work is discussed in more detail in Section 6.4.2.

3.4.2.4 Samples

It is possible that the reliance on medical students and residents in these samples plays a role in the results. The successful use of non-analytical reasoning is generally achieved with experience, and trainees earlier in their careers are more likely to use slower, more deliberate analytical methods (Coderre et al., 2003). In this way, students and residents, though a convenient sample for academic researchers, may not be the

most appropriate targets for interventions aimed at modifying dual process modes of reasoning; doctors later in their careers may respond differently to such efforts, and doctors with many years of experience are in a minority in these samples. Qualitative differences between novices and intermediates decision-makers have been demonstrated in other fields, including aviation (Wiggins, Stevens, Howard, Henley, & O'Hare, 2002). In medicine, students and residents also demonstrate different reasoning strategies in diagnosis (Ilgen et al., 2011), suggesting that the field would benefit from a more nuanced understanding of novice versus intermediate reasoning skills, expanding the usual distinction between novices and experts. This is particularly important when considering the introduction of interventions to educational curricula, as the timing of this training could profoundly impact its success. Therefore, developmental differences in reasoning constitute an important area for future research.

Specialty is also relevant: all of the studies included were in the area of general medicine, with cardiology being particularly well-represented. Mental health and psychiatry were conspicuously absent, which is important, as diagnosis in psychiatry is arguably quite different from diagnosis in general medicine. Psychiatric presentations may be regarded as inherently more uncertain, with increased reliance on self-report and a lack of objective tests available to diagnosticians (Pincus, 2014). Clear criteria for psychiatric diagnoses, such as those provided by the Diagnostic and Statistical Manual (DSM) and the International Statistical Classification of Diseases and Related Health Problems (ICD), are not consistently used by doctors (Meyer & Meyer, 2009; Morey & Benson, 2016), and patients with psychiatric presentations may be particularly at risk of diagnostic error (Croskerry, 2003b). These points will be considered in greater depth in Chapter 4 (see Section 4.1.3).

3.4.3 Limitations of the present review

3.4.3.1 Limitations of the search

The search was restricted to studies of medical professionals or trainees. As medical decision-making constitutes a special case of decision-making under uncertainty, it is likely that many valuable conceptual contributions and techniques may have been found in other disciplines, including economics, human factors engineering, and military scholarship. Much evidence exists to support the notion that deliberate thinking improves intuitive decision-making, and it is entirely probable that the same holds for medical decision-making (Moxley, Ericsson, Charness, & Krampe, 2012).

3.4.3.2 Limitations of included studies

The quality of methodology and reporting varied considerably across studies, somewhat hampering confident interpretation and synthesis. Medium- and long-term follow-ups in studies with educational components, evidence of adequate randomisation (see Table 3.4), reliable and validated measurement tools, and studies of experienced doctors were lacking. Effect sizes were reported in only four papers.

3.4.3.3 Limitations of the analysis

As with most meta-analyses, the intervention styles and outcome measurements were not consistent across studies. The relatively small number of studies, the heterogeneity of studies included in the review, and the smaller number still with appropriate and sufficient information for quantitative analysis all limit the strength of the conclusions that may be drawn, as the robustness of meta-analyses is reduced with smaller samples (Borenstein, Hedges, Higgins, & Rothstein, 2009).

3.4.4 Future research

A number of recommendations for future research can be made based on these findings. First, studies of more diverse populations of medics are required, particularly studies of the decision-making of very experienced doctors. All of the studies meeting inclusion criteria were in the area of general medicine; future research may examine whether similar interventions are successful with psychiatric diagnoses, which differ in important ways from general medical diagnoses but may still benefit from interventions such as those outlined in this review.

Second, further methodological refinement to ensure reliable and valid assessment of key outcomes, particularly in non-artificial settings, will lend strength to the body of work. The majority of studies use pen-and-paper-based vignettes and similar stimuli, which are static and do not allow for the sort of iterative, purposeful pursuit of information that characterises real-world clinical encounters. New techniques, such as high-fidelity patient simulation (Burbach, Barnason, & Thompson, 2015) or standardised patients (Maupomé, Schrader, Mannan, Garetto, & Eggertsson, 2010), offer researchers the opportunity to simulate the full complexity of diagnosis in real time, observing not only how doctors make sense of the information they are given but also how that information guides their search for more (Schubert, Denmark, Crandall, Grome, & Pappas, 2013).

Third, additional studies to confirm the effects of the most promising intervention styles detailed above, particularly by alternative research teams, are now needed. Further experimental study of guided reflection, checklists, and cognitive forcing strategies with diverse populations and more refined methodologies will aid

educators and designers of interventions in choosing strategies to promote and incorporate into educational curricula.

Fourth, the potential to improve the effectiveness of non-analytical processes and the potential value of switching between the two modes of reasoning in both directions constitute largely unexplored terrain. Accessing and manipulating non-analytical processes is difficult, due to their very nature as unconscious mechanisms, but improvements in how trainees build up their non-analytical and pattern recognition skills are possible through changes in medical education curricula and emerging computer and simulation technologies. These are exciting new possibilities for research in this area.

Finally, continued consideration of the role of theory will be important in guiding design of interventions and experiments alike. With greater appreciation of dual process theories and the specific psychological mechanisms by which diagnoses are reached, researchers will be better able to target intervention efforts and make sense of the outcomes.

3.4.5 Conclusions

This review reveals a burgeoning field of study. There is general enthusiasm for training in analytical and non-analytical reasoning, and trials of interventions have been emerging steadily over the last number of years. These have been accompanied by a wealth of opinion and commentary, though the extent to which this commentary is evidence-based and theoretically driven is unclear. The empirical work is therefore somewhat preliminary, and challenges include a lack of longitudinal work, standardised measurement tools, and research on experienced clinicians. Once-off styles of interventions using guided reflection, checklists, and cognitive strategies appear to have

some effectiveness in improving diagnostic performance, and this study has found modest evidence to recommend them to educators.

Chapter 4: Voluntary Use of Guided Reflection for Diagnostic Reasoning: A Think-Aloud Protocol Study

4.1 Introduction

The preceding chapter presented the results of a systematic review of dual process interventions for diagnostic accuracy. While many studies find some effect of interventions, guided reflection interventions, based on a model of reflective practice, emerged as the most consistently successful across five studies, while cognitive forcing strategies improved accuracy and confidence judgements.

While there has been some variation across a number of studies in the specific phrasing and framing of instructions, guided reflection interventions tend to follow a similar structure (Ilgen et al., 2011, 2013; Mamede, Schmidt, et al., 2010; Mamede, van Gog, et al., 2010, 2012; Mamede et al., 2008; Mamede, Splinter, et al., 2012; Myung et al., 2013; Schmidt et al., 2014). Participants are provided with a clinical scenario and asked to diagnose the patient using the following process:

- (1) Offer the first diagnosis to come to mind and re-examine the information in the clinical scenario
- (2) Find details in the scenario that support and refute the hypothesis
- (3) Offer alternative possible diagnoses
- (4) Find details in the scenario that support and refute each of these hypotheses
- (5) Rank the hypotheses in order of likelihood and offer a final diagnosis

While much of the existing research has focused on the utility of reflective practice as a learning tool (Mamede, van Gog, et al., 2012; Mamede et al., 2008), there is also interest in how these and related metacognitive techniques (which involve awareness of and reflection on one's own cognitive processes) may be employed in

practice (Croskerry, 2003a; Croskerry et al., 2013b; Graber et al., 2012; Sinclair & Croskerry, 2010). Reflection has been framed as both an educational strategy, to improve future diagnosis, and a workplace strategy, to improve diagnosis in the moment (Croskerry et al., 2013b). Guided reflection may be regarded as an enhanced differential diagnosis, which is expected in clinical practice. Differential diagnosis has been regarded as something of a ‘universal antidote’ for diagnostic error (Maude, 2014), and evidence suggests that this step is missing in the majority of cases of diagnostic error in primary care (Singh et al., 2013).

A recent review highlights the particular value of reflection at the verification stage of diagnosis, and suggests that reflective interventions that draw the diagnostician’s attention directly to the evidence in the case have greatest positive impact on diagnostic performance (Mamede & Schmidt, 2017). These results are promising, and additional questions may now be explored to understand the potential benefits of these interventions in more detail.

4.1.1 Unexplored aspects of guided reflection

First, as discussed in the previous chapter, guided reflection interventions are based on the oversight and monitoring function of System 2 reasoning, and in particular the rational override function of Croskerry’s dual process model of diagnosis (Croskerry, 2009a), whereby deliberate analytical reasoning is used to thoroughly consider evidence and catch errors resulting from non-analytical reasoning. This override process has received little empirical attention, apart from Moulton’s work on ‘slowing down when you should’, which illuminates the process by which doctors and surgeons transition from routine practice to effortful, deliberate practice (Moulton et al., 2010b, 2007) (see Section 2.4.2 for details). While some slowing down is proactively

planned in response to procedure- or patient-specific factors, others are initiated spontaneously in response to situational factors; a study of the phenomenon in surgery demonstrated that the shift to analytical reasoning may be initiated by surprising or unexpected factors in the presentation, prompting a stopping or slowing down in order to “regroup” and “reassess” (Moulton et al., 2010, p1022). Similar insights into the initiators and influences of cognitive shifting from one mode of reasoning to another, as happens with guided reflection, would be valuable in the realm of diagnosis, given the dominance of the dual process model as the conceptual framework by which diagnostic error is currently most widely understood (Balogh et al., 2015).

Second, studies of interventions for diagnostic error generally mandate a particular style of reasoning and deliberately induce the switch from one mode of reasoning to the other (almost uniformly from non-analytical to analytical reasoning). Only one study to date has explored the benefit of reflection when the decision to engage in reflection or not was left to the diagnostician (Monteiro et al., 2015). In that study, very few participants chose to revise a diagnosis when given the opportunity to do so, and revision did not significantly impact overall diagnostic accuracy. Continued study is required to explore voluntary use in particular of these interventions, to understand the factors that drive the choice to use or not use them, and to explore the impact of this choice on outcomes.

Third, studies of interventions for diagnostic reasoning in general, and reflective interventions in particular, seldom elicit participant feedback as to their experiences and opinions of the intervention or their thoughts on the implementation of interventions in real practice. As evidence builds for the benefits of reflection for diagnostic reasoning, it will be important to incorporate this feedback; interventions that users find burdensome or onerous to use are unlikely to translate well from the lab to the ward.

4.1.2 Think-aloud methods

To answer these questions about the components and impact of guided reflection interventions, it is useful to consider the potential benefits of moving beyond strictly quantitative, experimental methods and data. In particular, qualitative and think-aloud methods may offer richer insights into moment-to-moment information processing than retrospective reflection or pen-and-paper measures. Think-aloud studies by Elstein provided important early insights into the problem-solving processes of doctors (Elstein, Schulman, & Sprafka, 1978; Elstein & Schwartz, 1995), and these techniques are worthy of continued exploration in the field of diagnostic reasoning.

In think-aloud protocols, the participant speaks their thoughts aloud as they occur in short-term memory. The method is founded on the assumptions that human cognitions consist of information processing, and that the information currently being focused on is accessible as verbal data, which can be elicited through think-aloud methods (Ericsson & Simon, 1999). In this way, verbal data can illuminate what information is being concentrated upon in real time and how it is structured in problem-solving and decision-making tasks (Fonteyn, Kuipers, & Grobe, 1993).

Two types of think-aloud data may be identified: concurrent data, which includes synchronous verbalisations occurring during the completion of a task and may consist of unmediated or mediated thoughts, and retrospective data, which requires participants to verbalise their thoughts after the task has been completed (Ericsson & Simon, 1999).

Think-aloud methods have been utilised in a number of fields, including human factors engineering, and have yielded particular insights in the field of expert performance. Selective identification of relevant information in knowledge-rich

problems is a key component of expert performance, and concurrent verbal reports are an ideal way to examine this process (Ericsson & Simon, 1999). Think-aloud studies have demonstrated that the superior performance of experts is not due to mere accumulation of knowledge, but rather due to effective organisation of knowledge in memory to facilitate efficient recognition and retrieval. Verbal reports by experts tend to be shorter than those by novices for the same task, and also more extensive given the same length of exposure to stimuli, demonstrating greater use of retrieval (Pinheiro & Simon, 1992). Such qualitative differences in the cognitive processes of experts and novices are important in understanding both expert performance in its own right (experimental tasks should be designed in such a way as to access and capitalise on the distinct strategies of experts) and how novices may be progressed to expert status as quickly as possible (not simply by piling on content knowledge, but by facilitating expert-like organisation and use of knowledge in memory).

Protocol analysis is among the most commonly used techniques to analyse think-aloud data in clinical reasoning research (Lundgrén-Laine & Salanterä, 2010). Protocol analysis refers to analysis of “a verbal process, expressed aloud, which reveals a “step-by-step” progression of a person’s problem-solving ability” (Lundgrén-Laine & Salanterä, 2010, p566). In the analysis of think-aloud data using protocol analysis, the focus is on the process of the performance of the task, rather than the outcomes of the task.

A number of recent studies have implemented think-aloud methods to examine clinical reasoning. For example, Johnsen, Slettebø, and Fossum (2016) used a think-aloud methodology with protocol analysis to describe the cognitive processes and thinking strategies used by nurses engaging in home care. Through identifying the concepts on which participants focused and the assertions made by participants, the

researchers were able to make inferences about participants' clinical reasoning and identify simple and complex cognitive processes, such as assuming, gathering information, judging, planning, verifying, and shared decision-making. Inductive and deductive reasoning strategies were present, as participants both tested hypotheses and predictions and made judgements based on information they had gathered. The authors of a 2010 study using a think-aloud protocol combined with protocol analysis concluded that the method was appropriate for examining complex decision-making processes and rapid *ad-hoc* reasoning in critical care settings (Lundgrén-Laine & Salanterä, 2010). In another study of nurses, a think-aloud protocol was combined with high-fidelity patient simulation technology; both concurrent and retrospective think-aloud data were captured and distinct patterns in verbalisations were described, with the authors concluding that the method yielded rich, high quality data by which clinical reasoning may be better understood (Burbach et al., 2015). Think-aloud methodologies have also been used to examine the use, or lack thereof, of clinical guidelines in diagnosis (Patel, Arocha, Diermeier, How, & Mottur-Pilson, 2001; Skånér, Backlund, Montgomery, Bring, & Strender, 2005), and have been proposed as tools for teaching and assessing clinical reasoning, with some evidence that the method is popular and helpful for students (Pinnock, Young, Spence, & Henning, 2015).

4.1.3 The case of psychiatric diagnosis

As mentioned in Section 3.4.2.4, a number of factors make psychiatry something of a unique specialty where diagnosis is concerned. First, it is arguably the case that presentations of psychiatric symptoms are inherently more uncertain than presentations of physical symptoms. Many mental health diagnoses are made based on patient self-reports (which may be compromised by the nature of the mental health

problem itself), family collateral reports, and observations, and very few objective biological tests exist for the diagnosis of psychiatric conditions (Pincus, 2014). Doctors are often also faced with the task of distinguishing physical diagnoses from psychiatric diagnoses, as some mental illnesses can lead to somatic complaints, and some physical conditions can produce psychiatric symptoms.

Second, it is important to consider the criteria used to diagnose psychiatric conditions. Mental health diagnoses are made based on the Diagnostic and Statistical Manual of Mental Health Disorders (DSM), which lists a set of diagnostic criteria for each psychiatric diagnosis, including the nature and duration of symptoms that must be present. The DSM is a multi-axial system, which allows diagnosticians to account for an acute primary diagnosis within the patient's more global mental health presentation and context (American Psychiatric Association, 2013). Given the lack of objective tests for mental health conditions, the use of criteria for clinical evaluation becomes central to psychiatric diagnosis. However, evidence suggests that many diagnosticians do not adhere strictly to the criteria in their diagnoses, leading to concerns about the system's reliability (Meyer & Meyer, 2009; Morey & Benson, 2016). This has particular relevance for dual process models of diagnosis; while the DSM encourages deliberate consideration of individual features, this research suggests that psychiatrists instead frequently rely on global impressions and pattern recognition to make diagnoses.

Third, patients with mental health difficulties or symptoms that appear to be related to mental health may be particularly at risk of being harmed by diagnostic error. The nature of their illness may make it difficult for them to advocate for themselves in hospital, and certain cognitive biases can be accentuated with this population. Psychiatric patients are particularly at risk of fundamental attribution error, a cognitive bias that describes a tendency to attribute illness to dispositional rather than situational

causes; that is, patients can be blamed for the onset and perpetuation of their illness. This is especially relevant in the case of personality disorders; patients with these diagnoses may experience particular stigma from their healthcare providers, as they are often regarded as being manipulative, in control of their suicidal and challenging behaviours, and less deserving of care (Lewis & Appleby, 1988; Markham & Trower, 2003). Affective bias of this sort can emerge from the complex interaction of patient and practitioner factors and have a profound influence on clinical decision-making and delivery of care, including the potential for misdiagnosis for patients with psychiatric and behavioural diagnoses (Crookery, Abbass, & Wu, 2008). Relatedly, comorbid medical diagnoses can be missed when physical symptoms are attributed to a coexisting psychiatric condition; this is a particular instance of premature closure that has been described as “diagnostic overshadowing” (Thornicroft, Rose, & Kassam, 2007, p113). For example, symptoms of delirium, head injury, or infections of the central nervous system may be attributed to or misdiagnosed as a psychiatric condition (Crookery, 2003b).

As the patient safety movement grew in the early 2000s, psychiatry was somewhat left behind. It remains the case that little evidence on the prevalence and causes of error is available, and large-scale studies of medical error have consistently excluded considerations of mental health (Nath & Marcus, 2006). In short, psychiatry has had something of a “late arrival on the medical error scene” (Shore, 2003, p1600).

Some study has been made of medication error (see Procyshyn, Barr, Brickell, & Honer, 2010, for a review) and of diagnostic error (Martin, Hynes, Hatcher, & Colman, 2016; Meyer & Meyer, 2009) and diagnostic instability (Kotov, Fochtmann, Carlson, Tanenberg-Karant, & Ruggero, 2011) in psychiatry, but the literature is generally small and narrowly focused. Cognition in critical psychiatric care has received

some scant attention (T. Cohen, Blatter, Almeida, Shortliffe, & Patel, 2006), while diagnostic reasoning *per se* has been largely neglected in the literature (Crumlish & Kelly, 2009). Therefore, diagnostic error and interventions to reduce error in psychiatry constitute important areas for expanded research.

4.1.4 Aims and objectives

Given the above, diagnostic reasoning in psychiatry may be considered a ‘special case’ of diagnostic reasoning. However, as the systematic review presented in Chapter 3 demonstrates, it has been largely neglected in the wider literature on interventions for diagnostic reasoning generally and dual process interventions in particular. Questions remain about how doctors voluntarily use and experience the guided reflection process, and qualitative methods, particularly think-aloud protocols, offer an exciting opportunity to explore these questions in a rich, naturalistic way.

Therefore, the current study implemented a think-aloud protocol with junior doctors undergoing specialist training in psychiatry and general practice, in which participants were asked to diagnose fictional psychiatric cases aloud and provided with the guided reflection process as an additional tool, which they could use or not use at their discretion. The aims of the study, which is primarily exploratory in nature, were (1) to elicit the processes being used in guided reflection, (2) to elicit contextual triggers for the use of guided reflection and the transition to reflective reasoning, (3) to elucidate the usefulness of guided reflection in psychiatric diagnosis, and (4) to evaluate the acceptability of guided reflection to diagnosticians as a process for use in real practice.

4.2 Method

4.2.1 Design

A think-aloud procedure was utilised in order to elucidate how participants chose to use a guided reflection procedure. Participants were asked to diagnose three fictional clinical cases aloud and were provided with the guided reflection process as an additional tool, which they could use or not use at their discretion. A semi-structured interview was utilised to supplement the think-aloud procedure, in order to elicit participants' views of the guided reflection method and how it might be implemented in real practice.

4.2.2 Participants

A convenience sample of 14 medical doctors (12 female, 2 male) was sought from two large teaching hospitals in Dublin, Ireland. All participants were undertaking Basic Specialist Training ($n=10$) or Higher Specialist Training ($n=1$) in psychiatry or Specialist Training in general practice ($n=3$). The median number of years since graduation with a degree in medicine was 3. This sample size is in line with samples from other think-aloud studies in clinical reasoning, which tend to vary between approximately 8 and 30 participants (Burbach et al., 2015; Johnsen et al., 2016; Kim, 2015; Lundgrén-Laine & Salanterä, 2010).

A registrar in psychiatry at each of the two hospitals from which participants were drawn joined the research team to assist with recruitment. Following receipt of ethical approval, the registrars made initial contact with potential participants, briefly describing the study and inviting them to take part (see Appendix D for the recruitment message). The contact details of interested candidates were passed on to the researcher, who followed up by telephone or email to answer any questions about the study and to

arrange a think-aloud session. Twenty-eight doctors were approached by the registrars; of these, twenty-six agreed to be contacted by the researcher. Of these, seven doctors did not respond to further contacts from the researcher, five declined to participate, and fourteen gave consent and took part in the study.

4.2.3 Materials

4.2.3.1 Vignettes

A registrar and a consultant in psychiatry (a doctoral supervisor) developed vignettes in consultation with the researcher. The vignettes were modelled after those provided by Prof Charles Friedman and colleagues for use in the study described in Chapter 5 (Friedman et al., 1999), which have been used in similar research on clinical reasoning (Friedman et al., 1999; Payne, 2011) but do not contain psychiatric cases. The model cases average 450 words in length and contain information on the chief complaint and history of presenting illness, relevant past medical and social history, the physical examination, and laboratory data. The registrar in psychiatry reviewed the model cases and constructed vignettes in a similar style, loosely based upon real cases the registrar had encountered in practice.

The vignettes were designed to be challenging but not excessively difficult, with features and conditions that participants were likely to have encountered in practice. They were also designed to be diagnostically complex, with both psychiatric and medical features. A list of acceptable ‘correct’ diagnoses for each case was compiled by the research team.

Five vignettes were developed and reviewed by the research team and a senior registrar in psychiatry and were included in the pilot study. Following the pilot study,

two vignettes were excluded to allow sufficient time for participants to diagnose the remaining three vignettes in full. The selected vignettes offered an appropriate level of difficulty with minimal redundancy (see Appendix H). They formed a range of complexity; the first was psychiatrically and medically straightforward, the second was psychiatrically straightforward and medically ambiguous, and the third presented a complex and ambiguous overall picture with a somewhat less common correct diagnosis.

4.2.3.2 Interview schedule

A semi-structured interview schedule (see Appendix J) was constructed to supplement the think-aloud protocol. The schedule had two sections:

(1) Case reflection: After each case, participants were asked to reflect on how they reached their diagnosis, describe their degree of confidence in their chosen diagnosis, reflect on their decision to use or not use the guided reflection method for that case, and to report on their experience of the method if they had chosen to use it.

(2) General reflection: At the end of the interview, participants were asked to reflect generally on their use of the guided reflection method if they had chosen to use it, identify components they found helpful, comment on the potential for the method to be utilised in real-world practice, compare the process with their own natural diagnostic process, and comment on the issue of diagnostic error generally.

4.2.3.3 Guided reflection instructions

A written version of instructions for the guided reflection process was provided to participants as a visual aid (see Appendix L). The steps were outlined under three headings: Initial hypothesis, Alternatives, and Diagnose.

4.2.4 Data collection

The think-aloud sessions were conducted between July and October 2017 in testing rooms at the School of Psychology, Trinity College Dublin, or in offices or meeting rooms at the participants' places of work. The researcher conducted all sessions. Testing time varied in length between 40 and 85 minutes. Audio recordings were taken using an iPhone SE; files were moved from the device to a secure drive at the earliest opportunity. One participant did not consent to the audio recording; their responses were instead noted by hand by the researcher and were incorporated into the thematic analysis only.

4.2.5 Procedure

Participants were informed verbally of the nature of the study; they were told that they would be asked to diagnose a number of fictional cases aloud and that they would have the opportunity to make use of a reflective method for diagnosis, which would be explained to them in advance. They were told that they would then be asked some general questions about their experiences and reminded that the session would be audio recorded. Participants were also advised that they could withdraw from the study at any time without penalty. After reading the information sheet (see Appendix E), participants were asked to sign the consent form (see Appendix F) and given a copy of same for their own records. Once consent was given, participants were asked to indicate the number of years since graduation and their current standing in their training course. The researcher then began the recording.

Using a script (see Appendix K), the researcher introduced the guided reflection process as a method for diagnosis that is based on reflective practice and that may be particularly helpful when diagnosing complex cases. The researcher provided a written

version of the instructions (see Appendix L) and briefly described each step. Participants were told that the process was available to them as an extra tool, and that they were free to make use of it for some or all of the cases if they chose to do so. Participants were also provided with pen and paper to take notes if they chose to do so.

Participants were then presented with the three vignettes one by one. After each vignette, the researcher asked the case reflection questions from the semi-structured interview schedule. Participants were offered a break between the second and third vignettes. Vignettes were administered in the same order (A, B, and C) for each participant in order of increasing complexity, to allow the participant to get used to the protocol with more straightforward cases first.

Before reading the first vignette, participants were given the following instruction: “As you’re going through the case, remember to speak out loud, or you can read out loud if you like, and let me know what you’re thinking about, what you’re focusing on and what’s coming to mind for you.” Participants were given the following instruction before reading the second and third vignettes: “Remember to think out loud as you’re going through the case, and remember that the guided reflection process is there for you to use if you choose to do so.” If participants became quiet for more than one minute, they were asked, “What are you thinking about at the moment?”

After the vignettes, the researcher asked the general reflection questions from the semi-structured interview schedule. Once this was completed, the recording was stopped, the participants were thanked for their participation and debriefed (see Appendix G). They were informed that they would have the opportunity to review the transcript of their session once it was completed (one participant requested to view their transcript).

4.2.6 Analysis

Recordings were transcribed verbatim and segmented into seven sections: think-aloud and case reflection for each of three cases, and general reflection. A thematic analysis was conducted in accordance with the procedures outlined by Braun and Clarke (2013). Based on an in-depth reading of two transcripts and elements of the guided reflection process, a coding framework was developed based on guidance by Yoder and Simmons (2010).

The researcher and a research assistant independently coded three transcripts using this coding framework. Each transcript was read closely a number of times before beginning coding, and codes were assigned to all content that appeared significant. The coders collaborated to develop the coding framework in a flexible manner, remaining open to new and emerging themes. Codes and definitions were reviewed and revised collaboratively by the coders to finalise the coding framework (see Appendix M), which was then applied to all thirteen transcripts. Three transcripts (over 20% of the data, in accordance with previous guidelines (Bakeman & Quera, 2011)) were coded independently by both coders using the finalised framework. Cohen's κ was calculated to determine whether acceptable agreement had been reached between the coders, $\kappa = .69$. According to guidance in the statistical literature, this may be described as indicating moderate (McHugh, 2012), good (Altman, 1991; Fleiss, Cohen, & Everitt, 1969), or substantial (Landis & Koch, 1977) agreement. Where more than ten codes are used, the expected value for Cohen's κ is approximately .65 where observers are 85 per cent accurate (Bakeman & Quera, 2011).

Diagnostic accuracy was scored for each case as follows: 1 where the participant gave the correct psychiatric and medical diagnosis, determined by consensus among the research team for each case; 0.5 where the participant gave a correct psychiatric

diagnosis but not a correct medical diagnosis, or vice versa; and 0 where the participant did not give an acceptable or a correct diagnosis. (See Appendix I for the scoring key.)

4.3 Pilot study

4.3.1 Aims and objectives

As the materials for this study were novel and untested, a pilot study was conducted to explore four aspects of the protocol:

- (1) Feasibility of procedure: to determine how many vignettes could reasonably be diagnosed and how many general questions could reasonably be asked within one hour.
- (2) Appropriateness of vignettes: to explore participant responses to vignettes, to determine the difficulty of the vignettes, and to highlight specific areas of confusion or difficulty.
- (3) Acceptability of materials: to explore participant responses to the visual layout and comprehensibility of the vignettes and written guided reflection instructions.
- (4) Freedom of choice: to determine whether participants experienced a free choice of whether to use the guided reflection procedure or whether they felt undue pressure from the researcher due to her demeanour, her choice of words, or any other factor.

4.3.2 Method

4.3.2.1 Design

Similar to the main study, a think-aloud procedure was utilised, wherein participants were asked to diagnose up to five fictional clinical cases aloud. A semi-structured interview was utilised to supplement the think-aloud procedure, in order to elucidate participants' views of the guided reflection method and the think-aloud procedure.

4.3.2.2 Participants

Three doctors undertaking Basic Specialist Training in psychiatry were approached and agreed to take part in the pilot study. All were female.

4.3.2.3 Materials

4.3.2.3.1 Vignettes

The development of five fictional vignettes is outlined above in Section 4.2.3.1.

4.3.2.3.2 Interview schedule

A short semi-structured interview schedule was constructed to supplement the think-aloud procedure. The schedule had two sections:

- (a) Reflection on guided reflection: After all of the cases had been diagnosed, participants were asked to reflect generally on their use of the guided reflection method if they had chosen to use it and to comment on the potential for the method to be utilised in real-world practice.

(b) Reflection on procedure: Participants were asked to comment on the suitability and ease of use of each of the materials and on the clarity of their choice to use or not to use guided reflection, and to offer any advice for the research team on how the materials may be improved.

4.3.2.3.3 Guided reflection instructions

A written version of instructions for the guided reflection process was provided to participants as a visual aid (see Appendix L). The steps were outlined under three headings: Initial hypothesis, Alternatives, and Diagnose.

4.3.2.4 Procedure

Participants were informed verbally of the nature of the study; they were told that they would be asked to diagnose a number of fictional cases aloud and that they would have the opportunity to make use of a reflective method for diagnosis, which would be explained to them in advance. They were told that they would then be asked some general questions about their experiences and reminded that the session would be audio recorded. Participants were also advised that they could withdraw from the study at any time without penalty. After reading the information sheet, participants were asked to sign the consent form and given a copy of same for their own records. Once consent was given, the researcher then began the recording.

The researcher introduced the guided reflection process as a method for diagnosis that is based on reflective practice and that may be particularly helpful when diagnosing complex cases. The researcher provided a written version of the instructions (see Appendix L) and briefly described each step. Participants were told that the process was available to them as an extra tool, and that they were free to make use of it for some

or all of the cases if they chose to do so. Participants were also provided with pen and paper to take notes if they chose to do so.

Participants were then presented with the five vignettes one by one and asked to diagnose them aloud. Vignettes were administered in the same random order for each participant. After the vignettes, the researcher asked the general reflection questions from the semi-structured interview schedule. Once this was completed, the recording was stopped, the participants were thanked for their participation and debriefed. They were informed that they would have the opportunity to review the transcript of their session once it was completed (two participants requested to view their transcripts).

4.3.2.5 Analysis

Recordings were transcribed verbatim. The transcripts were reviewed by the researcher and comments relating to the study objectives were compiled and discussed with the research team to make adjustments to the materials and procedure. The number of vignettes completed by each participant was also noted.

4.3.3 Findings and actions taken

(1) Feasibility of procedure:

Findings: Participants completed a mean of three cases in the time available, and it was observed that they naturally reflected on each case after it was completed.

Actions taken: It was decided to allow 45 minutes for the completion of three cases and case-specific reflection, and 15 minutes for general reflection.

(2) Appropriateness of vignettes:

Findings: Participants generally responded positively to the vignettes. It was noted that some cases had similar content; two cases featuring patients with psychotic feature, and two cases featuring elderly patients with dementia-like symptoms.

Actions taken: In each pair of similar cases, the case that was less difficult, due to greater diagnostic clarity, was chosen for inclusion. The more difficult case in each pair was excluded.

(3) Acceptability of materials:

Findings: Participants responded positively to the visual layout and content of the vignettes and written guided reflection instructions, regarding them as clear and easy to follow. Participants indicated a lack of clarity around the number of alternatives that were called for by the instructions.

Actions taken: One phrase in the guided reflection instruction was changed to clarify that multiple alternatives could be pursued. The procedure also highlighted the need for the researcher to use a script to describe the study and the guided reflection instructions; this was developed for the main study (see Appendix K).

(4) Freedom of choice:

Findings: All participants stated that they experienced a free choice of whether to use the guided reflection procedure, and that they felt no undue pressure from the researcher to use it.

Actions taken: No adjustment to the procedure was required, and the phrasing the researcher used to communicate this choice to participants in the pilot study was incorporated into the script for the main study.

4.4 Results

Based on an in-depth reading of two transcripts and elements of the guided reflection process, a coding framework was developed based on guidance by Yoder & Simmons (2010) (see Appendix M). The framework comprised two broad categories of codes. First, diagnostic process codes were based on elements of the guided reflection process and other observed diagnostic reasoning behaviours, e.g., naming hypotheses, identifying and gathering different types of evidence, hypothesis selection, and other identifiable strategies, such as statistical reasoning, seeking further information, and seeking support from other professionals. These codes were developed based on the guided reflection instructions and in-depth reading of transcripts. Second, reflection codes were topics that arose during reflection on the diagnostic process and interview, e.g., reflection on the implementation, impact and experience of guided reflection, reflection on diagnostic error and diagnosis in psychiatry, and comments on the materials or experience of participating in the study. These codes were developed based on questions asked in the semi-structured interview and in-depth reading of transcripts.

4.4.1 Frequency analysis

Of the 13 participants for whom think-aloud data were available, 11 made use of the guided reflection for at least one case. Across 13 participants, 37 cases were diagnosed, and guided reflection was used in 25 (68%) of these.

Diagnostic accuracy was generally high, with a mean score of 0.70 across 35 cases for final diagnoses (no final diagnosis was provided for two cases).

Frequency statistics for the use of guided reflection and the components are displayed in Table 4.1.

Table 4.1

Frequency analysis of participants and cases using guided reflection

Component	Number of participants using guided reflection (n=11)		Number of cases using guided reflection (n=25)	
	<i>n</i>	%	<i>n</i>	%
Alternative diagnoses	11	100.0	25	100.0
Gathered supporting evidence	11	100.0	25	100.0
Gathered refuting evidence	11	100.0	25	100.0
Gathered missing evidence	9	81.8	14	56.0
Ranking	9	81.8	19	76.0

When using guided reflection, participants always gathered supporting and refuting evidence for their alternative hypotheses, and ranked their alternatives in a majority of cases.

The final diagnosis matched the initial hypothesis in 27 cases out of 35 (77%). The outcomes in cases where the final diagnosis differed from the initial hypothesis are summarised in Table 4.2. In six cases out of eight, the final diagnosis was superior to the initial hypothesis.

Table 4.2

Contingency table for changed diagnoses with and without use of guided reflection

Direction of change		Number of cases	
		With guided reflection (<i>n</i> =6)	Without guided reflection (<i>n</i> =2)
Positive	Incorrect to correct	1	1
	Incorrect to acceptable	3	-
	Acceptable to correct diagnosis	-	1
Negative	Correct to incorrect	2	-

4.4.2 Thematic analysis

The thematic analysis is presented diagrammatically in Figure 4.1. Five major themes were identified; three of these concerned the guided reflection process, and two concerned contextual factors in diagnosis. The themes and their sub-themes are described in detail below, followed by a general discussion of the analysis.

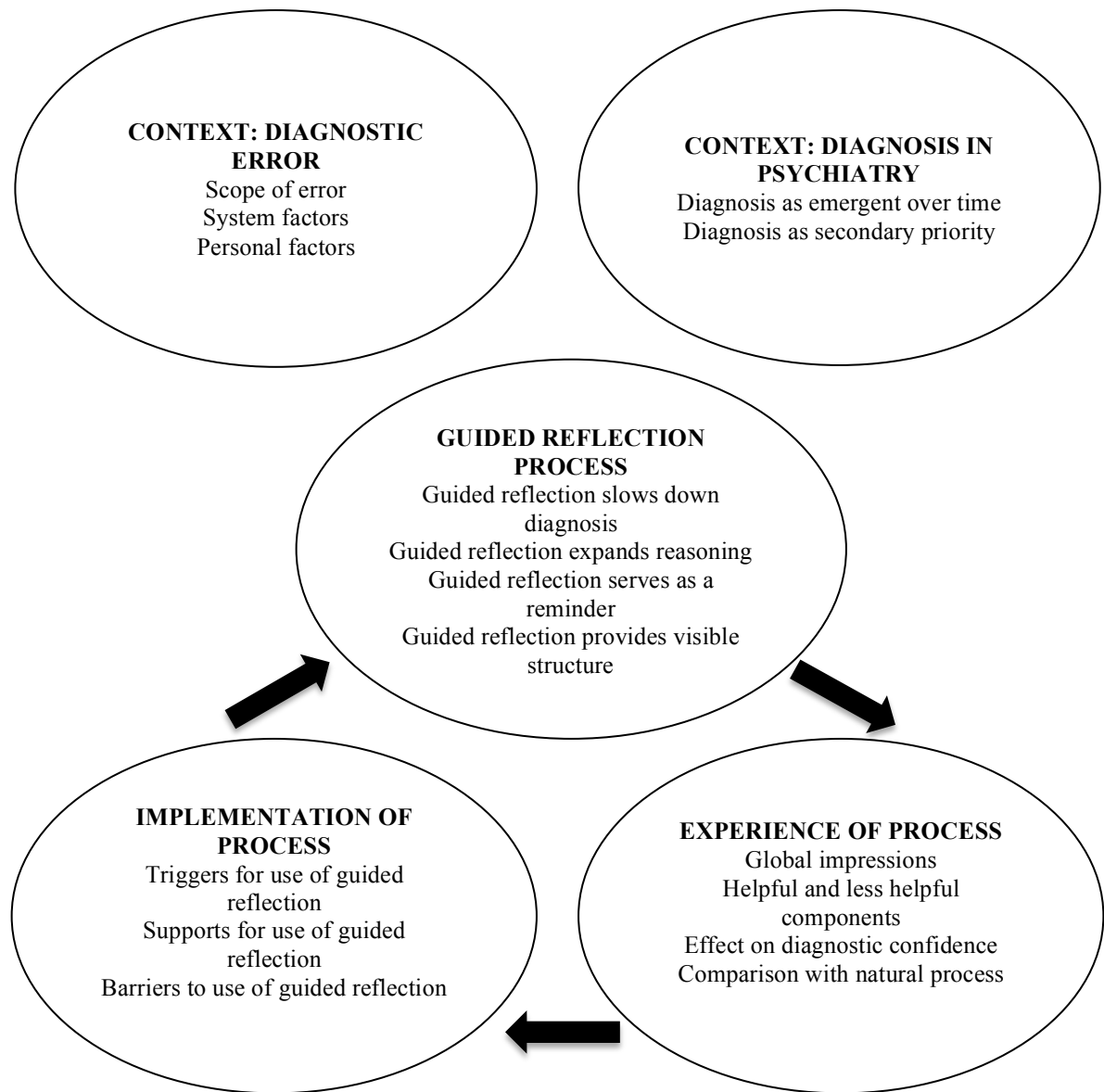


Figure 4.1. Diagrammatic illustration of the thematic analysis, showing major themes and sub-themes

4.4.2.1 Implementation of the process

4.4.2.1.1 Triggers for use of guided reflection

Participants spoke of triggers or reasons for using guided reflection in two contexts. First, regarding their decision to use reflection in diagnosing the cases during the think-aloud protocol, participants stated that they used guided reflection to deal with cases that they perceived to be complex or ambiguous, or that had multiple competing potential diagnoses.

“Yeah, definitely [the complexity influenced me to use the reflection process]. Em, can definitely see how it’s more useful in this case than for the other two, em, just because the other ones were just kind of more clear cut and lined out nicely, but this one was just a bit more all over the place.” (Participant 2)

This question of case complexity also accounted for most decisions *not* to use the process to aid diagnosis; given the description of the process at the outset as being potentially useful for complex cases, participants who chose not to use it generally made that choice because they felt that the cases were not of sufficient complexity to warrant its use.

“I definitely would do a lot more guided reflection em about cases which were less straightforward than these ... I guess you had mentioned that it was something that might be useful in difficult cases, and this doesn’t appear to be one” (Participant 5).

“I felt like I, it was pretty clear. I didn’t really need to reflect on it much.”
(Participant 13)

“It’s quite barn door actually.” (Participant 14)

Second, regarding situations in which they may use guided reflection in real practice, participants overwhelmingly mentioned ambiguous or complex presentations as candidates for use of guided reflection.

“If the diagnosis was unclear, that would be the biggest way for me.”
(Participant 7)

“In some cases ... something more complicated where you’re really like, oh god ... I don’t know, I need to sit down and actually do this consciously, em, then it might actually be time-saving ... it might actually be more efficient in a complex patient, to take a few minutes out and go through it.” (Participant 2)

Participants also described self-protective motivations to make use of the process to document and communicate their diagnostic reasoning.

“I suppose you are safeguarding yourself a bit by exploring all of the possible diagnoses and ruling them in or out as appropriate.” (Participant 4)

“I know that if someone’s looking at mine and the presentation’s completely different, as is constantly in psychiatry, they’re like, oh that makes sense, she saw this, this and this ... wanting your consultants and your peers to know that ... you are taking it seriously and you’re working hard.” (Participant 7)

4.4.2.1.2 Supports for use of guided reflection

A number of participants proposed environmental supports to enable doctors to put the process into practice; primary among these were *“a prompt or reminder”* (Participant 1). In this regard, participants most frequently mentioned existing pro-forma assessment and intake documents, which already contain sections for listing differential diagnoses; participants proposed expanding this section to allow room to list

supporting and refuting evidence for each differential. Opinions differed over whether this should be regarded as mandatory or optional.

“I suppose the difficulty if it’s added as an adjunct is it might get missed, whereas if it’s a part of the structured assessment, then people are kind of going through it that way.” (Participant 7)

“I mean if it was part of the template that we use you’d probably just do it ... If it was optional you might still do it ... if people want to do it they can.” (Participant 8)

“I think you, you need to force things upon people, especially doctors ... unless it’s forced in front of us, we won’t actually like take it on, and be like oh, it’s just an extra sheet of paper. It’ll just sit there forever and ever.” (Participant 9)

Participants also mentioned that the inclusion of the process in a pro-forma assessment would be an aid in communication with other team members, and perhaps serve to protect themselves in situations where their diagnosis is called into question.

“I think it’s probably useful, it’s always useful to show your thought process because maybe you mightn’t document something and someone would say how did you miss this? But maybe you’ve actually thought through, you know.” (Participant 8)

4.4.2.1.3 Barriers to use of guided reflection

Regarding factors that may discourage or prevent the use of guided reflection in clinical practice, system factors were primarily named as barriers to its use. Time poverty, which was also named as a potential cause of diagnostic error, was most frequently mentioned.

“In the simple cases, on a day-to-day basis, like in a practice or in ED or even on the wards, I think time doesn’t allow for it, most of the time, unless you’re really stuck and it’s really tricky.” (Participant 2)

Participants believed that guided reflection would generally be a useful tool, but impractical to use for every case.

“I would find it quite demanding if I had to do it every time and every patient.”
(Participant 10)

“Staggered access to information” (Participant 4) was also mentioned as a barrier, as the guided reflection process requires a certain amount of information to be available for processing at the time of diagnosis to fully consider alternatives and evidence. Lack of prompting (*“You just mightn't remember to”* (Participant 1)) and personal disinclination (*“probably just self-laziness, do you know, [laughing] like where you, you kind of go that's what it is, sure I won't think of anything else”* (Participant 12)) were also mentioned by a small number of participants.

4.4.2.2. Guided reflection process

4.4.2.2.1 Guided reflection slows down the diagnostic process

Participants described how using the guided reflection process encouraged them to slow down and fully consider symptoms and alternative hypotheses, and helped to prevent them from jumping to conclusions.

“It would encourage further eh reading and yeah, exploring the alternatives more so than just picking one and going with that.” (Participant 1)

“It's just another chance to process it and another chance to review what you have.” (Participant 7)

Participants also commented that guided reflection helped to focus their attention; this appeared to be particularly relevant for complex cases.

“It brings you back and you’re revisiting maybe parts of the history that didn’t seem important the first time but actually are ... I suppose you’re more tuned in to what you’re looking out for.” (Participant 7)

“Everyone kind of goes for a gut instinct about a diagnosis, which I suppose can often be right in a very obvious case, but say in the more complex ones, you know, it could be easy to, to brush over details that just aren’t slotting with what you think it is ... I suppose it focuses you to kind of actually acknowledge that.” (Participant 4)

4.4.2.2.2 Guided reflection expands reasoning

Participants said that alternative diagnoses came spontaneously and naturally to mind, but that guided reflection allowed them to consider a broader range of differentials before reaching a final decision than they otherwise would have; guided reflection encouraged *“more lateral thinking”* (Participant 1). A number of participants stated that particular diagnostic hypotheses came about only through guided reflection, including life-threatening conditions that, while unlikely, were important to rule out.

“I hadn’t thought about meningitis until I went back and thought about it, em, going through kind of the alternative em, which is obviously an important thing to be able to rule out, you know, in a, in a young person.” (Participant 4)

“I think it’s good for like keeping your mind open.” (Participant 9)

“Say, I don’t know, someone has chest pain, you’re like well, I don’t want to miss the heart attack so what, you know, what’s actually fitting with that, what doesn’t?” (Participant 12)

4.4.2.2.3 Guided reflection serves as a reminder

A number of participants described the role of the guided reflection process as a reminder or prompt to consider alternatives and broaden their diagnostic thinking.

“I do think it’s a good idea because it trains you to keep differentials in your mind as opposed to going that’s depression, grand, we’ll treat that.” (Participant 7)

“You can, you know, lead yourself towards believing that that’s what it is without fully taking in all the information that’s there in front of you, so I suppose you just ... need to remember to question yourself.” (Participant 13)

4.4.2.2.4 Guided reflection provides visible structure

Some of the most consistent comments related to the structure provided by the process. While many participants reported that they utilised similar processes in regular practice, guided reflection allowed them to go about their diagnosis in a more structured and explicit way: *“It’s just right there in front of our eyes”* (Participant 3). Participants described how the structure was *“almost like a checklist”* (Participant 12) that allowed them to *“clear your thoughts or organise your thoughts”* (Participant 7). The structure offered by the guided reflection process appeared to be a significant driver of its impact on diagnostic confidence, in particular the gathering of supporting evidence.

“It was a good way to think through the case and to logically I guess weigh up the pros and cons that you possibly do but maybe you don’t em, maybe when you actually articulate it and go through each of the diagnoses and what’s for it and against it, maybe it structures your final, you coming to a final diagnosis in a more confident way.” (Participant 8)

“It kind of makes me see almost in, in more structured form that [that diagnosis is] not strong enough to go with, which I feel like I knew but this gives me more confidence.” (Participant 7)

4.4.2.3 Experience of process

4.4.2.3.1 Global impressions

Participants generally regarded the guided reflection as useful and easy to use. They regarded the process as simple with clear, straightforward instructions.

“It’s a very, you know, simple thing to think through in your own head, you know, if you, if you have the information available to you.” (Participant 4)

“If you had it in your head, I think once you’ve used it a few times it’d be very easy to reflect back on it.” (Participant 12)

Concerns that the process may be time-consuming were contrasted with the potential benefits.

“It might actually be time-saving, so you’re less kind of chasing your tail, going round in circles and having all these different things going round in your head and it might actually be more efficient.” (Participant 2)

4.4.2.3.2 Helpful and less helpful components of the process

Participants generally described all of the steps in the process as helpful, and commented that gathering of supporting and refuting evidence came naturally and was a useful step. Participants especially praised the generation of alternatives and the gathering of refuting evidence, steps that they said were likely to be overlooked.

“Maybe don’t think enough about the things that refute it once you have a diagnosis in your mind.” (Participant 1)

They regarded the gathering of expected-but-missing evidence as a useful step, but acknowledged that it takes additional mental effort, depending on recall of diagnostic information and criteria from memory, rather than mere recognition of case features as they appear on the page in front of them.

“What’s there is there and you can use that and categorise that, but em, I suppose unless you have lots of experience or, you know, you have stuff in your hand or in front of you, sometimes you can actually just forget what’s missing. So that part is probably less useful kind of on a day-to-day basis, em, but you can always see what’s there in front of you and what, what supports this and that.” (Participant 2)

4.4.2.3.3 Effect on diagnostic confidence

Participants reported that using the guided reflection process made them more confident that they had selected the correct diagnosis, allowing them to demonstrate a clear pathway to the diagnosis and be sure that they had conducted a thorough search for alternatives. In particular, the gathering of supporting evidence for the chosen hypothesis appeared to increase confidence among participants.

“It probably made me a bit more stronger in my conviction of what was going on, em, being able to back up that oh, look, there’s ten reasons for that, there’s, you know, lots of other reasons missing for, or things that I would expect to see that are missing for other, other potential diagnoses.” (Participant 4)

The visible structure provided by the guided reflection process was often mentioned as a factor in promoting confidence in a diagnosis.

“When you actually articulate it and go through each of the diagnoses and what’s for it and against it, maybe it structures your final, you coming to a final diagnosis in a more confident way.” (Participant 8)

“I’m very visual so I prefer things to be written down, and the more I write down and see then the more I kind of feel confident that I’m not leaving anything out.” (Participant 9)

4.4.2.3.4 Comparison with natural process

When asked how the guided reflection process compared with their own natural diagnostic process, participants generally commented that guided reflection aligned closely with their natural tendencies. However, many commented that they carry out the process at an unconscious level.

“It’s good, I suppose, do we do it in practice? Eh, we probably do it but not in a c--, overtly conscious way.” (Participant 3)

Furthermore, they noted that the guided reflection process added clarity (*“It’s just right there in front of our eyes”* (Participant 3)) and structure to what they were already naturally doing (*“I think this isn’t unfamiliar, for sure, eh, but do we do it as structured as it’s written here?”* (Participant 7)).

In addition, they described that their natural diagnostic process regularly includes some features of guided reflection, such as initial and alternative hypotheses and gathering supporting evidence (*“probably trying to find symptoms that tick the most boxes for each diagnosis, is probably what I would do”* (Participant 12)), but that they frequently don't consider other aspects in great detail, such as refuting evidence and in some cases alternative hypotheses. This may be interpreted as confirmation bias (See Section 2.5.3).

“Maybe don’t think enough about the things that refute it once you have a diagnosis in your mind.” (Participant 1)

“I don’t know that I would necessarily go into the same level of details for every single one.” (Participant 7)

It is worth noting that many participants expressed a hope that their natural process was similar to guided reflection, or a desire to naturally diagnose in a more thorough way.

“I’d like to think so? Yeah, I’d like to think so.” (Participant 1)

“I probably need to get more in depth of how do I do it, eh, be more conscious about it.” (Participant 3)

4.4.2.4 Diagnostic error

4.4.2.4.1 Scope of error

A small number of participants indicated that they believed diagnostic error to be common, and that the scale of the problem may be more significant than it appears.

“God, well I think it’s really common for starters, anyway, much more common than we admit or than is publicised ... I think there’s a lot of errors happening all the time that aren’t kind of picked up on or reported on.” (Participant 2)

“I think it’s something that happens all the time.” (Participant 4)

4.4.2.4.2 System factors

System factors were frequently described as causes of diagnostic error. Time pressure was the most commonly cited issue in this regard (*“I think the time constraints and the, the pressures kind of day to day are a massive factor in it”* (Participant 2)). The

busyness and distractions of the clinical environment are particularly important in psychiatry, as assessments tend to be long and rapport is of critical importance.

“If I can get an hour and a half to two hours int--, uninterrupted with someone, I feel like I can get a really good sense of what’s going on, but again, in any sort of on-call or acute setting, you know, your bleep’s going off and you’re called away and something more urgent comes up, so your time with someone is really fractured, and each time you’re trying to apologise, re-develop the rapport, so for me uninterrupted time is helpful.” (Participant 7)

4.4.2.4.3 Personal factors

On the level of the individual doctor, lack of confidence emerged as an important factor in error. Indecision was described as a potential cause of errors; importantly, this was commonly linked to system factors, where junior doctors are left unsupported.

“That’s sometimes true in a hospital environment that the support isn’t there and the lack of confidence can kind of lead you to make mistakes as well, maybe being too hesitant or, or that kind of thing. Spending too long mulling it over as opposed to just getting on with it, you know.” (Participant 2)

“A feeling of ... incompetence or shame, when you can’t accurately make the diagnosis” (Participant 1) was also commonly mentioned as a factor in diagnostic error, and participants described reluctance to admit difficulties or seek assistance from colleagues or superiors.

“Not asking because you’re too embarrassed to ask, you’d be like where is that piece of paper or am I’m supposed to ask someone something, but I, you know, and like you’re too embarrassed to be like where is that?” (Participant 9)

Participants also described types of errors that may be described as cognitive biases. For example, the potential for emotional reactions to individual patients to bias an assessment was also acknowledged by a number of participants (this has been called “visceral bias” in the literature (Croskerry, 2003b, p778)).

“That comes with some judgemental view of the case, oh my god, this is an emotionally unstable personality, you know, you know, and you prepare yourself in a way, but obviously you don’t let that to overcome, to, to interact too much in, you know, in your assessment.” (Participant 3)

“If someone comes and talks to me and they’re talking about their mood being low and I’m very aware of how strongly I dislike them ... I’d wonder if there’s something else going on there.” (Participant 5)

Most commonly described was the phenomenon of diagnostic momentum or anchoring; these specific terms were seldom used, but many participants described this concern.

“Knowing the presenting complaint would push me towards a certain line of questioning.” (Participant 5)

“I think I was kind of hell-bent on my first one being the right one, so I kind of lost interest in my other differentials.” (Participant 9)

“If someone comes in with a GP letter saying that they have low mood, I’m going in there thinking about depression and trying to tick off the depression boxes and not necessarily asking all the questions about, you know, anxiety or something like that.” (Participant 4)

4.4.2.5 Diagnosis in psychiatry

A number of participants commented on the specific nature of psychiatric diagnosis in contrast with diagnosis in other disciplines; participants were also briefly asked to comment on this specifically at the end of the interview.

4.4.2.5.1 Diagnosis as emergent over time

Participants noted that in psychiatry, diagnoses tend to be made over a longer period of time than in general medicine. For instance, most participants commented that in the case of Patient A, a young man presenting with manic and delusional symptoms, it was difficult to diagnose either bipolar affective disorder or schizophrenia on the basis of a single presentation; there were features in the case to support either diagnosis, and observations over a period of weeks or months would be required to determine which was most appropriate.

“Uncertainty in the definite diagnosis at this point, seeing as it’s his first presentation.” (Participant 1)

“We may not put a label on it as depression until we have kind of a longitudinal assessment.” (Participant 14)

“In psychiatry, we need to write em diagnoses with eh a pencil more than a pen.” (Participant 3)

4.4.2.5.2 Diagnosis as secondary priority

The value of a single diagnosis, at least early in the process, appears to be less significant in psychiatry than in general medicine, taking secondary importance to risk management and treatment. Diagnosis was described as *“a bit more of a moving*

picture” (Participant 4). Particularly in emergency settings, risk management is prioritised over reaching a definitive diagnosis.

“Obviously at three o’clock in the morning you just, you’re not, you’re not very concerned about em diagnosis, your, your main concern is about em risk and, you know, just gathering enough information to make a decision, a safe decision. ... If the psychotic symptoms are within the picture of schizophrenia or schizoaffective, it wouldn’t be such, to me, it wouldn’t be obviously the main thing to work on an emergency basis.” (Participant 3)

Treatment is often begun in the absence of a formal diagnosis; the treatment process may in itself inform the final diagnosis, if symptoms respond or fail to respond to a chosen treatment in the expected way. Additionally, *“there is a lot of crossover even in the treatment regimes with, you know, if someone comes in with a psychosis, they’re going to treat that the same no matter what the cause initially”* (Participant 4).

4.5 Discussion

4.5.1 Guided reflection enhances natural processes

The findings presented here indicate that the guided reflection process has applicability to psychiatric diagnosis; furthermore, it was positively regarded on the whole by the junior doctors in this sample, who made extensive use of the process (on 25 out of 37 cases, with all but two participants using it for at least one case) and found it straightforward and helpful. In considering how use of the guided reflection process modifies a diagnostician’s natural thinking process, it is significant that participants generally described the steps of guided reflection as being closely aligned with their own natural tendencies. In slowing down, expanding and serving as a reminder for

diagnostic reasoning, the effect of guided reflection is essentially to enhance and augment rather than alter doctors' natural process *per se*. Of crucial importance is the fourth effect of guided reflection explored here, which is to offer *visible* structure to the diagnostic process. The process was not described as burdensome and, with the exception of being prompted to consider some additional evidence that they might otherwise have neglected, such as refuting evidence, participants did not describe their process as being extensively changed by the use of guided reflection. Rather, the diagnostic steps they naturally take anyway were illuminated and clarified by the process, an effect they regarded as beneficial.

This is in line with some of the interventions discussed in Chapter 3 that also affect reasoning by enhancing processes that participants naturally perform, such as thorough consideration of alternatives that naturally come to mind (Feyzi-Behnagh et al., 2014; Regehr et al., 1994), and by providing visibility for reasoning processes through metacognitive feedback in the same way as guided reflection (Feyzi-Behnagh et al., 2014; Payne, 2011). Others, meanwhile, centre on adding extra steps to the diagnostic process (e.g., use of checklists (Graber et al., 2014; Shimizu et al., 2013), listing of clinical features (Kulatunga-Moruzi et al., 2001, 2011), mental removal of salient features (Arzy et al., 2009)). The ease with which guided reflection seems to fit with doctor's natural diagnostic processes, demonstrated by this study, may go some way to explaining its success relative to other interventions.

4.5.2 The case of psychiatry

The findings also lend support for regarding diagnosis in psychiatry as different from diagnosis in general medicine in some important ways; naming a diagnosis early on in the presentation is not highly prioritised in psychiatry, and much depends on the

clinical judgement of the diagnostician in the room with the presenting patient. The clarification of the diagnostic picture over a longer period of time was also highlighted in these findings, reflecting the uncertain nature of psychiatric diagnosis that has been explored in some previous literature (Kotov et al., 2011). This has important implications for the study of diagnostic error in mental health. In the absence of a gold standard test for a given diagnosis, such as the biomarker tests that exist in other fields of medicine, and with the understanding that diagnostic labels are subject to change for an individual patient over time (and subject to redefinition for the field as a whole), clarity around how diagnostic error in psychiatry may be detected is needed going forward (Phillips, 2014). In a similar way, the assumptions of designers of interventions to reduce diagnostic error must align with the assumptions of clinicians; interventions must be designed with awareness of the fact that a psychiatric diagnosis may reasonably be expected to shift over time and that a delay in diagnosis is not necessarily an indicator of poor care.

4.5.3 Correction of diagnoses and diagnostic confidence

Diagnostic accuracy across the sample was very high, and doctors largely stayed with their initial hypothesis, which was usually stated within the first few sentences of the vignette and generally correct. This mirrors Monteiro and colleagues' findings, outlined earlier in this chapter, that when given the opportunity to revise a diagnosis, physicians rarely choose to do so (Monteiro et al., 2015). It also reflects previous literature demonstrating the importance of the initial diagnostic hypothesis (Kostopoulou, Sirota, Round, Samaranayaka, & Delaney, 2017).

When the final diagnosis did differ from the initial hypothesis, the accuracy of the diagnosis improved in 6 out of 8 cases. This was true whether the participant made

explicit use of guided reflection (improved accuracy on 4 out of 6 cases) or not (improved accuracy on 2 out of 2 cases). The small number of cases here precludes meaningful statistical analysis, but the trend reflects previous findings in the literature that confirm the usefulness of reflecting. Hess and colleagues demonstrated that, even in the absence of a guided reflective process, reconsidering answers on complex questions in multiple choice exams can improve scores (Hess, Lipner, Thompson, Holmboe, & Graber, 2015).

Also notable is the related finding that guided reflection instilled confidence in participants. This effect appears to be rooted in the gathering of supporting evidence and the visible structure provided by the guided reflection process; participants could ‘see’ that they were adequately accounting for all the necessary information.

Reinforcement of diagnostic confidence may be something of a double-edged sword, as overconfidence has been identified as a potential cause of diagnostic error (Berner & Graber, 2008), and autopsy studies have indicated that diagnostic confidence does not correlate with diagnostic accuracy (Podbregar et al., 2001). As described in Section 2.3.1, lack of feedback on diagnostic accuracy and lack of visibility of errors lead to poor calibration of doctors’ confidence judgements. This is an important challenge for the field of patient safety going forward.

Arguably, the effect of guided reflection on diagnostic confidence (however well-placed that confidence may be) is not sufficient reason to recommend its widespread implementation in diagnostic practice. That being said, the findings must be interpreted with caution; the sample here was small and homogenous, and the materials used were novel and idiosyncratic. Further study is required to fully establish whether, and under what circumstances, reflective interventions of this sort may help doctors to correct their diagnostic course, or, indeed, if such interventions can have unintended

negative consequence for diagnosis, through overconfidence or other ill effects. This is explored in the experimental study presented in Chapter 5.

4.5.4 Theoretical implications

The verbal data provides support for a number of the features of the dual process model of diagnosis outlined by Croskerry (2009a). Most participants stated an initial diagnosis early, demonstrating pattern recognition and intuitive diagnostic processes often within the first few sentences of the vignette. Where they believed the case to be straightforward or “barn door”, participants allowed this initial diagnosis to stand, unchecked by analytical processes. However, when cued by complexity or ambiguity in the presentation, participants also demonstrated an ability to move between the two modes, switching to a more deliberate, analytical approach by means of guided reflection. This arguably serves as an example of the rational override described by Croskerry (2009a); even in the presence of reasonable hypotheses, System 2 monitoring and oversight was invoked to verify the diagnosis.

Furthermore, participants described errors in thinking that closely mirror cognitive biases described in the literature; anchoring and the failure to adjust from an initial hypothesis, diagnostic momentum, confirmation bias, and elements of fundamental attribution bias all may be identified in the verbal reports. This suggests that the heuristics and biases described in the extensive commentary on the topic have some basis in the lived experience of diagnosticians, and provides support for their conceptualisation as potential sources of error (Saposnik et al., 2016).

Participants regarded the guided reflection process as helpful in overcoming these biases, serving as a reminder to fully consider alternatives and not become anchored or fixed on an initial hypothesis, although participants seldom ultimately

changed their minds about a diagnosis. As mentioned above in the discussion of diagnostic confidence, the subjective perception of an intervention's usefulness may not map closely to its actual impact on objective outcomes.

4.5.5 Limitations

The present study is subject to a number of limitations. First, the sample was predominantly female, and relatively small. All participants were undergoing specialist training, and so the findings may not be generalizable to more experienced clinicians. Including senior doctors, even in small numbers, would have added nuance to the findings and potentially allowed for useful comparisons. However, it was decided that efforts should be focused on recruitment of junior doctors to ensure an adequate sample size within the timeframe of the study. Second, the sample was also self-selecting; this is of note in relation to the question of whether and when doctors choose to make use of guided reflection, as those who chose to take part may have found the idea of the intervention particularly compelling. This is difficult to avoid, particularly with the relatively high participant burden of this study (i.e., time commitment of one hour, potential discomfort of thinking aloud and being recorded). However, higher participation rates and a more representative sample could be achieved through increased compensation or incentive to participate (e.g., reimbursement of travel expenses, provision of refreshments, payment in vouchers). Future studies may also benefit from exploration of the characteristics of participants and non-participants, if metrics exist for relevant variables, to understand how the two groups may differ. Third, participants were not given a 'warm-up' session prior to data collection to get used to the process and receive feedback on their verbalisations from the experimenter; the quality of data elicited could be improved in future studies by including this step.

Fourth, the data collection sessions were subject to tight time constraints, which in some cases limited the number of cases participants were able to complete and may have hindered full use of the guided reflection for some participants. (However, time constraints are a daily reality for doctors engaged in diagnosis, and so this limitation may not be so troubling.) Fifth, there is scope for inter-rater agreement to be improved, as it falls slightly short of the .70 or .80 score that would indicate strong agreement according to many guidelines (McHugh, 2012). Improvement in agreement may be achieved through greater use of clarifying questions during the interview on the part of the researcher, development of the initial coding frame based on a larger sample of data, and agreement between the coders on mutual expectations around the level of detail required in the coding. Sixth, this study is subject to the same limitations as other studies that use fictional written vignettes as stimuli for diagnostic reasoning. Although the vignettes used were regarded by participants as being appropriate and of a high quality, participants commented on the artificiality of the format; the relevant information in a case is seldom presented in so organised a fashion and is seldom available in its entirety to doctors at the outset. Instead, information emerges over time as collateral is gathered, observations are carried out, and tests are completed. The vignettes used were therefore especially conducive to the use of guided reflection in a way that real-life cases may not be; this is worthy of further exploration in future studies.

4.5.6 Future research

This was an exploratory study, and as such, no control group was included for comparison purposes. This is an important potential next step in examining how doctors choose to use guided reflection and other interventions like it, and will allow

researchers to examine whether the process alters the types of evidence doctors consider or the steps they go through in reaching a diagnosis. (Although there were cases where doctors in this study chose not to use the guided reflection intervention, the fact that they had the process explained to them at the outset means that these cases cannot serve as true comparisons; the influence of the guided reflection approach may still be present in how doctors approached them.) Doctors in this study claimed that their natural diagnostic process was similar to the guided reflection process; an expanded version of the present study with a control group would make this comparison possible.

Future studies of guided reflection may also benefit from recruiting broader samples, including more experienced doctors (who are generally absent from this literature) and doctors from other medical specialties, to understand how diagnostic reasoning and the effects of interventions for same differ across fields.

Designers of future studies should also be mindful of the role of case complexity and difficulty. Support for the effectiveness of guided reflection has been primarily identified for complex cases (Mamede et al., 2008); therefore, vignettes used for experimental study should be sufficiently complex to allow this effect to be detected (e.g., comorbid diagnoses, ambiguity, red herrings). Case difficulty should also be pitched appropriately for the target population. If cases are too easy and familiar, participants may solve them immediately and accurately without reflection. If cases are too difficult, or feature diagnoses that might be reasonably expected to go beyond the content knowledge of the participants, participants' ability to effectively utilise the guided reflection process may be limited. In this way, cases should be thoughtfully calibrated in order to avoid ceiling and floor effects, where the stimuli themselves limit the range of possible responses. This is arguably a strength of the present study; the

cases ranged appropriately in difficulty and complexity, reflected in both participant comments and the range of diagnostic responses.

Certain components of the process are highlighted as being more helpful or more natural to use. As evidence is gathered for the basic efficacy of this style of intervention, the next logical step may be to consider refining the intervention by streamlining or simplifying it, to maximise its efficiency and usability. Further to this, Chapter 5 will present an experimental study comparing a short and long version of the guided reflection intervention to examine the relative benefits of instructions to generate more or fewer alternatives.

Basic pen-and-paper vignette studies do not allow for the sort of iterative, purposeful pursuit of information that characterises real-world clinical encounters; information is presented in a static form in its entirety. More sophisticated techniques, such as high-fidelity patient simulation (e.g., Burbach, Barnason, & Thompson, 2015) and standardised patients (e.g., Maupomé, Schrader, Mannan, Garetto, & Eggertsson, 2010), while highly resource-intensive, are interactive and offer researchers the opportunity to simulate the full complexity of diagnosis in real time, observing not only how doctors make sense of the information they are given but also how that information guides their search for more (Schubert et al., 2013). Methods such as eye-tracking protocols or interactive computerised presentations of the vignettes were explored for this study, but unfortunately ruled out due to resource limitations; these techniques may be used in future studies to enhance visibility of the information doctors seek out and attend to in real-time. Observational studies of behaviour have already begun to shed light on the transition from intuitive, non-analytical thinking to deliberate, effortful thinking in surgical settings (Moulton, Regehr, Lingard, Merritt, & MacRae, 2010a);

think-aloud methods offer the possibility of exploring the same transition in diagnosis with fine-grain, moment-to-moment verbal data.

4.5.7 Conclusions

Guided reflection interventions for diagnostic accuracy have had some success to date in the literature; this study has expanded on this work by using a think-aloud protocol to illuminate how doctors make use of the reflection process in real time. The findings reveal that, given the option under laboratory conditions, doctors used the process in most cases, with case complexity being a key driver of this choice, and that they regarded the intervention as straightforward and easy to use with potential for transfer to real-world settings. Improved accuracy was noted in the few cases where doctors changed their minds about a diagnosis in the course of the exercise, both with and without guided reflection. All components of the intervention were regularly used, particularly the gathering of supporting and refuting evidence for given hypotheses. The results show that think-aloud protocols can offer rich insights into not only the outcomes but also the processes of cognitive interventions; future research should continue to explore more sophisticated techniques in order to address the limitations of this study and to expand the generalisability of the findings therein.

Chapter 5: Impact of Short and Long Guided Reflection Interventions for Diagnostic Reasoning: An Experimental Study²

5.1 Introduction

A growing body of literature supports the use of reflection as a powerful tool to aid and verify diagnosis (Mamede & Schmidt, 2017). As described in Chapter 3, specific guided reflection interventions to improve diagnostic accuracy have been found to be successful in experimental vignette-based studies (e.g., Mamede et al., 2010; Mamede, Schmidt, & Penaforte, 2008; Mamede, Splinter, van Gog, Rikers, & Schmidt, 2012; Myung, Kang, Phyo, Shin, & Park, 2013). In Chapter 4, it was found that guided reflection interventions can be used in diagnosing psychiatric cases, and that junior doctors found the process accessible and helpful in enhancing their natural diagnostic process.

However, questions remain to be answered before these interventions can be confidently recommended to medical educators and practitioners. There has been limited commentary on the potential of interventions of this sort to have unintended or adverse consequences, or on the trade-offs inherent in promoting one mode of reasoning over another (Norman & Eva, 2003). The aim of this study is to examine three potential pitfalls that may arise in using guided reflection as a diagnostic process.

5.1.1 Potential for unintended consequences

First, there is the concern that adopting reflective methods of diagnosis is time-consuming. The adoption of a more reflective approach to diagnosis may lead to delays

² A version of this chapter has been submitted for publication in *Frontiers in Psychology* and is under review at the time of thesis submission. The chapter expands on the paper by the inclusion of full details of the pilot study and a more complete description of the methods and discussion.

in diagnosis in both the short term (longer history-taking, second-guessing by clinicians, bedside consideration) and long term (additional testing, ‘paralysis-by-analysis’) (Scott, 2009). These delays have obvious implications for the patient awaiting treatment, and also take time from other patients requiring attention, impacting on the clinician’s general efficiency (Berner & Graber, 2008). Therefore, it is important to design the most efficient version of any such diagnostic intervention and to identify the ‘active ingredients’. Specifically in relation to guided reflection, it will be useful to determine whether it is the analysis of symptoms that confirm or disconfirm a diagnosis that improves diagnostic performance, rather than the generation of an extensive list of alternatives. This study therefore compares a short version of the guided reflection task with a long version, wherein diagnosticians are asked to generate either two or six alternative diagnoses. If the two versions produce similar effects, it may suggest that diagnostic benefits are possible even with a relatively short reflective process. The study also compares these reflective interventions with a control condition, wherein participants diagnose cases based on their first impressions, to determine the effectiveness of the reflective interventions in improving diagnostic performance among novice diagnosticians.

Second, there is the concern that specific attempts at debiasing can have deleterious effects on reasoning and compound other biases (Lilienfeld, Ammirati, & Landfield, 2009). For example, a clinician attempting to make use of a ‘consider the alternatives’ cognitive forcing strategy (Croskerry et al., 2013b) may find that alternative diagnoses do not come easily to mind, leading them to conclude that there are not many good alternatives – and thereby strengthening their confidence that their working diagnosis is correct, by means of the availability bias. This effect has been confirmed in the psychology literature; Sanna, Schwarz and Stocker (2002) found that

participants who were asked to generate many counterfactual alternatives demonstrated increased hindsight bias, compared to participants who were asked to generate only a few (see Schwarz (2005) for further discussion; for an overview of hindsight bias, see Section 2.5.3). Therefore, this study examines whether diagnosticians who must generate a large number of alternatives become more confident in their initial diagnosis (indicating a hindsight bias) than those who must generate a smaller number of alternatives.

Third, unnecessary correction or overcorrection can occur when people are unsuccessful in their attempts to correct a judgement that they believe to have been biased (T. D. Wilson, Centerbar, & Brekke, 2002). Support for these effects may be found in the psychological literature: participants adjusted their initial ratings of holiday destinations unnecessarily when they received cues that their initial judgements may have been biased, even though no manipulation had been made to bias judgements (Petty & Wegener, 1993). In another study, when a manipulation to bias judgements had been made and participants were subsequently made aware of this, they over-adjusted their ratings to compensate (Stapel, Martin, & Schwarz, 1998). However, as discussed in the previous chapter, recent work by Hess and colleagues has demonstrated the benefit of reconsidering responses on multiple choice exams among residents, who were twice as likely to change their answers from incorrect to correct as to change from correct to incorrect (Hess et al., 2015). On this basis, the present study also examines what happens when an initial diagnosis is changed as the result of extensive reflection, and compares whether the short or long version of the task is more likely to lead to correction of an incorrect initial diagnosis or a correct initial diagnosis being discounted in favour of an incorrect diagnosis (which shall be described as ‘backfiring’ for the purpose of this study).

5.1.2 Aims and hypotheses

Therefore, the hypotheses under examination are as follows:

- (1) The accuracy of final diagnoses will differ between participants who are required to generate a greater number of alternative diagnoses and those who are required to generate fewer alternatives, and between both these sets of participants and those who are required to generate no alternatives.
- (2) Participants who are required to generate a greater number of possible alternative diagnoses will be less likely to lose confidence in their original diagnosis over the course of the guided reflection task than those who are required to generate fewer alternatives.
- (3) Where a final diagnosis differs from the initial diagnostic hypothesis, rates of correction (change from an incorrect initial diagnosis to a correct final diagnosis) or backfiring (change from a correct initial diagnosis to an incorrect final diagnosis), will differ between participants who are required to generate a greater number of alternative diagnoses and those who are required to generate fewer alternatives.

This study contributes a number of novel points to the literature. First, the study represents an attempt to identify how the guided reflection method may be translated from an extensive educational aid into a practical, concise tool for use in diagnosis. To the author's knowledge, this is the first empirical study of whether a shorter, less demanding version of the guided reflection method may be as effective as a longer, more thorough version. Second, as outlined above, the potential for unintended consequences or even diagnostic backfiring has been largely neglected in the literature to date; this study attempts to address some of these concerns empirically. Third, the

study draws more extensively than most studies to date on perspectives from psychological science and highlights in a more nuanced way the potential for interaction between individual psychological mechanisms (specifically, the interaction between the availability bias and hindsight bias as a potential area of vulnerability for diagnosticians).

5.2 Method

5.2.1 Design

An experiment was conducted using a between-groups design. Participants were asked to diagnose four fictional clinical cases, by first impressions (control condition) or by using a short or long guided reflection process. Participants rated their confidence in their initial diagnostic hypothesis at intervals throughout the process, on a scale of 1-6, where 1 indicated complete uncertainty and 6 indicated complete certainty. There were two outcomes of interest: 1) change in confidence judgements in the initial diagnostic hypothesis, and 2) accuracy of final diagnosis.

5.2.2 Participants

One hundred and eighty-six students were recruited from an undergraduate medical course during their psychiatry rotation. Students were in their fourth ($n=101$) and fifth ($n=85$) years of study. Although no gender information was gathered about the students who took part in the study, approximately 58% of students overall in each year group were female. Power calculations indicated a sample size of 165 (55 per group) for a medium effect size with 0.80 power for ANOVA; this is in line with some studies in the literature upon which the present study builds (e.g., Ibiapina, Mamede, Moura, Elói-

Santos, & van Gog, 2014; Myung et al., 2013; Reilly, Ogdie, von Feldt, & Myers, 2013). Participants were randomly allocated to control ($n=64$) or one of two experimental conditions, completing either a short ($n=58$) or long ($n=64$) diagnostic table.

5.2.3 Materials

5.2.3.1 Vignettes

Participants were asked to diagnose a series of four fictional clinical case scenarios. These vignettes are drawn from a bank compiled by Charles P. Friedman and colleagues at University of Michigan and have been used in similar research on diagnostic reasoning (Friedman et al., 1999; Payne, 2011). The vignettes are based on real patient cases and represent both common and uncommon or atypical presentations. The vignettes contain information on the chief complaint and history of presenting illness, relevant past medical and social history, the physical examination and laboratory data, and average 450 words in length. The case authors provide a definitive correct diagnosis for each case, which is used as the gold standard for participant responses; the correct diagnoses were appendicitis, amoebic liver, colon cancer and Guillaine Barré Syndrome. A member of the research team adapted some of the language in the cases for Irish participants. In order to maintain the integrity of future studies with these vignettes, they are not included in the Appendices of this thesis; however, the vignettes constructed for the think-aloud study presented in Chapter 4 follow a similar format and are included in Appendix H.

5.2.3.2 Diagnostic tables

For each clinical scenario, participants in the control condition were asked to write down the first diagnosis that comes to mind and to rate their confidence in this diagnosis (see Appendix R).

For each clinical scenario, participants in the experimental conditions were asked to complete a guided reflection table in order to reach a final diagnosis (see Appendix R). The table followed the procedure laid out in previous guided reflection studies (Mamede et al., 2008) and asked participants to write down the first diagnosis that comes to mind, along with a number of alternative diagnoses. Participants wrote down details to support and refute each diagnosis, as well as details that would be expected if the hypothesis were true, but that were missing from the scenario. Finally, participants were asked to rank their diagnostic hypotheses and to make a final diagnosis. They rated their confidence in the original hypothesis at four times: immediately after stating their initial hypothesis; after gathering evidence for this hypothesis; after generating alternatives; and after ranking their alternatives. They rated their confidence in their final diagnosis immediately after stating it.

Participants in the first experimental group (“short”) received a table asking them to generate two possible alternative diagnoses. Participants in the second experimental group (“long”) received a table asking them to generate six possible alternative diagnoses.

5.2.4 Procedure

Following receipt of ethical approval, a recruitment message was sent to potential participants (see Appendix O). Recruitment and data collection took place during four lecture sessions. Participants were verbally informed about the research by

two members of the research team, provided with an information sheet (see Appendix P), and invited to take part. Booklets containing the instructions, vignettes, and diagnostic tables (see Appendix R for instructions and tables) were randomly distributed to the students, and the students were instructed to work through the cases in silence. Students were given 60 minutes to complete the booklet and were advised to spend no more than 15 minutes on each case; they were notified of the time remaining at 15-minute intervals. Following participation, students were debriefed (see Appendix Q) and a short explanation of the key principles of diagnostic error was provided.

5.2.5 Analysis

Drawing on previous studies of guided reflection interventions (e.g., Mamede et al., 2008), diagnostic accuracy was scored in three ways.

- (1) Under ‘first impression’ scoring, only the initial diagnostic hypothesis for each case was considered. The case was counted as having been diagnosed correctly only if the participant selected the correct diagnosis as their *first diagnostic hypothesis*.
- (2) Under strict scoring, a case was counted as having been diagnosed correctly only if the participant selected the correct diagnosis as their *final diagnosis*.
- (3) Under lenient scoring, a case was counted as having been diagnosed correctly if the participant included the correct diagnosis as *one of their alternatives*, whether or not they ultimately selected this as their final diagnosis.

The gold standard diagnosis for each case provided by the case authors was used as the correct diagnosis; for three cases, a small number of additional diagnoses, which represented more specific diagnoses that fitted under the gold standard diagnosis, were also scored as correct (see Appendix S).

Change in confidence was calculated by subtracting the participant's confidence rating in their final diagnosis from their confidence rating in their first diagnosis. This comparison was chosen as it represents the most important shift in confidence (confidence at the beginning of the process, where the initial hypothesis has been shown to be particularly powerful, (Kostopoulou et al., 2017), and at the end, when action will presumably be taken) and most succinctly reflects the change in confidence over the course of the reflection exercise.

With regard to parametric assumptions, homogeneity of variance was tested using Levene's statistic, and normality was tested through visual inspection of histograms. Where data did not comply with parametric assumptions, parametric and corresponding non-parametric tests were run. These instances are noted below. Statistical significance was set at $p < .05$ for all tests.

5.3 Pilot study

5.3.1 Aims and objectives

As the materials for this study were untested with this population (Irish senior medical students), a pilot study was conducted to explore three aspects of the protocol:

- (1) Feasibility of procedure: to determine how many vignettes could reasonably be diagnosed using the short, long and control diagnostic tables within one hour.
- (2) Appropriateness of vignettes: to explore participant responses to vignettes, to determine the difficulty of the vignettes, and to highlight specific areas of confusion or difficulty.

- (3) Acceptability of materials and instructions: to explore participant responses to the visual layout and comprehensibility of the vignettes and diagnostic tables, and to examine the clarity of the instructions given.

5.3.2 Method

5.3.2.1 Design

Similar to the main study, an experiment was conducted using a between-groups design. Participants were asked to diagnose five fictional clinical cases, by first impressions (control condition) or by using a short or long guided reflection process.

5.3.2.2 Participants

Volunteers for the study were sought at a lecture for fourth-year medical students. Four students agreed to take part in the pilot study and were randomly allocated to control ($n=1$) or one of two experimental conditions, completing either a short ($n=1$) or long ($n=2$) diagnostic table. These students were then excluded from the sample for the main study.

5.3.2.3 Materials

5.3.2.3.1 Vignettes

The vignettes used for the study are drawn from a bank compiled by Charles P. Friedman and colleagues at University of Michigan and have been used in similar research on diagnostic reasoning (Friedman et al., 1999; Payne, 2011). Five vignettes

were chosen for the pilot study: the correct diagnoses were appendicitis, amoebic liver, colon cancer, Guillaine Barré Syndrome, and ulcerative colitis.

5.3.2.3.2 Diagnostic tables

The diagnostic tables used for the pilot study were identical to those used in the main study; these are outlined above in Section 5.2.3.2 (see also Appendix R).

5.3.2.4 Procedure

Data collection took place in a seminar room with four participants. Booklets containing the vignettes and diagnostic tables were randomly distributed to the participants, and the participants were instructed to work through the cases in silence. Participants were given 60 minutes to complete the booklet. As each participant completed the booklet, the researcher and participant adjourned to a neighbouring room, where the researcher asked for the participant's opinions and feedback on the diagnostic table, each of the vignettes, and the procedure as a whole, which were recorded in writing. Following participation, participants were debriefed and thanked for their time.

5.3.2.5 Analysis

Diagnostic accuracy was scored as for the main study (see Section 5.2.5). Notes on participant feedback were reviewed by the researcher, and comments relating to the pilot study objectives were compiled and discussed with the research team to make adjustments to the materials and procedure. The number of vignettes completed by each participant was also noted.

5.3.3 Findings and actions taken

(1) Feasibility of procedure:

Findings: Participants completed a mean of 3.75 cases in the time available, and participants in the short and long condition commented that there was insufficient time available to complete all five cases to the level of detail required.

Actions taken: It was decided to allow 60 minutes for the completion of four cases, to advise participants to spend no more than 15 minutes on each case, and to notify participants of the time remaining at 15-minute intervals.

(2) Appropriateness of vignettes:

Findings: Participants generally responded positively to the vignettes, describing the cases as being challenging but of appropriate difficulty for their level of experience and knowledge. No changes to language or the amount of detail presented were recommended. Diagnostic accuracy was similar across all five cases.

Actions taken: For one case (ulcerative colitis), when participants diagnosed incorrectly, they consistently named the same incorrect, though somewhat related, diagnosis (diverticulitis). As such, the research team determined that there may be some factor in the case that systematically biased participants to this modal incorrect diagnosis. This case was therefore chosen for exclusion from the main study.

(3) Acceptability of materials and instructions:

Findings: Participants generally described the instructions and diagnostic tables as straightforward. The instruction to consider evidence that would be expected if the hypothesis were true, but that is in fact missing from the case, was highlighted as being somewhat more difficult to understand; this reflects similar observations by participants in the think-aloud study presented in Chapter 4. Two participants reported that they did not proceed from one case to the next in sequential fashion, instead diagnosing all of the cases based on first impressions and then returning to complete the diagnostic tables.

Actions taken: Explicit instructions to proceed sequentially through the booklet were incorporated into the instructions, along with additional text to clarify that the first diagnosis given should be the very first diagnosis to come to mind (rather than a diagnosis chosen following early differential diagnosis).

5.4 Results

5.4.1 Confirmation of manipulation

An independent samples t-test confirmed that the mean number of alternatives generated by participants in the long condition ($n=58$, $M=3.03$) was significantly higher than the mean number generated by participants in the short condition ($n=56$, $M=1.86$), $t(61.92) = 6.76$, $p < .001$, two-tailed.

As the number of alternatives generated did not meet parametric assumptions of normality and homogeneity of variance, a Mann-Whitney U-test was also performed; this also revealed a significant difference in the number of alternatives generated by

participants in the short condition ($n=56$, $Mdn=2.0$) and the long condition ($n=58$, $Mdn=3.0$), $U = 548$, $z=6.31$, $p<.001$, $r=.59$.

5.4.2 Diagnostic accuracy

Table 5.1 presents descriptive statistics for the proportions of cases diagnosed correctly under first impressions, strict scoring, and lenient scoring for each of the experimental conditions and each of the four cases.

Table 5.1

Means and standard deviations obtained for proportions of Cases 1, 2, 3 and 4 diagnosed correctly

Group	First impression		Strict (Correct final diagnosis)		Lenient (Correct diagnosis included in alternatives)	
	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>
<i>Case 1: Appendicitis</i>						
Control	64	0.41 (0.50)	64	0.41 (0.50)*	64	0.41 (0.50)*
Short	56	0.59 (0.50)	52	0.54 (0.50)	56	0.86 (0.35)
Long	59	0.56 (0.50)	55	0.45 (0.50)	58	0.91 (0.28)
<i>Case 2: Amoebic liver</i>						
Control	63	0.52 (0.50)	63	0.52 (0.50)*	63	0.52 (0.50)*
Short	56	0.50 (0.50)	52	0.35 (0.48)	54	0.67 (0.48)
Long	59	0.37 (0.49)	53	0.47 (0.50)	58	0.69 (0.47)
<i>Case 3: Colon cancer</i>						
Control	63	0.54 (0.50)	63	0.54 (0.50)*	63	0.54 (0.50)*
Short	56	0.59 (0.50)	53	0.58 (0.50)	56	0.77 (0.43)
Long	59	0.55 (0.50)	53	0.51 (0.51)	58	0.72 (0.45)
<i>Case 4: Guillaine Barré syndrome</i>						
Control	62	0.43 (0.50)	63	0.43 (0.50)*	63	0.43 (0.50)*
Short	49	0.33 (0.47)	45	0.36 (0.48)	49	0.59 (0.50)
Long	57	0.37 (0.49)	49	0.35 (0.48)	56	0.57 (0.50)
<i>Note:</i>						
* Under the control condition, the first impression diagnosis also constitutes the final diagnosis and no alternative diagnoses are offered. Therefore, the scores for the control condition under first impression, strict scoring, and lenient scoring are all equal.						

Table 5.2 presents descriptive statistics for the proportions of all cases diagnosed correctly under first impressions, strict scoring, and lenient scoring for each of the experimental conditions.

Table 5.2

Means and standard deviations obtained for proportions of all cases diagnosed correctly

Group	First impression		Strict (Correct final diagnosis)		Lenient (Correct diagnosis included in alternatives)	
	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>
Control	64	0.47 (0.26)	64	0.47 (0.26)*	64	0.47 (0.26)*
Short	56	0.49 (0.24)	55	0.42 (0.24)	56	0.70 (0.23)
Long	59	0.45 (0.24)	57	0.41 (0.26)	58	0.72 (0.25)

Note:

* Under the control condition, the first impression diagnosis also constitutes the final diagnosis and no alternative diagnoses are offered. Therefore, the scores for the control condition under first impression, strict scoring, and lenient scoring are all equal.

A one-way between-groups analysis of variance revealed no significant differences in the accuracy of final diagnosis (strict scoring) between any of the three conditions, $F(2, 173)=0.841, p>.05$.

A one-way between-groups analysis of variance revealed a significant difference in accuracy of diagnostic alternatives (lenient scoring) between the three conditions: $F(2, 175)=18.992, p<.001$. The effect size, calculated using eta squared, was 0.18. *Post-hoc* comparisons using the Tukey HSD test indicated that the mean scores for the short and long conditions ($M=0.70, SD=0.23$ and $M=0.72, SD=0.25$ respectively) did not differ significantly from one another, but were significantly higher than the control condition ($M=0.47, SD=0.26$).

5.4.3 Change in confidence

An independent samples t-test revealed no significant difference in change in confidence between the short ($M=0.11$, $SD=0.59$) and long ($M=0.10$, $SD=0.61$) groups: $t(109)=0.52$, $p>.05$.

5.4.4 Rates of correction and backfiring

Across both the short and long form groups, participants changed their final diagnosis from their initial diagnostic hypothesis in 17% of cases ($n=70$ cases). Table 5.3 presents the contingency table for diagnoses in Cases 1, 2, 3 and 4 changed from incorrect to correct (corrected) and diagnoses changed from correct to incorrect (backfired) in the short and long form conditions.

Table 5.4 presents the contingency table for diagnoses in all cases changed from incorrect to correct (corrected) and diagnoses changed from correct to incorrect (backfired) in the short and long form conditions

A Chi-square test for independence indicated no significant association between group and correction/backfiring of diagnosis, $\chi^2(1, n=70)=0.972$, $p>.05$, $\phi=.118$.

A binomial test indicated that across both conditions, the proportion of backfired changed diagnoses (57%) was not significantly greater than the proportion of corrected changed diagnoses (43%), $p>.05$ (two-tailed).

Table 5.3

Contingency table for changed diagnoses in Cases 1, 2, 3 and 4 in short and long conditions (n=70 cases)

	Corrected		Backfired	
	<i>n</i>	%	<i>n</i>	%
<i>Case 1: Appendicitis</i>				
Short	4	36%	7	64%
Long	2	22%	7	78%
Case 1 Total	6	30%	14	70%
<i>Case 2: Amoebic liver</i>				
Short	3	23%	10	77%
Long	4	100%	-	-
Case 2 Total	7	41%	10	59%
<i>Case 3: Colon cancer</i>				
Short	3	43%	4	57%
Long	4	44%	5	56%
Case 3 Total	7	44%	9	56%
<i>Case 4: Guillaine Barré syndrome</i>				
Short	6	55%	5	45%
Long	4	67%	2	33%
Case 4 Total	10	59%	7	41%

Table 5.4

Contingency table for changed diagnoses in all cases in short and long conditions (n=70 cases)

	Corrected		Backfired	
	<i>n</i>	%	<i>n</i>	%
Short	16	38%	26	62%
Long	14	50%	14	50%
Total	30	43%	40	57%

5.5 Discussion

5.5.1 *Diagnostic accuracy*

Results showed that neither participants using the short nor long version of the guided reflection intervention were more accurate in their final diagnoses than participants who diagnosed based on their first impressions only. This is in contrast with some previous studies that found a beneficial effect for diagnosis, particularly of complex cases (Ilgen et al., 2011; Mamede, van Gog, et al., 2010; Mamede et al., 2008; Myung et al., 2013). Two factors may explain these results. First, the lack of improvement with reflection may arise due to the reasoning strategies employed by medical students in particular; there is some evidence that novice diagnosticians rely more on analytical, rule-based reasoning processes than experts (Kulatunga-Moruzi et al., 2001). If this is the case, participants in this study may have offered a ‘first impression’ diagnosis that was in fact the product of an analytical reasoning style that more closely resembles the guided reflection model than the non-analytical, pattern-based mode of reasoning intended to be elicited by requesting a ‘first impressions’ diagnosis. Future studies may benefit from methodological adjustments to parse these effects, for example restricted time limits for non-analytical ‘control’ conditions or think-aloud protocols. Second, although the cases were clearly difficult, given the low accuracy scores overall, they were not structurally complex; no secondary diagnosis was present in any case. Previous studies indicate a particular beneficial effect of reflective practice on accuracy of final diagnosis for complex cases only (Mamede et al., 2008). This may explain the absence of an effect on accuracy of final diagnoses in this study.

Results showed that the guided reflection intervention was associated with the generation of more accurate hypotheses than diagnosis based on first impressions only; this finding is unsurprising, as it is logical to expect a correct diagnosis to be named with three or seven ‘chances’ compared to just one. However, it was also shown that participants who were asked to generate two alternative diagnostic hypotheses provided an accurate diagnostic hypothesis as frequently as those who were asked to generate six alternatives. In this way, instructions to generate many alternatives did not increase the accuracy of alternatives, suggesting that a relatively short version of the guided reflection process may yield similar accuracy rates as a longer, more burdensome version.

Previous studies of the guided reflection method have not instructed participants to generate a particular number of alternative diagnoses for test cases, and these findings suggest that the number of diagnostic alternatives generated may not be a crucial component of the success of guided reflection interventions. The metacognitive processes induced by the instruction to generate alternatives and the evidence-gathering process to investigate those alternatives may be of more value; that is, the quality of reasoning processes may contribute more than the exhaustiveness or extent of reasoning to diagnostic success. This is in line with existing research; while ‘consider the alternatives’ is a commonly suggested strategy, it has seldom been examined empirically and the evidence that it is, in itself, sufficient to improve accuracy is lacking (Feyzi-Behnagh et al., 2014; Regehr et al., 1994). Specific reasoning instructions that draw attention to the evidence in the case have been found to be substantially more effective than more general metacognitive instructions to take care and consider all the data (Mamede & Schmidt, 2017). Not all reflective interventions are alike, and not all studies find evidence for their effectiveness; it is therefore important to identify the

‘active ingredient’ in the guided reflection process to ensure that the most efficient version of the task may be recommended to diagnosticians and medical educators. A number of variables may be targeted for future experimentation to identify which aspects of the process are most critical, including the timing of the reflective exercise within the diagnostic process, the extent of evidence gathered, verification of hypotheses, and total time spent on the task (Mamede & Schmidt, 2017). These findings suggest that the optimal version of a guided reflection intervention may include instructions to generate only a few alternatives.

It is possible that the generation of a greater number of alternatives offered no additional benefit to diagnostic accuracy due to the extra cognitive load required by a more extensive differential diagnosis (Payne, 2011). This may be particularly the case for the novice diagnosticians who took part in this study; even with a pen-and-paper booklet to assist them, they may have struggled to consider more than two or three alternatives at a time. Indeed, even when asked to offer six alternatives, participants offered an average of only three, and six alternatives were offered in only 12% of cases ($n=236$ cases).

5.5.2 Diagnostic confidence

Participants who completed a longer version of the task exhibited a similar change in confidence as those who completed the shorter version of the task; that is, the change in confidence across both groups was minimal. This study therefore finds no support for concerns raised in the literature (Lilienfeld et al., 2009) that an availability bias induced by such a reflective exercise may compromise diagnostic accuracy, nor does it find evidence that generating alternatives bolsters or diminishes diagnostic confidence among this participant group.

5.5.3 Correction and backfiring

Where a final diagnosis differed from the initial diagnostic hypothesis, participants who were required to generate a greater number of alternatives were no more likely to correct an incorrect initial diagnostic hypothesis than those who were required to generate fewer alternatives.

Two phenomena are of interest here. First, participants seldom changed their minds about their initial diagnosis; in 83% of cases, the original diagnostic hypothesis was retained as the final diagnosis. Second, participants were as likely to change from an incorrect to a correct diagnosis as they were to change from a correct to an incorrect diagnosis. This contrasts with the findings in a previous study with residents, who changed from an incorrect to correct answer twice as often as the reverse on a multiple-choice exam (Hess et al., 2015). This suggests that, for this student participant group, the utility of the guided reflection exercise in catching errors is relatively limited, and opens up inexperienced diagnosticians to the risk of ‘talking themselves out of’ a correct diagnosis. These students may not yet have attained sufficient content knowledge to reliably identify cases where they may be wrong; more senior doctors tend to be poorly calibrated judges of their own accuracy (Podbregar et al., 2001), and the subjective experience of being wrong “feels like being right” (Schulz, 2011, p18).

These findings, around diagnostic accuracy generally and rates of correction and backfiring specifically, highlight the importance of the first impression in diagnosis. Participants were able to identify accurate alternatives to their initial diagnostic hypothesis, but seldom actually changed their minds from their first impressions. This reflects findings from the think-aloud study presented in Chapter 4 and also existing research; studies have previously found a strong association between the first diagnostic

impression and subsequent diagnostic and treatment choices (Kostopoulou et al., 2017), and doctors tend to be highly confident in their initial diagnosis (Monteiro et al., 2015). While this potent effect of the first impression may leave doctors vulnerable to certain cognitive pitfalls, such as an anchoring bias or confirmation bias (Croskerry et al., 2013a), the results of this study suggest that the novice diagnostician's initial diagnosis is as likely to be correct as a final diagnosis chosen after a relatively lengthy reflective exercise. As such, reflection strategies do not reliably help the novice to choose the correct final diagnosis, even when they identify the correct answer as a possibility.

5.5.4 Theoretical implications

Some implications for the dual process model as it has been applied to diagnostic reasoning may also be identified. First, assuming that the initial hypothesis offered by participants was truly the product of non-analytical reasoning processes, the findings suggest that such processes are strongly influential and even dominant in determining the accuracy of the final diagnosis. In this study, the first diagnosis to come to mind served as a powerful anchor from which participants seldom deviated. While a rational override, wherein analytical processes led to reconsideration and subsequent selection of a correct diagnosis, is possible, this does not appear to be the norm, even with mandated and extensive analytical reasoning. In this way, the findings challenge the widely held hypothesis that reconsideration of a diagnosis using analytical reasoning reduces error.

Second, the findings add support to the growing commentary that both modes of reasoning have value, and that non-analytical reasoning strategies must not be dismissed as error-prone and ineffective. While many errors were made by participants using this mode of reasoning in this study, analytical reasoning was not shown to be superior in reaching a correct final diagnosis. Non-analytical processes, therefore, must be brought

out of the shadow of their analytical counterparts. The challenge of improving these largely unconscious non-analytical processes, through training pattern recognition and increasing experience with a variety of presentations, is significant, but meeting it may yield dividends for the design of effective interventions.

5.5.5 Limitations and future research

A number of limitations should be acknowledged in the interpretation of these findings. First, participants were senior medical students, and not expert diagnosticians; as reliance on intuitive reasoning in diagnosis may increase over the course of a doctor's career (Ilgen et al., 2011), these findings may not be generalisable to doctors with more expertise. Medical students were selected as the population for this study as potential users of an educational intervention based on guided reflection, and also to ensure recruitment of an adequate sample size for statistical analysis; recruiting a sufficiently large sample of qualified doctors was not deemed to be feasible within the timeframe of the study. Second, given the relatively low accuracy scores across all conditions, the cases used in this study may have been too challenging for this participant group, which may have introduced a ceiling effect. Third, participants in the short and long conditions had equal amounts of time to complete the task. It is possible that participants in the short condition spent additional time on each case, compromising their adherence to the instruction to name the first diagnosis that came to mind, or that participants in the long condition did not have sufficient time to complete the task in full for each case. Fourth, participants in the long condition on average provided only three out of the required six alternatives. Although this is a statistically larger number of alternatives than provided by participants in the short condition, it is worth considering why participants provided fewer alternatives than required; it is possible that the cases were too clear-cut to generate a large number of reasonable

alternatives, or that the time allowed was not sufficient to carry out a full analysis of each alternative. These issues may be corrected with future studies. Fifth, as with the majority of studies in this area, the use of fictional vignettes and the laboratory environment in which this study was conducted limits the generalisability of findings to a clinical setting. Sixth, in considering the potential for extensive reflection to ‘backfire’, only cognitive factors were considered, such as availability bias, in the somewhat artificial context of all information being available to the diagnostician. The study does not allow us to determine the precise process by which participants change their mind about a diagnosis, only to observe the frequency with which this occurs. Non-cognitive issues, such as additional tests leading to inflated risk of false positives, overtesting, and overdiagnosis, are also enormously significant factors in potential backfiring (Scott, 2009). These factors are beyond the scope of this study but provide fertile ground for future studies.

An additional limitation reflects a potential confound in the study (and some others in the literature upon which the methodology was modelled); participants in the control condition only had the opportunity to list a single hypothesis, when in reality it is possible that these participants may have considered more than one diagnosis very rapidly from the early stages of each vignette. This study is unable to determine whether the greater accuracy of the alternatives generated by participants in the short and long conditions is an effect of the reflective intervention or merely an artefact of the restrictions placed on participants in the control condition. As such, future research should include a condition in which participants offer their first impressions without restrictions or imposed structure; comparison of accuracy in this condition compared with accuracy in the reflective condition(s) will help to elucidate the intervention’s effect and eliminate the present confound.

5.5.6 Conclusions

This study finds no evidence to support the use of the guided reflection method as a diagnostic aid for novice diagnosticians using standard methods in the literature. While our findings suggest that a shorter, less demanding version of the process can elicit accurate diagnostic hypotheses at a similar rate as a longer version, the accuracy of final diagnosis was ultimately not improved by any version of the reflection task with this participant group. The findings also highlight the durability of an initial diagnosis and the relative inertia of diagnostic decisions.

Future studies should continue to test alternative versions of the task to identify those factors that are most effective in improving accuracy in order to create the most efficient possible intervention. Identifying cases and situations in which reflection may offer the greatest benefit should also be a priority for this research area. It is also important that such studies examine the generalisability of such efforts to trainees and practitioners at all levels; understanding development differences in diagnostic reasoning may provide important insights into the mechanisms by which diagnostic expertise develops, as well as point to the most appropriate interventions for reducing error and enhancing reasoning at every stage of a doctor's career. Finally, methodological improvements to parse the effect of the intervention from other factors in experimental studies and to eliminate potential confounds, as outlined above, will be important in clarifying and strengthening the case for continued development of guided reflection interventions.

Chapter 6: General Discussion

6.1 Aims and contributions

The aim of the thesis was to examine a range of interventions for medical and psychiatric diagnostic reasoning using the dual process model of reasoning as an organising framework and to examine the extent to which the model has been utilised in intervention design. Following from the findings of a systematic review of interventions based on this model (Chapter 3), the thesis evolved into a consideration of guided reflection interventions specifically, utilising an empirical, mixed-methods approach and taking the dual process model as a theoretical foundation.

The thesis provides a significant contribution to the diagnostic reasoning literature, as previous research has largely neglected to fully utilise the dual process model as a framework to substantially inform intervention design and interpretation of outcomes. These studies use the dual process model as a lens to examine both previous research and the direct impact of interventions on diagnostic reasoning processes. Additionally, the work advances the use of qualitative methods in examining diagnostic interventions, and represents, to the author's knowledge, the first use of a think-aloud protocol to examine the impact of a reflective intervention for diagnosis. In this way, the theoretical framework and novel methods add greatly to the literature not only around which interventions work, but specifically around how they achieve their effects.

What follows is a discussion of the key implications of the research for theory, practice, education, and future research. An overall summary of the research from previous chapters will be presented first. The implications of these results and findings

will be highlighted, with suggestions for future research. Finally, the general limitations of the research presented in the thesis will be discussed.

6.2 Summary of research

Chapter 3 presented a systematic review of interventions to enhance analytical and non-analytical reasoning among medical trainees and doctors, with the aims of identifying, describing and appraising these studies. The review also aimed to assess the effectiveness of the interventions in improving diagnostic accuracy and to assess the extent to which the included interventions incorporated dual process theory as a conceptual basis for their design. The review of twenty-eight studies concluded that while many of studies found some effect of interventions, guided reflection interventions emerged as the most consistently successful in improving diagnostic accuracy, compared with educational interventions, checklists, cognitive forcing strategies, and diagnostic instructions. In addition, it was clear that the use of theory as a foundational rationale was relatively shallow for most studies. Only interventions based on the dual process model were included; other cognitive interventions for diagnostic reasoning, such as second opinions or use of decision support systems, were not reviewed here but also show promise in the literature (Graber et al., 2012).

Chapter 4 described a think-aloud protocol, in which thematic analysis was used to analyse how trainees in psychiatry and general practice made voluntary use of a guided reflection tool in diagnosing complex psychiatric clinical vignettes. The tool was utilised by a majority of participants and was described in positive terms. The findings highlighted the impact of reflection on the process of diagnosis and on diagnostic confidence, and participants broadly indicated that the tool could be successfully implemented in real-world diagnosis, particularly for ambiguous or complex cases.

Chapter 5 presented an experimental study comparing a short and long version of the guided reflection intervention, undertaken with 186 senior medical students. The results indicated that the reflective intervention did not elicit more accurate final diagnoses than diagnosis based on first impressions only, and suggested that a short version of the guided reflection process elicits similar performance as a long version. Notably, the results also revealed that participants seldom changed their minds, highlighting the importance of the first impression in diagnosis and the relative lack of utility of the guided reflection process as a tool to ‘catch’ errors, at least for inexperienced diagnosticians. It was posited that the lack of effect observed here may have been due to the relative structural simplicity of the stimuli, as evidence points to the utility of guided reflection in complex cases (Mamede et al., 2008).

6.3 Implications

6.3.1 Theoretical implications

While these studies did not set out to test the dual process model as it has been applied to diagnostic reasoning, the findings offer some support for its validity and relevance. The verbal data presented in Chapter 4 provides support for a number of the features of the dual process model of diagnosis outlined by (Croskerry, 2009a); pattern recognition and early intuitive diagnostic processes were manifest, and these were checked or, in relatively few cases, overridden by analytical processing when participants perceived the task to be of sufficient complexity, ambiguity, or difficulty. Furthermore, participants described errors in thinking that closely mirror cognitive biases described in the literature; anchoring and the failure to adjust from an initial hypothesis, diagnostic momentum, confirmation bias, and elements of fundamental

attribution bias all were identified in the verbal reports. This suggests that the heuristics and biases described in the extensive commentary on the topic are recognised by doctors, and provides support for their conceptualisation as potential sources of error (Saposnik et al., 2016).

Findings from both the think-aloud study presented in Chapter 4 and the experimental study presented in Chapter 5 provide evidence for the powerful influence of the first impression and, by extension, of intuitive pattern-recognition processes. While a rational override, wherein analytical processes led to reconsideration and subsequent selection of a correct diagnosis, is possible, this does not appear to be the norm, even with extensive analytical reasoning.

Results from the systematic review of interventions based on the dual process model are arguably too mixed, and the targeted processes too vaguely defined, to make clear statements about the theoretical process and constructs most successfully targeted by the interventions. As the field matures, insights from implementation science may be incorporated into the literature on diagnostic reasoning in order to design interventions clearly grounded in psychological theory (Michie et al., 2011).

In relation to the specific case of psychiatry, the findings point to the continued need to fully integrate mental health and psychiatric diagnoses into the discourse around diagnostic error. Psychiatric diagnoses may reasonably be expected to shift over time, and assigning a label to a presentation may be deliberately delayed to allow for medium- to -long-term observation. Definitions of error that may be used in general medicine may require modification or further specification before being applied to mental health presentations (Phillips, 2014). Psychiatry provides a particularly interesting field in which to further examine the operations of the dual processes of diagnostic reasoning, where the diagnostic criteria of the DSM or ICD meet the self-

reported symptoms of patients and the holistic impression formation of diagnosticians; there is fertile ground here for hypothesis generation and exploration.

The findings lend support for the dual process model as a useful, if not perfect, framework for describing the diagnostic process. Crucially, it has significantly outperformed other models (e.g., hypothetic-deductive reasoning) in terms of stimulating research, commentary, and educational and practical use. Normative models, such as Bayesian reasoning, feature in the literature but are almost never used in practice; clinicians rely on instead on pattern recognition and ‘gist’ to make sense of diagnostic information, and in the main are able to do so successfully (Jain, 2016; Lloyd & Reyna, 2009).

6.3.2 Practical implications

The work presented in this thesis provides support for the use of dual process interventions in real-world diagnosis. The systematic review of interventions revealed a mixed picture of the efficacy of these interventions under laboratory conditions, while their effectiveness in real diagnostic settings remains almost entirely unexplored. There is precedent for the successful transfer of such interventions from laboratory study to field study and on to widespread implementation; the use of surgical checklists, such as the World Health Organisation’s Surgical Safety Checklist, provides a recent example of an evidence-based intervention that has been adopted in a wide variety of settings with demonstrable impact on care (Treadwell, Lucas, & Tsou, 2014).

The think-aloud protocol suggested a positive impact of a guided reflection intervention on diagnostic accuracy, and the method was generally popular among participants, who indicated that it would be beneficial as an optional tool in their real practice. As discussed in Section 4.5.3, a positive reception and feelings of increased

confidence alone do not necessarily lend support for the intervention's widespread adoption. It is important not to conflate satisfaction with effectiveness, a distinction commonly invoked in usability studies in the field of human factors (Frøkjær, Hertzum, & Hornbæk, 2000). When an intervention is ineffectual but well-liked, it is also important to consider the reason for satisfaction. A 2000 study in mental health provides an illustrative example, wherein respondents' satisfaction with post-trauma debriefing contrasted with a lack of improvement on objective measures of traumatic stress. It is possible that participant's positive feelings regarding the intervention resulted from iatrogenic effects, rather than the intervention itself (Kagee, 2002).

When considering the potential for guided reflection and similar interventions to be implemented in clinical settings, contextual factors are also important; participants believed that the intervention would be best suited to complex cases, and that its implementation would be hampered by time poverty. Opinions were mixed around whether the process should be incorporated on a mandatory or optional basis into assessment documentation. Reflection, even if included in a pro-forma assessment, is unlikely to be used by doctors for cases that appear simple due to doctors' poorly calibrated diagnostic confidence. Demanding that doctors apply the process to cases that appear simple carries risks; excessive time spent on straightforward diagnoses may reduce overall efficiency, and mandatory reflection may lead to frustration and dismissal of the process on the parts of doctors. However, given that doctors' diagnostic confidence is poorly calibrated, that it is difficult to determine which apparently simple cases require further examination, and that many diagnostic errors do arise in what appear to be very simple cases, an argument could be made for its inclusion despite these challenges. Indeed, evidence-gathering and ranking of alternatives fit neatly with

existing sections of pro-forma assessments that already require a brief differential diagnosis.

6.3.3 Educational implications

The use of the dual process model in general and the guided reflection intervention in particular as educational tools is an important avenue for continued research. Findings from the systematic review presented in Chapter 3 indicate that evidence for education about the dual process model, including biases, heuristics, and Systems 1 and 2 processing, as an intervention for diagnostic accuracy among early career medics is, to date, underdeveloped. Although studies showed positive effects of educational interventions on a range of outcome measures, evidence for their impact on diagnostic accuracy in particular is lacking. Medical trainees receive very little instruction in diagnostic reasoning, and there currently exists no dedicated curriculum on diagnosis in medical training. Even if comprehensive education to orient future practitioners to the cognitive underpinnings of clinical reasoning proves ineffective in inoculating clinicians from errors owing to cognitive bias, clinicians may learn to recognise biased cognition where it is employed by others. In light of increased enthusiasm for second opinions as a cognitive strategy to prevent diagnostic errors, this presents an exciting opportunity for continued work; if doctors cannot reliably debias themselves, perhaps they can debias one another (Dror, 2011; Elstein, 2009).

The experimental study presented in Chapter 5 indicated that diagnostic accuracy among senior medical students was not significantly improved by use of a guided reflection process; however, this contrasted with findings from other studies in the systematic review, which found support for the efficacy of some dual process interventions among medical students and junior doctors. It is possible that the

structural simplicity of the cases used for the study presented here, despite being relatively challenging, can account for these differences (see Section 5.5.1 for further details). Participants in the think-aloud study presented in Chapter 4 indicated that the guided reflection method was easy to learn and that they believed that it would be helpful in their practice going forward. As such, reflective and some other dual process interventions may be considered for inclusion in medical curricula at multiple stages of professional development, though the optimal time to introduce this material to diagnosticians remains unclear and further research is required.

6.4 Limitations

The limitations associated with the individual studies can be found in the respective chapters (see Sections 3.4.3, 4.5.5, 5.5.5). This general limitations section will consider the overall limitations of the methods and approaches described in the thesis.

6.4.1 Theoretical limitations

The dual process model has achieved widespread acceptance in the medical reasoning literature and is the dominant framework by which diagnostic error is currently understood (Balogh et al., 2015). However, three limitations of the model may be noted, both in the features of the model itself and in how it has been applied to diagnostic reasoning.

First, the model provides a useful and comprehensive framework for understanding how diagnostic knowledge is retrieved and processed, but does not make clear statements about how diagnostic knowledge is organised. Alternative theories offer more clarity on this point; for example, mental model theory posits that a doctor

stores information as mental representations, and that these knowledge structures undergo change over time as doctors acquire expertise, transitioning from elaborated causal networks (causes and consequences of pathophysiological processes) to abridged networks (simplified causal models, incorporating signs and symptoms observed by the doctor in real patients) to illness scripts (list-like structures or prototypes) to instance scripts (real patient encounters that do not conform to the prototype) (for a summary, see Payne, 2011). This thesis focused on dual process interventions due to the greater volume of scholarship devoted to that model in the diagnostic reasoning literature; however, the mental model theory provides fertile ground for future research, and incorporation of this understanding of knowledge structures may contribute substantially to intervention design, in allowing researchers to target mental representations of disease that give rise to errors at many stages of doctors' development (Coderre et al., 2003).

Second, while a clearly conceptualised application of the model to diagnostic reasoning, with relatively detailed components, has been formulated (Croskerry, 2009a) and endorsed by leaders in the field (Balogh et al., 2015), this conceptualisation has yet to gain real traction in the empirical literature on interventions. A more generic version of the model is usually invoked instead, lacking specificity and clear mechanisms that may be targeted. In this way, the full potential of the model to prompt experimental hypotheses and theory-based interventions (Michie et al., 2011) is yet to be realised.

Third, non-analytical or System 1 processes constitute a vital aspect of the dual process model and serve an essential function in the provision of swift, accurate diagnosis; however, these processes are seldom the focus of interventions or commentary, often unfairly maligned as the primary source of diagnostic error or regarded as unconscious, inaccessible, and therefore of little empirical interest (Elstein,

1999). The potential for these processes to be studied as key drivers of diagnostic excellence and possible targets for meaningful intervention, beyond the frame of heuristics and biases, is an important avenue for future research (see Section 3.4.4).

6.4.2 Population and sampling limitations

As with much research in the area of diagnostic reasoning, the samples recruited for the studies in this thesis represent a relatively narrow group of diagnosticians. Doctors for the think-aloud study were drawn from the psychiatry departments of two Dublin-based public hospitals; medical students for the experimental study were drawn from a single cohort at one Dublin-based university.

More broadly speaking, in line with many studies in the area (see Section 3.4.2.4), the range of expertise of the participants was quite limited, ranging from senior medical students to doctors undertaking Higher Specialist Training. It is quite possible that different results would have been obtained in these studies with more experienced doctors; the successful use of non-analytical reasoning is generally achieved with experience, and trainees earlier in their careers are more likely to already use the types of slower, more analytical reasoning methods promoted by reflective interventions (Coderre et al., 2003). This aligns with the findings of the think-aloud study presented in Chapter 4; junior doctors reported that the guided reflection process closely maps on to their natural process for diagnosis. The logistical challenge of accessing more senior and more experienced doctors is an important one for researchers to meet going forward (see Section 6.5).

6.4.3 Methodological limitations

Regarding stimuli, the vignettes-based tasks employed in the experimental study and think-aloud protocol are closely modelled on established pen-and-paper stimuli from the existing literature, and therefore these studies do not make significant strides in overcoming the limitations of these methods. Vignettes are widely used in diagnosis research and there is some evidence to support their value as a tool to measure physician performance (Peabody et al., 2000), and the research team made significant efforts to ensure their face validity through consultation with experts and piloting with participant groups. However, feedback from participants outlined in Section 4.5.5 indicated that, while verisimilitudinous, the vignettes used were, by their nature, unlike real-life cases in important ways: information was presented tidily and in full, without the possibility of asking follow-up questions or of forming a global impression of the patient from the interpersonal interaction, which was particularly pertinent for psychiatric presentations. Vignettes have clear advantages for researchers, being easy and inexpensive to administer and versatile in many settings (Peabody et al., 2000). Still, alternative stimuli for studies may offer significant benefits; these options are discussed in more detail below in Section 6.5.

Regarding outcome measures, in Section 3.4.2.3, it was noted that the vast majority of intervention studies in this area examine outcomes at a level higher than Level 2b (improvement in knowledge or skills) on Kirkpatrick's adapted hierarchy (Barr et al., 2000); that is, outcomes are overwhelmingly measured using clinical scenarios and vignettes, not real-world decision-making in the workplace. Owing to ethical and logistical challenges, resource limitations, and the relatively short period of time available to carry out the work, it was not considered feasible for this programme of research to directly address this. Indeed, even much better-resourced programmes of

research face significant methodological challenges in conducting controlled experiments and detecting experimental effects in hospital settings with real patients, as discussed below. The work presented in this thesis therefore reflects this general trend, and the potential for a transfer of effects to real-world practice, and ultimately to the level of patient outcomes, remains unexplored.

That being said, while efforts to improve patient care are ultimately directed at improving actual outcomes for patients, it is important to consider the appropriateness of measures at the level of patient outcomes for studies in this field. The experimental paradigm, and particularly the randomised control trial that is held up as the ‘gold standard’ for intervention studies, is undoubtedly effective for examining linear, tightly coupled causal relationships and the impact of interventions under ideal or tightly controlled conditions (Pawson & Tilley, 1997). However, utilising outcome measures such as mortality, morbidity, and length of stay in an experimental study of an intervention for diagnostic accuracy would assume that all other sources of variation can be accounted for, but these are neither specific nor sensitive measures of quality of care (O’Dea, 2017). Transferring an intervention out of the lab and onto the ward places the experiment in usual, not ideal, circumstances, and the experimental paradigm is poorly equipped to capture the mechanisms and contextual factors that profoundly impact on outcomes (Berwick, 2008).

In this way, researchers advancing the study of interventions for diagnostic reasoning must anticipate the unique challenges of studying multifactorial interventions in complex environments, and adapt their methods accordingly. In particular, observational methods, rather than experiments intended to be tightly controlled, aimed at examining local quality improvement, wherein broad generalisation takes on relatively less importance, may prove useful tools in understanding whether these

interventions are effective and, crucially, how and why they work (O'Dea, 2017). In that regard, this thesis has advanced the literature, by using a qualitative methodology (think-aloud protocol) to present a clearer picture of how and why the guided reflection intervention is voluntarily implemented by diagnosticians and the specific processes underlying its use. The potential of qualitative methods is further discussed in the next section.

6.5 Future research

Future research may seek to build upon the findings from this research in four broad domains: use of theory, recruitment of diverse populations, inclusion of psychiatry in the diagnostic error discussion, exploration of emerging and qualitative methods, and continued examination of guided reflection interventions specifically.

First, the dual process interventions reviewed in Chapter 3 reflect a common difficulty with intervention design, whereby interventions that are said to be guided by theory or make nominal reference to a theoretical framework are in fact only minimally theoretically informed, if at all; formal analysis of the target behaviour or the purported underlying mechanisms is often absent from the design process (Michie et al., 2011). The dual process model has attained wide acceptance within the diagnostic reasoning literature and some work has already been done to develop a conceptualisation of the process as it applies specifically to diagnosis (Croskerry, 2009b); however, theory arguably remains underutilised in the development of interventions for diagnostic reasoning. Increased use of theory and insights from the implementation and psychological sciences will offer profound benefits to researchers, allowing for more systematic design of interventions and identification of how intervention modalities actively target specific diagnostic mechanisms (e.g., the rational override function of the

dual process model may be targeted through research examining the initiators of the transition between non-analytical and analytical reasoning, informed by existing work on ‘slowing down when you should’ (Moulton et al., 2010b)).

Second, as mentioned elsewhere, studies of more diverse populations of medical trainees and practitioners are required, both in terms of specialism and of experience level. In the literature to date, studies of diagnosis in internal medicine and cardiology have largely dominated (see Chapter 3). As diagnostic practices may differ across specialisms (e.g., psychiatry), researchers endeavouring to understand and improve diagnostic reasoning generally in medicine must take account of this pluralism. As discussed elsewhere in this thesis, studies are dominated by participants who are relatively early in their medical careers, with very experienced doctors being significantly under-represented. The successful use of non-analytical reasoning is generally acknowledged to be a hallmark of expertise and experience, while novices are more likely to rely on more deliberate analytical methods (Coderre et al., 2003). Moreover, there is reason to believe that novice and intermediate diagnosticians also demonstrate different reasoning strategies, and achieving a more nuanced understanding of developmental trends in diagnostic reasoning across the career trajectory represents a potentially fertile area for future research (Ilgen et al., 2011). More senior doctors, largely absent from studies of dual process interventions, may respond quite differently to their less experienced counterparts. If these interventions are to be recommended as educational or workplace tools for trainees and doctors, it will be important to understand how diagnostic reasoning changes across years of experience in order to identify the most appropriate stage at which to introduce these methods to diagnosticians.

Third, the exploration of emerging and qualitative methods as a serious option for research in this area has the potential to yield dividends in illuminating specific psychological processes in diagnostic reasoning (O'Dea, 2017). This benefit may be realised in two ways: through improved stimuli and improved measurement of outcomes.

Regarding stimuli, limitations of ubiquitous vignette-based stimuli may be overcome by high-fidelity patient simulation (e.g., Burbach, Barnason, & Thompson, 2015) or standardised patients (e.g., Maupomé, Schrader, Mannan, Garetto, & Eggertsson, 2010). Basic pen-and-paper vignette studies do not allow for the sort of iterative, purposeful pursuit of information that characterises real-world clinical encounters; information is presented in a static form in its entirety. These new techniques, while highly resource-intensive, are interactive and offer researchers the opportunity to simulate the full complexity of diagnosis in real time, observing not only how doctors make sense of the information they are given but also how that information guides their search for more (Schubert et al., 2013).

Regarding measurement of outcomes, the potential for think-aloud and qualitative methods to offer additional insights to traditional pen-and-paper assessments has been explored in Chapter 4. While it is arguably the case that reasoning processes may only be inferred and not observed directly from think-aloud methods, such data offers a richer insight into moment-to-moment information processing than retrospective reflection or pen-and-paper measures, and remains underutilised in the field of diagnostic reasoning. Observational studies of behaviour have already begun to shed light on the transition from intuitive, non-analytical thinking to deliberate, effortful thinking in surgical settings (Moulton et al., 2010a); think-aloud methods offer the possibility of exploring the same transition in diagnosis with fine-grain, moment-to-

moment verbal data. Eye-tracking studies also offer detailed insight into the visual information being attended to on a moment-to-moment basis and have been used to illuminate reasoning processes in fields such as radiography (Drew, Williams, Aldred, Heilbrun, & Minoshima, 2018).

Notwithstanding the financial and logistical barriers to the use of such methods, they represent exciting opportunities for future research.

Fourth, this thesis presents support for the continued examination of guided reflection as a promising intervention for diagnostic reasoning (for a more up-to-date and comprehensive review of reflective interventions, see Mamede & Schmidt, 2017). Further research is required to clarify its applicability to real-world diagnosis, the ‘active ingredients’ that contribute most to its effects, and how the process may be modified or perhaps simplified to be optimally effective. In particular, clarity is required around the most appropriate audience for the intervention and the circumstances under which it may be most effective. Based on the findings in previous literature and in this thesis, medical trainees very early in their careers may not be well-equipped to take full advantage of the intervention; intermediate diagnosticians, such as registrar-level doctors, may yield more compelling results.

In addition to these domains, as experimental evidence emerges for interventions that are effective in improving diagnostic reasoning, researchers must address the challenge of successfully transferring evidence-based interventions to real-world settings and integrating them into standard care. A significant research-to-practice gap exists; first, relatively few healthcare interventions have been successfully transferred into community organisations, and second, a “voltage drop” (Kilbourne, Neumann, Pincus, Bauer, & Stall, 2007, p2) is frequently observed after transfer, whereby interventions are rendered less effective as fidelity to the intervention is

compromised. An additional challenge lies in ensuring fidelity while allowing for the methods to be adapted to local organisations and integrated into workflows (Kilbourne et al., 2007). This relationship between efficacy in the lab and effectiveness on the ward is complex, and innovative approaches, such as hybrid designs to combine assessment of clinical effectiveness and implementation, are now required to close this research-to-practice gap (Curran, Bauer, Mittman, Pyne, & Stetler, 2012).

6.6 Summary and conclusions

The aim of the research was to examine a range of interventions for medical and psychiatric diagnostic reasoning using the dual process model of reasoning as an organising framework, to examine the extent to which the model has been utilised in intervention design, and to add nuance to our understanding of how interventions impact on diagnostic processes. This aim was achieved through the application of a range of quantitative and qualitative research methods and the findings make significant novel contributions to the literature in this field, bringing a psychological lens to bear on work that has traditionally been the purview of medical researchers. The research underscores the value of theoretical perspectives in the design and interpretation of intervention studies and highlights specific opportunities for expansion of future research, with the guided reflection process showing promise as a tool for implementation in the diagnostic workplace.

References

- Altman, D. G. (1991). *Practical statistics for medical research*. Washington, DC: Chapman & Hall / CRC.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed). Washington, DC: American Psychiatric Association.
- Arzy, S., Brezis, M., Khoury, S., Simon, S. R., & Ben-Hur, T. (2009). Misleading one detail: A preventable mode of diagnostic error? *Journal of Evaluation in Clinical Practice*, 15(5), 804–806. <http://doi.org/10.1111/j.1365-2753.2008.01098.x>
- Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. New York, NY: Cambridge University Press.
- Baker, G. R., Norton, P. G., Flintoft, V., Blais, R., Brown, A., Cox, J., ... Tamblyn, R. (2004). The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal*, 170(11), 1678–1686. <http://doi.org/10.1053/cmaj.1040498>
- Balogh, E. P., Miller, B. T., Ball, J. R., & Press, T. N. A. (2015). *Improving diagnosis in health care*. Washington: Institute of Medicine. <http://doi.org/10.17226/21794>
- Baron, J. (2007). *Thinking and deciding* (4th ed.). Cambridge: Cambridge University Press.
- Barr, H., Freeth, D., Hammick, M., Koppel, I., & Reeves, S. (2000). *Evaluations of interprofessional education: A United Kingdom review of health and social care*. London: Centre for the Advancement of Interprofessional Education (CAIPE).
- Bates, D. W. (2003). Examining the evidence. *Psychiatric Services*, 54(12), 1599–1603.
- Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, 121(5), S2–S23. <http://doi.org/10.1016/j.amjmed.2008.01.001>
- Berwick, D. M. (2008). The science of improvement. *JAMA*, 299(10), 1182–1184.
- Blendon, R. J., DesRoches, C. M., Brodie, M., Benson, J. M., Rosen, A. B., Schneider, E., ... Steffenson, A. E. (2002). Views of practicing physicians and the public on medical errors. *New England Journal of Medicine*, 347(24), 1933–1940. <http://doi.org/10.1056/NEJMs022151>
- Bond, W. F., Deitrick, L. M., Arnold, D. C., Kostenbader, M., Barr, G. C., Kimmel, S. R., & Worrlow, C. C. (2004). Using simulation to instruct emergency medicine residents in cognitive forcing strategies. *Academic Emergency Medicine*, 79, 438–446.
- Bordage, G. (1999). Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Academic Medicine*, 74(10), S138–S143.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. London, England: Sage Publications Ltd.
- Brazil, K., Ozer, E., Cloutier, M. M., Levine, R., & Stryer, D. (2005). From theory to practice: Improving the impact of health services research. *BMC Health Services Research*, 5, 1–5. <http://doi.org/10.1186/1472-6963-5-1>
- Burbach, B., Barnason, S., & Thompson, S. A. (2015). Using “Think Aloud” to capture clinical reasoning during patient simulation. *International Journal of Nursing Education Scholarship*, 12(1). <http://doi.org/10.1515/ijnes-2014-0044>
- Chimowitz, M. I., Logigian, E. L., & Caplan, L. R. (1990). The accuracy of bedside neurological diagnoses. *Annals of Neurology*, 28(1), 78–85.
- Coderre, S., Mandlin, H., Harasaym, P. H., & Fick, G. H. (2003). Diagnostic reasoning

- strategies and diagnostic success. *Medical Education*, 37, 695–703.
- Coderre, S., Wright, H., & McLaughlin, K. (2010). To think is good: querying an initial hypothesis reduces diagnostic error in medical students. *Academic Medicine*, 85, 1125–1129.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, T., Blatter, B., Almeida, C., Shortliffe, E., & Patel, V. (2006). A cognitive blueprint of collaboration in context: Distributed cognition in the psychiatric emergency department. *Artificial Intelligence in Medicine*, 37(2), 73–83. <http://doi.org/10.1016/j.artmed.2006.03.009>
- Cook, M., Noyes, J., & Masakowski, Y. (Eds.). (2007). *Decision making in complex environments*. Burlington, VT: Ashgate Publishing.
- Croskerry, P. (2000a). The cognitive imperative: Thinking about how we think. *Academic Emergency Medicine*, 7(11), 1223–1231.
- Croskerry, P. (2000b). The feedback sanction. *Academic Emergency Medicine*, 7, 1232–1238.
- Croskerry, P. (2002). Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine*, 9(11), 1184–1204.
- Croskerry, P. (2003a). Cognitive forcing strategies in clinical decisionmaking. *Annals of Emergency Medicine*, 41(1), 110–120. <http://doi.org/10.1067/mem.2003.22>
- Croskerry, P. (2003b). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, 78(8), 775–780.
- Croskerry, P. (2009a). A universal model of diagnostic reasoning. *Academic Medicine*, 84(8), 1022–1028.
- Croskerry, P. (2009b). Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in Health Sciences Education*, 14(S1), 27–35. <http://doi.org/10.1007/s10459-009-9182-2>
- Croskerry, P. (2010). To err is human -- and let's not forget it. *Canadian Medical Association Journal*, 182(5), 524. <http://doi.org/10.1503/cmaj.100270>
- Croskerry, P., Abbass, A. A., & Wu, A. W. (2008). How doctors feel: affective issues in patients' safety. *The Lancet*, 372(9645), 1205–1206. [http://doi.org/10.1016/s0140-6736\(08\)61500-7](http://doi.org/10.1016/s0140-6736(08)61500-7)
- Croskerry, P., & Sinclair, D. (2001). Emergency medicine: A practice prone to error? *Canadian Journal of Emergency Medicine*, 3(4), 271–276.
- Croskerry, P., Singhal, G., & Mamede, S. (2013a). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality & Safety*, 22(Suppl 2), ii58–ii64. <http://doi.org/10.1136/bmjqs-2012-001712>
- Croskerry, P., Singhal, G., & Mamede, S. (2013b). Cognitive debiasing 2: impediments to and strategies for change. *BMJ Quality & Safety*, 22(Suppl 2), ii65–ii72. <http://doi.org/10.1136/bmjqs-2012-001713>
- Crumlish, N., & Kelly, B. D. (2009). How psychiatrists think. *Advances in Psychiatric Treatment*, 15(1), 72–79. <http://doi.org/10.1192/apt.bp.107.005298>
- Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-Implementation hybrid design: Combining elements of clinical effectiveness and implementation research to enhance public health. *Med Care*, 50(3), 217–226. <http://doi.org/10.1097/MLR.0b013e3182408812>
- Custers, E. J. F. M. (2013). Medical Education and Cognitive Continuum Theory. *Academic Medicine*, 88(8), 1074–1080. <http://doi.org/10.1097/ACM.0b013e31829a3b10>

- Davis, P., Lay-Yee, R., Briant, R., Ali, W., Scott, A., & Schug, S. (2002). Adverse events in New Zealand public hospitals I: Occurrence and impact. *New Zealand Medical Journal*, 115(1167), U271.
- De Neys, W. (2006a). Automatic-heuristic and executive-analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology*, 59, 1070–1100.
- De Neys, W. (2006b). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, 17, 428–433.
- Drew, T., Williams, L. H., Aldred, B., Heilbrun, M. E., & Minoshima, S. (2018). Quantifying the costs of interruption during diagnostic radiology interpretation using mobile eye-tracking glasses. *Journal of Medical Imaging*, 5(03), 1. <http://doi.org/10.1117/1.JMI.5.3.031406>
- Dror, I. (2011). A novel approach to minimize error in the medical domain: Cognitive neuroscientific insights into training. *Medical Teacher*, 33(1), 34–38. <http://doi.org/10.3109/0142159X.2011.535047>
- Elstein, A. S. (1999). Heuristics and biases: Selected errors in clinical reasoning. *Academic Medicine*, 74(7), 791–794.
- Elstein, A. S. (2009). Thinking about diagnostic thinking: A 30-year perspective. *Advances in Health Sciences Education*, 14(1 SUPPL), 7–18. <http://doi.org/10.1007/s10459-009-9184-0>
- Elstein, A. S., Schulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press.
- Elstein, A. S., & Schwartz, A. (1995). Clinical reasoning in medicine. In J. Higgs (Ed.), *Clinical reasoning in the health professions* (pp. 49–59). Oxford: Butterworth-Heinemann Ltd.
- Ely, J. W., Graber, M. L., & Croskerry, P. (2011). Checklists to reduce diagnostic errors. *Academic Medicine*, 86(3), 307–313. <http://doi.org/10.1097/ACM.0b013e31820824cd>
- Ericsson, K. A., & Simon, H. A. (1999). *Protocol analysis. Verbal reports as data* (2nd ed.). Cambridge, MA: The MIT Press.
- Eva, K. W. (2002). The aging physician: changes in cognitive processing and their impact on medical practice. *AAMC Academic Medicine Journal of the Association of American Medical Colleges*, 77(10 Suppl), S1–S6.
- Eva, K. W., Hatala, R. M., LeBlanc, V. R., & Brooks, L. R. (2007). Teaching from the clinical reasoning literature: Combined reasoning strategies help novice diagnosticians overcome misleading information. *Medical Education*, 41(12), 1152–1158. <http://doi.org/10.1111/j.1365-2923.2007.02923.x>
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59(1), 255–278. <http://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11, 382–389.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241. <http://doi.org/10.1177/1745691612460685>
- Feyzi-Behnagh, R., Azevedo, R., Legowski, E., Reitmeyer, K., Tseytlin, E., & Crowley, R. S. (2014). Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system. *Instructional Science*, 42(2), 159–181. <http://doi.org/10.1007/s11251-013-9275-4>

- Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72(5), 323–327.
<http://doi.org/10.1037/h0028106>
- Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research*, 3(4), 430–441.
<http://doi.org/10.1177/104973239300300403>
- Forster, A. J., Asmis, T. R., Clark, H. D., Al Saied, G., Code, C. C., Caughey, S. C., ... van Walraven, C. (2004). Ottawa Hospital Patient Safety Study: Incidence and timing of adverse events in patients admitted to a Canadian teaching hospital. *Canadian Medical Association Journal*, 170(8), 1235–1240.
<http://doi.org/10.1053/cmaj.1030683>
- Friedman, C. P., Elstein, A. S., Wolf, F. M., Murphy, G. C., Franz, T. M., Heckerling, P. S., ... Abraham, V. (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation: A multisite study of 2 systems. *Journal of the American Medical Association*, 282(19), 1851–1856.
- Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 345–352). The Hague, Netherlands.
- Gigerenzer, G., Czerlinski, J., & Martignon, L. (2002). How good are fast and frugal heuristics? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 559–581). New York, NY: Cambridge University Press.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, W. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53–96.
- Graber, M. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety*, 22, i21–i27. <http://doi.org/10.1136/bmjqs2012-001615>
- Graber, M., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine*, 165, 1493–1499.
- Graber, M., Gordon, R., & Franklin, N. (2002). Reducing diagnostic errors in medicine: What's the goal? *Academic Medicine*, 77(10), 981–992.
- Graber, M., Kissam, S., Payne, V. L., Meyer, A. N. D., Sorensen, A., Lenfestey, N., ... Singh, H. (2012). Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Quality & Safety*, 21(7), 535–557. <http://doi.org/10.1136/bmjqs-2011-000149>
- Graber, M., Sorensen, A. V., Biswas, J., Modi, V., Wackett, A., Johnson, S., ... Singh, H. (2014). Developing checklists to prevent diagnostic error in Emergency Room settings. *Diagnosis*, 1(3). <http://doi.org/10.1515/dx-2014-0019>
- Halligan, M., & Zecevic, A. (2011). Safety culture in healthcare: A review of concepts, dimensions, measures and progress. *BMJ Quality and Safety*, 20(4), 338–343.
<http://doi.org/10.1136/bmjqs.2010.040964>
- Hershberger, P. J., Markert, R. J., Part, H. M., Cohen, S. M., & Finger, W. W. (1997). Understanding and addressing cognitive bias in medical education. *Advances in Health Sciences Education*, 1, 221–226.
- Hershberger, P. J., Part, H. M., Markert, R. J., Cohen, S. M., & Finger, W. W. (1995). Teaching awareness of cognitive bias in medical decision making. *Academic Medicine*, 70(8), 661.
- Hess, B. J., Lipner, R. S., Thompson, V., Holmboe, E. S., & Graber, M. L. (2015).

- Blink or Think: Can Further Reflection Improve Initial Diagnostic Impressions? *Acad Med*, 90. <http://doi.org/10.1097/ACM.0000000000000550>
- Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. The Cochrane Collaboration, 2011.
- Ibiapina, C., Mamede, S., Moura, A., Elói-Santos, S., & van Gog, T. (2014). Effects of free, cued and modelled reflection on medical students' diagnostic competence. *Medical Education*, 48(8), 796–805. <http://doi.org/10.1111/medu.12435>
- Ilgen, J. S., Bowen, J. L., McIntyre, L. A., Banh, K. V., Barnes, D., Coates, W. C., ... Eva, K. W. (2013). Comparing diagnostic performance and the utility of clinical vignette-based assessment under testing conditions designed to encourage either automatic or analytic thought. *Academic Medicine*, 88(10), 1545–1551. <http://doi.org/10.1097/ACM.0b013e3182a31c1e>
- Ilgen, J. S., Bowen, J. L., Yarris, L. M., Fu, R., Lowe, R. A., & Eva, K. (2011). Adjusting our lens: Can developmental differences in diagnostic reasoning be harnessed to improve health professional and trainee assessment? *Academic Emergency Medicine*, 18, S79–S86. <http://doi.org/10.1111/j.1553-2712.2011.01182.x>
- Jain, B. (2016). An investigation into method of diagnosis in clinicopathologic conferences (CPCs). *Diagnosis*, 3(2), 61–64. <http://doi.org/10.1515/dx-2015-0034>
- Johnsen, H. M., Slettebø, Å., & Fossum, M. (2016). Registered nurses' clinical reasoning in home healthcare clinical practice: A think-aloud study with protocol analysis. *Nurse Education Today*, 40, 95–100. <http://doi.org/10.1016/j.nedt.2016.02.023>
- Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: the case of multiple quantification. *Psychological Review*, 96(4), 658–673. <http://doi.org/10.1037//0033-295X.96.4.658>
- Kagee, A. (2002). Concerns about the effectiveness of critical incident stress debriefing in ameliorating stress reactions. *Critical Care*, 6(1), 88. <http://doi.org/10.1186/cc1459>
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss & Giroux.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kassirer, J. P., & Kopelman, R. I. (1989). Cognitive errors in diagnosis: instantiation, classification and consequences. *The American Journal of Medicine*, 86(4), 433–441.
- Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, 4(6), 533–550.
- Kilbourne, A. M., Neumann, M. S., Pincus, H. A., Bauer, M. S., & Stall, R. (2007). Implementing evidence-based interventions in health care: Application of the replicating effective programs framework. *Implementation Science*, 2(1), 1–10. <http://doi.org/10.1186/1748-5908-2-42>
- Kim, H. Y. (2015). *Modeling the reasoning processes in experts and novices' argument diagramming task: Sequential analysis of diagramming behavior and think-aloud data*. Florida State University.
- Kohn, K. T., Corrigan, J. M., & Donaldson, M. S. (1999). *To err is human: Building a safer health system*. Washington, DC: National Academy Press.
- Kostopoulou, O., Sirota, M., Round, T., Samaranayaka, S., & Delaney, B. C. (2017). The role of physicians' first impressions in the diagnosis of possible cancers without alarm symptoms. *Medical Decision Making*, 37, 9–16.
- Kotov, R., Fochtmann, L. J., Carlson, G. A., Tanenberg-Karant, M., & Ruggero, C.

- (2011). Diagnostic shifts during the decade following first admission for psychosis. *American Journal of Psychiatry*, 168, 1186–1194.
- Kovacs, G., & Croskerry, P. (1999). Clinical decision making: An emergency medicine perspective. *Academic Emergency Medicine*, 6(9), 947–952.
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, 118(1), 97–109. <http://doi.org/10.1037/a0020762>
- Kuhn, G. J. (2002). Diagnostic errors. *Academic Emergency Medicine*, 9(7), 740–750.
- Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine*, 13(2), 110–116. http://doi.org/10.1207/s15328015t1302_6
- Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2011). Teaching posttraining: Influencing diagnostic strategy with instructions at test. *Journal of Experimental Psychology: Applied*, 17(3), 195–209. <http://doi.org/10.1037/a0024461>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Leape, L. L., & Berwick, D. M. (2005). Five years after To Err Is Human: What have we learned? *Journal of the American Medical Association*, 293(19), 2384–2390.
- Lewis, G., & Appleby, L. (1988). Personality disorder: The patients psychiatrists dislike. *British Journal of Psychiatry*, 153, 44–49. <http://doi.org/10.1192/pb.bp.115.052456>
- Lilienfeld, S. O., Ammirati, E., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, 4(4), 390–398.
- Lloyd, F. J., & Reyna, V. F. (2009). Clinical gist and medical education: Connecting the dots. *JAMA - Journal of the American Medical Association*, 302(12), 1332–1333. <http://doi.org/10.1001/jama.2009.1383>
- Lundgrén-Laine, H., & Salanterä, S. (2010). Think-aloud technique and protocol analysis in clinical decision-making research. *Qualitative Health Research*, 20(4), 565–575. <http://doi.org/10.1177/1049732309354278>
- Mamede, S., & Schmidt, H. G. (2004). The structure of reflective practice in medicine. *Medical Education*, 38(12), 1302–1308. <http://doi.org/10.1111/j.1365-2929.2004.01917.x>
- Mamede, S., & Schmidt, H. G. (2017). Reflection in medical diagnosis: A literature review. *Health Professions Education*, 3(1), 15–25. <http://doi.org/10.1016/j.hpe.2017.01.003>
- Mamede, S., Schmidt, H. G., & Penaforte, J. C. (2008). Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education*, 42(5), 468–475. <http://doi.org/10.1111/j.1365-2923.2008.03030.x>
- Mamede, S., Schmidt, H. G., Rikers, R. M. J. P., Custers, E. J. F. M., Splinter, T. A. W., & van Saase, J. L. C. M. (2010). Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychological Research*, 74(6), 586–592. <http://doi.org/10.1007/s00426-010-0281-8>
- Mamede, S., Splinter, T. A. W., van Gog, T., Rikers, R. M. J. P., & Schmidt, H. G. (2012). Exploring the role of salient distracting clinical features in the emergence of diagnostic errors and the mechanisms through which reflection counteracts mistakes. *BMJ Quality & Safety*, 21(4), 295–300. <http://doi.org/10.1136/bmjqs-2011-000518>

- Mamede, S., van Gog, T., Moura, A. S., de Faria, R. M. D., Peixoto, J. M., Rikers, R. M. J. P., & Schmidt, H. G. (2012). Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Medical Education*, 46(5), 464–472. <http://doi.org/10.1111/j.1365-2923.2012.04217.x>
- Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M. J. P., van Saase, J. L. C. M., van Guldener, C., & Schmidt, H. G. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Journal of the American Medical Association*, 304(11), 1198–1203.
- Markham, D., & Trower, P. (2003). The effects of the psychiatric label 'borderline personality. *The British Journal of Clinical Psychology*, 42, 243–256.
- Martin, M. S., Hynes, K., Hatcher, S., & Colman, I. (2016). Diagnostic error in correctional mental health: Prevalence, causes, and consequences. *Journal of Correctional Health Care*, 22(2), 109–117. <http://doi.org/10.1177/1078345816634327>
- Maude, J. (2014). Differential diagnosis: The key to reducing diagnosis error, measuring diagnosis and a mechanism to reduce healthcare costs. *Diagnosis*, 1(1), 107–109. <http://doi.org/10.1515/dx-2013-0009>
- Maupomé, G., Schrader, S., Mannan, S., Garetto, L., & Eggertsson, H. (2010). Diagnostic thinking and information used in clinical decision-making: a qualitative study of expert and student dental clinicians. *BMC Oral Health*, 10(1), 11. <http://doi.org/10.1186/1472-6831-10-11>
- McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, 306, 503–507.
- McDonald, K., Matesic, B., Contopoulos-Iannidis, D., Lonhart, J., Schmidt, E., Pineda, N., & Ioannidis, J. P. A. (2013). Patient safety strategies targeted at diagnostic errors: A systematic review. *Annals of Internal Medicine*, 158(5), 381–389.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 276–282. <http://doi.org/10.11613/BM.2012.031>
- Meyer, F., & Meyer, T. D. (2009). The misdiagnosis of bipolar disorder as a psychotic disorder. *Journal of Affective Disorders*, 112(1–3), 174–183.
- Michie, S., & Prestwich, A. (2010). Are interventions theory-based? Development of a theory coding scheme. *Health Psychology*, 29(1), 1–8. <http://doi.org/10.1037/a0016939>
- Michie, S., van Stralen, M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(42).
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, 339(jul21 1), b2535–b2535. <http://doi.org/10.1136/bmj.b2535>
- Monteiro, S. D., Sherbino, J., Patel, A., Mazzetti, I., Norman, G. R., & Howey, E. (2015). Reflecting on diagnostic errors: Taking a second look is not enough. *Journal of General Internal Medicine*, 30(9), 1270–1274. <http://doi.org/10.1007/s11606-015-3369-4>
- Morey, L. C., & Benson, K. T. (2016). An investigation of adherence to diagnostic criteria, revisited: Clinical diagnosis of the DSM-IV/DSM-5 Section II Personality Disorders. *Journal of Personality Disorders*, 30(1), 130–144. http://doi.org/10.1521/pedi_2015_29_188
- Moulton, C. E., Regehr, G., Lingard, L., Merritt, C., & MacRae, H. (2010a). Slowing down to stay out of trouble in the operating room: Remaining attentive in

- automaticity. *Academic Medicine*, 85(10), 1571–1577.
<http://doi.org/10.1097/ACM.0b013e3181f073dd>
- Moulton, C. E., Regehr, G., Lingard, L., Merritt, C., & MacRae, H. (2010b). ‘Slowing down when you should’: Initiators and influences of the transition from the routine to the effortful. *Journal of Gastrointestinal Surgery*, 14, 1019–1026.
- Moulton, C. E., Regehr, G., Mylopoulous, M., & MacRae, H. (2007). Slowing down when you should: A new model of expert judgment. *Academic Medicine*, 82(10), S109–S116.
- Moxley, J. H., Ericsson, K. A., Charness, N., & Krampe, R. T. (2012). The role of intuition and deliberative thinking in experts’ superior tactical decision-making. *Cognition*, 124, 72–78.
- Myung, S. J., Kang, S. H., Phyo, S. R., Shin, J. S., & Park, W. B. (2013). Effect of enhanced analytic reasoning on diagnostic accuracy: A randomized controlled study. *Medical Teacher*, 35(3), 248–250.
<http://doi.org/10.3109/0142159x.2013.759643>
- Nath, S. B., & Marcus, S. C. (2006). Medical errors in psychiatry. *Harvard Review of Psychiatry*, 14(4), 204–211. <http://doi.org/10.1080/10673220600889272>
- Nendaz, M. R., Gut, A. M., Louis-Simonet, M., Perrier, A., & Vu, N. V. (2011). Bringing explicit insight into cognitive psychology features during clinical reasoning seminars: A prospective, controlled study. *Education for Health*, 24(1), 496.
- Norman, G. R., & Eva, K. W. (2003). Doggie diagnosis, diagnostic success and diagnostic reasoning strategies: an alternative view. *Medical Education*, 37, 676–677.
- Norman, G. R., Monteiro, S., & Sherbino, J. (2013). Is clinical cognition binary or continuous? *Academic Medicine*, 88(8).
<http://doi.org/10.1097/ACM.0b013e31829a3c32>
- Norman, G. R., Sherbino, J., Dore, K., Wood, T., Young, M., Gaissmaier, W., ... Monteiro, S. (2014). The etiology of diagnostic errors. *Academic Medicine*, 89(2), 277–284. <http://doi.org/10.1097/acm.0000000000000105>
- O’Cionnaith, F. (2012). €81m bill for incidents in hospitals. *The Irish Examiner*. Cork.
- O’Dea, A. (2017). Should human factors & QI research be held to the same evidentiary standards as clinical research? In *Safety in Health Systems Inaugural Conference*. Dublin, Ireland.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29, 261–265.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11(6), 988–1010. <http://doi.org/10.3758/BF03196730>
- Patel, V. L., Arocha, J. F., Diermeier, M., How, J., & Mottur-Pilson, C. (2001). Cognitive psychological studies of representation and use of clinical practice guidelines. *International Journal of Medical Informatics*, 63(3), 147–167.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, England: Sage Publications Ltd.
- Payne, V. L. (2011). *Effect of a metacognitive intervention on cognitive heuristic use during diagnostic reasoning*. University of Pittsburgh, Pittsburgh, PA.
- Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association*, 283(13), 1715–1722.
- Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. New York,

- NY: Basic Books.
- Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: Correcting for context-induced contrast. *Journal of Experimental Social Psychology*, 29, 137–165.
- Phillips, J. (2014). Detecting diagnostic error in psychiatry. *Diagnosis*, 1(1), 75–78. <http://doi.org/10.1515/dx-2013-0032>
- Pincus, H. A. (2014). Diagnostic errors issues in behavioral health. In *Presentation to the Committee on Diagnostic Error in Health Care*.
- Pinheiro, V. E. D., & Simon, H. A. (1992). An operational model of motor skill diagnosis. *Journal of Teaching in Physical Education*, 11(3), 288–302.
- Pinnock, R., Young, L., Spence, F., & Henning, M. (2015). Can think aloud be used to teach and assess clinical reasoning in graduate medical education? *Journal of Graduate Medical Education*, 7(3), 334–337. <http://doi.org/10.4300/JGME-D-14-00601.1>
- Podbregar, M., Voga, G., Krivec, B., Skale, R., Parežnik, R., & Gabršček, L. (2001). Should we confirm our clinical diagnostic certainty by autopsies? *Intensive Care Medicine*, 27(11), 1750–1755. <http://doi.org/10.1007/s00134-001-1129-x>
- Procyshyn, R. M., Barr, A. M., Brickell, T., & Honer, W. G. (2010). Medication errors in psychiatry: A comprehensive review. *CNS Drugs*, 24(7), 595–609. <http://doi.org/10.2165/11533710-000000000-00000>
- Quirk, M. (2006). *Intuition and metacognition in medical education: Keys to developing expertise*. (C. J. Bland & S. Jones, Eds.), *Springer Series on Medical Education*. New York: Springer Publishing Company.
- Rajkomar, A., & Dhaliwal, G. (2011). Improving diagnostic reasoning to improve patient safety. *The Permanente Journal*, 15(3), 68–73.
- Reason, J. (1990). *Human error*. New York, NY: Cambridge University Press.
- Redelmeier, D. A. (2005). The cognitive psychology of missed diagnoses. *Annals of Internal Medicine*, 142, 115–120.
- Regehr, G., Cline, J., Norman, G., & Brooks, L. R. (1994). Effect of processing strategy on diagnostic skill in dermatology. *Academic Medicine*, 69(10), S34–S36.
- Reilly, J. B., Ogdie, A. R., von Feldt, J. M., & Myers, J. S. (2013). Teaching about how doctors think: A longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ Quality & Safety*, 22, 1044–1050. <http://doi.org/10.1136/bmjqs-2013-001987>
- Round, A. P. (1999). Teaching clinical reasoning: A preliminary controlled study. *Medical Education*, 33, 480–483.
- Saber Tehrani, A. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E. (2013). 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: An analysis from the National Practitioner Data Bank. *BMJ Quality & Safety*, 22(8), 672 LP-680.
- Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 497–502. <http://doi.org/10.1037/0278-7393.28.3.497>
- Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: A systematic review. *BMC Medical Informatics and Decision Making*, 16(1). <http://doi.org/10.1186/s12911-016-0377-1>
- Sari, A. B. A., Sheldon, T. A., Cracknell, A., Turnbull, A., Dobson, Y., Grant, C., ... Richardson, A. (2007). Extent, nature and consequences of adverse events: Results of a retrospective casenote review in a large NHS hospital. *Quality and Safety in*

- Health Care*, 16(6), 434–439. <http://doi.org/10.1136/qshc.2006.021154>
- Schiff, G. D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B. L., ... Krosnjar, N. (2009). Diagnostic error in medicine: Analysis of 583 physician-reported errors. *Advances in Health Sciences Education*, 14(1 SUPPL), 1–5. <http://doi.org/10.1007/s10459-009-9187-x>
- Schmidt, H. G., Mamede, S., van den Berge, K., van Gog, T., van Saase, J. L. C. M., & Rikers, R. M. J. P. (2014). Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. *Academic Medicine*, 89(2), 285–291. <http://doi.org/10.1097/acm.000000000000107>
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, 65(10), 611–621.
- Schnapp, B. H., & Shah, K. (2014). Teaching decision making errors: Lecture-based format is limited. *Academic Emergency Medicine*, 21(5, Suppl. 1), S72-73.
- Schubert, C. C., Denmark, T. K., Crandall, B., Grome, A., & Pappas, J. (2013). Characterizing novice-expert differences in macrocognition: An exploratory study of cognitive work in the emergency department. *Annals of Emergency Medicine*, 61(1), 96–109. <http://doi.org/10.1016/j.annemergmed.2012.08.034>
- Schulz, K. (2011). *Being wrong: Adventures in the margin of error*. New York, NY: Harper-Collins Publishers.
- Schwarz, N. (2005). When thinking feels difficult: Meta-cognitive experiences in judgment and decision making. *Medical Decision Making*, 25(1), 105–112. <http://doi.org/10.1177/0272989x04273144>
- Scott, I. (2009). Errors in clinical reasoning: Causes and remedial strategies. *BMJ*, 339, 22–25.
- Shafir, E., & LeBouef, R. A. (2002). Rationality. *Annual Review of Psychology*, 53, 491–517.
- Shappell, S. A., & Wiegmann, D. A. (2000). *The Human Factors Analysis and Classification System – HFACS*. Washington, DC. <http://doi.org/10.1177/1062860613491623>
- Sherbino, J., Dore, K. L., Siu, E., & Norman, G. R. (2011). The effectiveness of cognitive forcing strategies to decrease diagnostic error: An exploratory study. *Teaching and Learning in Medicine*, 23(1), 78–84. <http://doi.org/10.1080/10401334.2011.536897>
- Sherbino, J., Kulasegaram, K., Howey, E., & Norman, G. R. (2014). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. *Canadian Journal of Emergency Medicine*, 16(1), 34–40. <http://doi.org/10.2310/8000.2013.130860>
- Shimizu, T., Matsumoto, K., & Tokuda, Y. (2013). Effects of the use of differential diagnosis checklist and general de-biasing checklist on diagnostic performance in comparison to intuitive diagnosis. *Medical Teacher*, 35(6), e1218–e1229. <http://doi.org/10.3109/0142159x.2012.742493>
- Shojania, K. G., Burton, E. C., McDonald, K. M., & Goldman, L. (2002). *The autopsy as an outcome and performance measure*. Agency for Healthcare Research and Quality. Rockville, MD.
- Shore, M. F. (2003). Examining the evidence: Contributing factors. *Patient Safety Forum*, 54(12), 1599–1603.
- Sibbald, M., & de Bruin, A. B. H. (2012). Feasibility of self-reflection as a tool to balance clinical reasoning strategies. *Advances in Health Sciences Education*, 17(3), 419–429. <http://doi.org/10.1007/s10459-011-9320-5>

- Sibbald, M., de Bruin, A. B. H., Cavalcanti, R. B., & van Merrienboer, J. J. G. (2013). Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Quality & Safety*, 22, 333–338. <http://doi.org/10.1136/bmjqs-2012-001537>
- Sibbald, M., de Bruin, A. B. H., & van Merrienboer, J. J. G. (2013). Checklists improve experts' diagnostic decisions. *Medical Education*, 47(3), 301–308. <http://doi.org/10.1111/medu.12080>
- Sibbald, M., McKinney, J., Cavalcanti, R. B., Yu, E., Wood, D. A., Nair, P., ... Hatala, R. M. (2012). Cardiac examination and the effect of dual-processing instruction in a cardiopulmonary simulator. *Advances in Health Sciences Education*, 18(3), 497–508. <http://doi.org/10.1007/s10459-012-9388-6>
- Sinclair, D., & Croskerry, P. (2010). Patient safety and diagnostic error: Tips for your next shift. *Canadian Family Physician*, 56, 28–30.
- Singh, H., Giardina, T. D., Meyer, A. N. D., Forjuoh, S. N., Reis, M. D., & Thomas, E. J. (2013). Types and origins of diagnostic errors in primary care settings. *JAMA Internal Medicine*, 173(6), 418–425. <http://doi.org/10.1001/jamainternmed.2013.2777>
- Singh, H., Graber, M. L., Kissam, S. M., Sorensen, A. V., Lenfestey, N. F., Tant, E. M., ... LaBresh, K. A. (2011). System-related interventions to reduce diagnostic errors: A narrative review. *BMJ Quality & Safety*, 21(2), 160–170. <http://doi.org/10.1136/bmjqs-2011-000150>
- Skånér, Y., Backlund, L., Montgomery, H., Bring, J., & Strender, L. E. (2005). General practitioners' reasoning when considering the diagnosis heart failure: A think-aloud study. *BMC Family Practice*, 6, 1–10. <http://doi.org/10.1186/1471-2296-6-4>
- Sonderegger-Iseli, K., Burger, S., Muntwyler, J., & Salomon, F. (2000). Diagnostic errors in three medical eras: a necropsy study. *The Lancet*, 355(9220), 2027–2031. [http://doi.org/10.1016/S0140-6736\(00\)02349-7](http://doi.org/10.1016/S0140-6736(00)02349-7)
- Soop, M., Fryksmark, U., Koster, M., & Haglund, B. (2009). The incidence of adverse events in Swedish hospitals: A retrospective medical record review study. *International Journal for Quality in Health Care*, 21(4), 285–291.
- Stanovich, K. E., & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking & Reasoning*, 4, 193–230.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences*, 23, 645–726.
- Stapel, D. A., Martin, L. L., & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality and Social Psychology Bulletin*, 24(8), 797–806.
- Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1), S164–S187. <http://doi.org/10.1086/380085>
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *The American Economic Review*, 93(2), 175–179.
- Thornicroft, G., Rose, D., & Kassam, A. (2007). Discrimination in health care against people with mental illness. *International Review of Psychiatry*, 19(2), 113–122. <http://doi.org/10.1080/09540260701278937>
- Treadwell, J. R., Lucas, S., & Tsou, A. Y. (2014). Surgical checklists: A systematic review of impacts and implementation. *BMJ Quality and Safety*, 23(4), 299–318. <http://doi.org/10.1136/bmjqs-2012-001797>
- Trowbridge, R. L. (2008). Twelve tips for teaching avoidance of diagnostic errors.

- Medical Teacher*, 30(5), 496–500. <http://doi.org/10.1080/01421590801965137>
- Trowbridge, R. L., Dhaliwal, G., & Cosby, K. S. (2013). Educational agenda for diagnostic error reduction. *BMJ Quality & Safety*, 22(Suppl 2), ii28–ii32. <http://doi.org/10.1136/bmjqs-2012-001622>
- Tsujii, T., & Watanabe, S. (2009). Neural correlates of dual-task effect on belief-bias syllogistic reasoning: A near-infrared spectroscopy study. *Brain Research*, 1287, 118–125.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Vincent, C., Neale, G., & Woloshynowych, M. (2001). Adverse events in British hospitals: Preliminary retrospective record review. *BMJ*, 322, 517–519.
- Waddington, L., & Morley, S. (2000). Availability bias in clinical formulation: The first idea that comes to mind. *British Journal of Medical Psychology*, 73, 117–127.
- Wears, R. L., & Nemeth, C. P. (2007). Replacing hindsight with insight: Toward better understanding of diagnostic failures. *Annals of Emergency Medicine*, 49(2), 206–209. <http://doi.org/10.1016/j.annemergmed.2006.08.027>
- Wiggins, M., Stevens, C., Howard, A., Henley, I., & O'Hare, D. (2002). Expert, intermediate and novice performance during simulated pre-flight decision-making. *Australian Journal of Psychology*, 54(3), 162–167. <http://doi.org/10.1080/00049530412331312744>
- Wilson, R. M., Runciman, W. B., Gibberd, R. W., Harrison, B. T., Newby, L., & Hamilton, J. D. (1995). The Quality in Australian Health Care Study. *The Medical Journal of Australia*, 163, 458–471.
- Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 185–200). New York, NY: Cambridge University Press.
- Wolpaw, T., Papp, K. K., & Bordage, G. (2009). Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: A randomized comparison group trial. *Academic Medicine*, 84(4), 517–524.
- Yoder, P., & Simmons, F. (2010). *Observational measurement of behavior*. New York, NY: Springer Publishing Company.

Appendix A: Additional search strategies (Systematic review, Chapter 3)

Table A.1

Search strategy for PsycInfo (Ovid, 1990-present)

Diagnosis / Decision-making	Cognition	Population	Studies	Not
judgment/	cognition/	exp medical education/	exp experimental design/	clinical trial/
OR medical diagnosis/	OR intuition/	OR physicians/	OR quantitative methods/	
OR “clinical judgment (not diagnosis)”/	OR metacognition/	OR medical personnel/	OR experiment\$.mp	
OR misdiagnosis/	OR heuristics/	OR medical students/	OR control\$.mp	
OR diagnostic reasoning.mp	OR dual process models/	OR teaching/	OR trial\$.mp	
OR clinical reasoning.mp	OR cognitive bias\$.mp	OR physician\$.mp	OR intervention\$.mp	
OR diagnostic accuracy.mp	OR cognitive error\$.mp	OR doctor\$.mp	OR effect.mp	
OR diagnostic error\$.mp	OR (cognitive adj2 strateg\$.mp	OR clinician\$.mp	OR effects.mp	
OR medical error\$.mp	OR (reasoning adj2 strateg\$.mp	OR medical student\$.mp	OR effectiveness.mo	
OR quality of care/	OR metacogniti\$.mp	OR consultant\$.mp	OR efficacy.mp	
OR decision making/	OR reasoning/	OR intern.mp	OR exp empirical methods/	
OR problem solving/	OR analytical adj2 reasoning.mp	OR interns.mp		
OR diagnosis/	OR non-analytical adj2 reasoning.mp	OR resident.mp		
OR reflective practice.mp	OR intuiti\$.mp	OR residents.mp		
		OR medical training.mp		
		OR curriculum.mp		

Note:

/ indicates Subject Heading

.mp searches for term in Title, Original Title, Abstract, Subject Heading, Name of Substance, and Registry Word fields

\$ indicates truncation

adj searches for both terms next to each other in the order presented

adj 2 searches for both terms within two words of each other, in any order

exp indicates exploded search

Table A.2

Search strategy for Embase (Elsevier, 1990-present)

Diagnosis / Decision-making	Cognition	Population	Studies	Not
clinical competence'/de	cognition'/de	medical education'/exp	controlled study'/de	cognitive near/1 deficit
OR 'medical error'/exp	OR 'intuition'/de	OR 'medical personnel'/exp	OR 'experimental study'/de	OR cognitive near/1 decline
OR 'diagnosis'/exp	OR analytical next/2 reasoning	OR medical near/2 training	OR 'comparative study'/de	OR cognitive near/1 impairment
OR 'medical decision making'/de	OR non- analytical next/2 reasoning	OR curriculum	OR intervention*	
OR 'clinical decision making'/de	OR metacogniti*	OR medical next/2 student*	OR experiment*	
OR 'diagnostic reasoning'/de	OR cognitive next/2 bias*	OR physician*	OR effect?	
OR 'problem solving'/de	OR cognitive next/2 error*	OR doctor*	OR effectiveness	
OR “reflective practice”	OR cognitive next/2 strateg*	OR clinician*	OR efficacy	
	OR reasoning near/2 strateg*	OR consultant*		
		OR intern		
		OR interns		
		OR resident		
		OR residents		

Note:

/ indicates preferred term (Subject Heading)

\$ indicates truncation

? indicates wildcard

near/2 searches for both words within two words of each other, in any order

next/2 searches for both terms within two words of each other, in the order presented

exp indicates exploded search

Table A.3

Search strategy for Education Resource Information Centre (ERIC) (ProQuest, 1990-present)

Diagnosis / Decision-making	Cognition	Population	Studies
“medical error”	“analytical reasoning”	physician	intervention
OR “medical errors”	OR “non-analytical reasoning”	OR physicians	OR interventions
OR “diagnostic error”	OR metacognition	OR (doctor OR doctors	OR effect
OR “diagnostic errors”	OR metacognitive	OR (clinician OR clinicians)	OR effects
OR “diagnostic reasoning”	OR “cognitive bias”	OR (“medical student” OR “medical students”)	OR effectiveness
OR “reflective practice”	OR “cognitive biases”	OR (“consultant” OR “consultants”)	OR efficacy
OR diagnosis	OR “cognitive strategy”	OR (“intern” OR “interns”)	OR experiment
OR su.Exact(“Clinical Diagnosis”)	OR “cognitive strategies”	OR (“resident” OR “residents”)	OR experiments
OR su.Exact(“Error Patterns”)	OR su.Exact(“Intuition”)	OR curriculum	OR trial
	OR su.Exact(“Metacognition”)	OR (su.Exact(“Medical students”)	OR control
	OR (su.Exact(“Problem Solving”)	OR (su.Exact(“Medical education”)	OR controlled
	OR su.Exact(“Thinking Skills”)	OR (su.Exact(“Expertise”)	OR (su.Exact(“Experiments”)
		OR (su.Exact(“Teaching Methods”)	OR (su.Exact(“Investigations”)
		OR (su.Exact(“Medical Schools”)	OR (su.Exact(“Evidence”)
		OR (su.Exact(“Clinical Teaching (Health Professions”)	
		OR (su.Exact(“Science Education”)	
		OR (su.Exact(“Physicians”)	

Note:

su.Exact indicates Subject Heading

Table A.4

Search strategy for the Cochrane Database of Controlled Trials (Wiley, 1990-present)

Diagnosis / Decision-making	Cognition	Population	Studies
physician's practice patterns/ OR decision support systems, clinical/ OR decision support techniques/ OR judgment/ OR diagnosis/ OR clinical competence/ OR exp diagnostic errors/ OR exp medical errors/ OR problem solving/ OR diagnostic reasoning.mp OR clinical reasoning.mp OR reflective practice.mp OR reflective reasoning.mp OR diagnostic accuracy.mp OR cognitive adj2 strateg\$.mp OR reasoning adj strateg\$.mp. OR metacogniti\$.mp OR diagnostic error\$.mp OR medical error\$.mp	cognition/ OR intuition/ OR thinking/ OR cogniti\$.mp OR cognitive bias\$.mp OR cognitive error\$.mp	exp students, medical/ OR exp medical staff/ OR internship and residency/ OR educational measurement/ OR exp education, medical/ OR teaching/ OR physician\$.mp OR doctor\$.mp OR clinician\$.mp OR medical student\$.mp OR consultant\$.mp OR intern.mp OR interns.mp OR resident.mp OR residents.mp	evaluation studies/ OR longitudinal studies/ OR follow-up studies/ OR prospective studies/ OR cross-sectional studies/ OR comparative studies/ OR experiment\$.mp OR control\$.mp OR trial\$.mp OR intervention\$.mp OR effect.mp OR effects.mp OR effectiveness.mp OR efficacy.mp

Note:

/ indicates MeSH descriptor

.mp searches for term in Title, Original Title, Abstract, Subject Heading, Name of Substance, and Registry Word fields

\$ indicates truncation

adj searches for both terms next to each other in the order presented

adj 2 searches for both terms within two words of each other, in any order

exp indicates exploded search

Appendix B: Coding frameworks: Use of theory in intervention design (Systematic review, Chapter 3)

Table B.1

Coding framework for elements of dual process model mentioned in included papers

Elements	Language coded
Analytical reasoning	Executive, deliberate, counterfactual, hypothetical, statistical, rational
Biases / heuristics	Bias, heuristic
Dual processes	Dual process model, dual process theory
Metacognition	Metacognition, monitoring
Non-analytical reasoning	Intuition, automatic, instinct, gut, pattern recognition
Systems 1 and 2	System 1, System 2

Table B.2

Coding framework for mechanisms / processes of dual process models targeted by interventions

Mechanism / Process	Definition	Example
Cognitive load	Cognitive load, information overload or working memory	“Checklists can combat the information overload involved in system 2 processing by limiting conscious attention to a small number of relevant variables.”
Interrogation	Deliberate attention drawn to a specific aspect of reasoning, going beyond the more general metacognitive monitoring function	“Pause to reflect ... Did I consider the inherent flaws of heuristic thinking?”
Metacognition	General metacognition or real-time awareness of cognitive processes	“This intervention was designed to induce the subject to “think about how they think””; To help participants learn to recognize their own errors by watching themselves perform the task while reflecting on it and then watching the correct solution path for the case.”
Switching	Switching from one mode of reasoning to another or deliberate slowing down	“Residents were given ‘balancing’ reasoning instructions.”
System 2 oversight / monitoring	The use of System 2 reasoning to monitor or override System 1 reasoning, to repair errors in reasoning during the diagnostic process or to verify a diagnostic hypothesis	“When this initial diagnosis is wrong, it can only be “repaired” by further conscious analysis of the other, less salient features of the case, leading to the activation of alternative hypotheses and, eventually, recognition of the correct diagnosis.”
System 2 reasoning	System 2 processes, including hypothesis generation, evidence gathering, statistical reasoning, counterfactual reasoning, hypothetical reasoning, reflection	“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 the diagnosis were true but were not described in the case.”

Appendix C: Ethical approval (Think-aloud study, Chapter 4)



Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

F.A.O. Kathryn Lambe

School of Psychology Research Ethics Committee

19th December 2017

Dear Kathryn,

The School of Psychology Research Ethics Committee has reviewed your application entitled "Reflection in diagnosis – think-aloud protocol" and I am pleased to inform you that it was approved.

Please note that you will be required to submit a completed **Project Annual Report Form** on each anniversary of this approval, until such time as the research is complete and the thesis is submitted. The form is available for download from the Ethics section of the School website.

Adverse events associated with the conduct of this research must be reported immediately to the Chair of the Ethics Committee.

Please ensure that you keep this letter safe for inclusion with final project/dissertation/thesis.

Yours sincerely,

Dr Clare Kelly
Acting Chair
School of Psychology Research Ethics Committee

SCHOOL OF PSYCHOLOGY
Arás an Phiarsaigh
Trinity College
Dublin 2

Scoil na Síceolaíochta

Dámh na nEolaíochtaí Sóisialta agus Daonna,
Áras an Phiarsaigh, Coláiste na Tríonóide,
Baile Átha Cliath 2, Éire.

School of Psychology

Faculty of Arts, Humanities and Social Sciences,
Áras an Phiarsaigh, Trinity College,
Dublin 2, Ireland.

T +353 1 896 1886
F +353 1 671 2006
psychology@tcd.ie
www.tcd.ie/psychology

Appendix D: Recruitment message (Think-aloud study, Chapter 4)

Title of study: Voluntary use of guided reflection for diagnostic reasoning: A think-aloud protocol study

To whom it concerns,

I am a postgraduate researcher from the School of Psychology at Trinity College Dublin, working with the Discipline of Psychiatry and the School of Medicine, and I am conducting a study that explores the way in which doctors make decisions. I would like to invite you to take part in my research.

Participation involves reading some medical scenarios and offering your opinions aloud on the situations they describe.

It is hoped that the results of the study will help to advance our understanding of the factors that contribute to medical decision-making and to design useful educational materials to improve decision-making.

The study will take under one hour to complete, at a time and place convenient to you. Participation is completely voluntary and so you are not obliged to take part.

If you have any questions about the research, feel free to contact me at lambeka@tcd.ie.

Kind regards,
Kathryn Lambe

Appendix E: Information sheet (Think-aloud study, Chapter 4)

Title of study: Voluntary use of guided reflection for diagnostic reasoning: A think-aloud protocol study

Introduction:

Existing research suggests that the majority of medical errors arise from errors of thinking, rather than errors of technique, and that it may be possible to reduce such error through cognitive strategies. This study aims to examine the choices doctors make to use or not use these strategies, and to identify how these strategies can help or hinder doctors in their diagnostic decisions.

Your participation should take no longer than one hour. Your participation is entirely voluntary, and your participation or non-participation will have no implications for your employment or future job opportunities.

Procedures:

You are eligible to take part in this study if you are a practicing medical doctor at registrar grade or above, with the necessary fluency in English.

If you decide to take part, you will be asked to read a series of fictional patient cases and to make diagnoses based on the information contained therein. You will be asked to speak your thoughts aloud as you consider the cases, and then we will ask you to briefly reflect on the experience. The session will be audio-recorded and transcribed for analysis. You will be given the opportunity to review the transcript once it is complete.

Benefits:

The study offers no direct benefits to participants. However, it is hoped that the findings will help to advance our understanding of how doctors make decisions and how we may further educational efforts to improve medical decision-making. The implications of this research may help educators to improve curricula and teaching methods, thereby contributing to the quality of care experienced by patients in the future.

The study may also offer you the opportunity to reflect on your own decision-making processes, which participants in similar studies often report as being an interesting experience that benefits their practice.

Risks:

It is not anticipated that you will be exposed to any physical, psychological or social risks through this study that you would not encounter in everyday life. As the study asks you to make diagnoses based on some fictional patient cases, you might find this portion of the study to be exam-like or stressful. At the end of the study, you will be provided with the details of some agencies that may offer support if you feel upset or

distressed. Again, your participation is entirely voluntary, and your participation or non-participation will have no implications for your employment or future job opportunities.

Exclusion from participation:

No exclusion criteria apply.

Confidentiality:

Your identity will remain confidential. Your name will not be published and will not be disclosed to anyone outside the research team. Your information will not be linked to your name, will be stored securely, and will be accessible only to the research team and the Principal Investigator's PhD examination board. Your information will be used only for the purposes of this study as outlined in this leaflet; it will not be used in future unrelated studies without your express permission.

Compensation:

This study is covered by standard institutional indemnity insurance. Nothing in this document restricts or curtails your rights.

Voluntary Participation:

If you decide to volunteer to participate in this study, you may withdraw at any time. If you decide not to participate, or if you withdraw, you will not be penalised and will not give up any benefits that you had before entering the study. Your participation or non-participation will have no implications for your studies, grades or future job opportunities.

On completion of the interview, you may withdraw your information at any time during the four weeks following the day of the interview session by contacting the research team; after this time, it will not be possible to withdraw your information.

Stopping the study:

You understand that the investigators may withdraw your participation in the study at any time without your consent.

Permission:

The study has received approval from the Trinity College Dublin School of Medicine Research Ethics Committee.

Further information:

You can get more information or answers to your questions about the study, your participation in the study, and your rights from any of the research team listed below.

If your Principal Investigator learns of important new information that might affect your desire to remain in the study, you will be informed of this.

Kathryn Lambe
PhD Candidate, TCD School of Psychology
Email lambeka@tcd.ie

Prof Brendan Kelly
TCD School of Medicine
Email brendan.kelly@tcd.ie
Tel 01 896 3799

Dr David Hevey
TCD School of Psychology
Email heveydt@tcd.ie
Tel 01 896 2406

Appendix F: Consent form (Think-aloud study, Chapter 4)

Title of research study: Voluntary use of guided reflection for diagnostic reasoning: A think-aloud protocol study

This study and this consent form have been explained to me.

My Principal Investigator has answered all my questions to my satisfaction.

I believe that I understand what will happen if I agree to be part of this study.

I have read, or had read to me, this consent form.

I have had the opportunity to ask questions and all my questions have been answered to my satisfaction.

I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.

I understand that the session will be audio-recorded and transcribed for analysis, and that I will be given the opportunity to review the transcript once it is complete.

I understand that I may withdraw my information at any time during the four weeks following the day of the interview session by contacting the research team and that it will not be possible to withdraw my information after this time.

PARTICIPANT'S NAME:

PARTICIPANT'S SIGNATURE:

Date:

Date on which the participant was first furnished with this form:

Statement of investigator's responsibility: I have explained the nature, purpose, procedures, benefits, risks of, or alternatives to, this research study. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

Principal Investigator's signature:

Date:

Appendix G: Debriefing sheet (Think-aloud study, Chapter 4)

Title of study: Voluntary use of guided reflection for diagnostic reasoning: A think-aloud protocol study

Thank you very much for taking part in this research.

Existing research suggests that the majority of medical errors arise from errors of thinking, rather than errors of technique, and that it may be possible to reduce such error through acknowledgement of cognitive biases and deliberate implementation of cognitive strategies by individual doctors. This study aims to identify how these strategies can help doctors in their diagnostic decisions.

You will be assigned a participant code, which will be used to mark the information you provide to us. Your information will not be linked to your name and will be stored securely. Your information will be accessible only to the research team and the Principal Investigator's PhD examination board. Your information will be used only for the purposes of this study as outlined in this leaflet; it will not be used in future unrelated studies without your express permission.

If you have found participating in this study stressful or feel in any way upset or distressed following participation, we recommend making contact with your GP, the Occupational Health department of your base hospital, or the Samaritans (www.samaritans.org Tel 116 123).

If you have any further questions about the study, please feel free to contact the research team.

Kathryn Lambe

PhD Candidate, TCD School of Psychology

Email lambeka@tcd.ie

Prof Brendan Kelly

TCD School of Medicine

Email brendan.kelly@tcd.ie

Tel 01 896 3799

Dr David Hevey
TCD School of Psychology
Email heveydt@tcd.ie
Tel 01 896 2406

Appendix H: Vignettes (Think-aloud study, Chapter 4)

Patient A

Chief Complaint: This 21 year old male presented to Accident and Emergency Department with his mother, who describes him as acting bizarrely for the last two weeks, with new onset inability to communicate verbally today.

History of Presenting Complaint: His mother describes a history of unusual behaviour beginning two weeks prior to this presentation. Normally a reserved and generally uncommunicative young man, he developed signs of increasing animation: initiating conversations with strangers and being noticeably more talkative for a period of about ten days. During this time he seemed to show evidence of elevated mood, punctuated by outbursts of anger. A business student in UCD, he neglected to attend lectures and didn't feel he needed to study for upcoming exams. Instead he spent his time working on business ideas, excitedly telling everyone that he believed he had invented a number of patents that were worth millions. His behaviour has changed over the past three days, becoming increasingly suspicious and bizarre. He is worried that he is being monitored by a specific company, and he believes that they want to steal his ideas. He is suspicious of anyone wearing the colour yellow, paranoid that they are spies sent by this company. His mother took him to see a GP today, whom she says wanted her son to go to A/E to be assessed by psychiatry for possible admission. Unfortunately, the GP was wearing a yellow tie, so the patient refused to go to A/E, and told the GP that she was 'a voodoo witch doctor in the pocket of that company'. The GP prescribed Haloperidol 5mg OD and 5mg PRN. His mother says that she convinced him to take 10mg Haloperidol, and it seemed to have a calming effect, although he did communicate a worry that he was talking too much, and that he felt he needed to shut-up before all his ideas were stolen. Over the nine hours that have passed since taking the medication, he has developed signs of increasing distress. He is now unable or unwilling to speak. He can communicate by writing, and tells that he can't speak because his neck is very painful. He also tells of a difficulty breathing, and he feels that his inhaler isn't working properly. He denies any perceptual abnormality.

Past Medical History: Asthma

Medication: Salbutamol inhaler PRN; No known drug allergies

Family History: Father Deceased age 55, Myocardial Infarction; Paternal Uncle Schizophrenia

Personal History: Normal delivery; normal milestones. Only child.

Social History: Lives with his mother. Third year business student in UCD. His mother says that he rarely drinks alcohol, and is adamant that he is unlikely to have ever tried illicit drugs. He plays football for his local club.

Physical Examination: revealed a young, adult white male who appeared fidgety and worried looking. He was unable to communicate verbally during the initial assessment, except to utter 'aahh' and 'naahh' which seemed to stand for yes and no respectively. He did not appear confused, at times nodding assent in agreement with his mother's telling of the history. His pulse was 110. His blood pressure 115/80. His temperature was 36.9°. His respiratory rate was 20, and his spO2 96% on room air. He was diaphoretic, and his face was flushed. His mouth remained open, at times sticking his tongue out as if attempting to articulate correctly, and he had profuse sialorrhoea. He showed evidence of mild dyspnoea. His shoulders and his neck were held in extension. Auscultation of the chest revealed normal air entry and normal vesicular breath sounds were heard bilaterally.

Laboratory Information

HB	13.8	g/dl	(13.0- 18.5)
PLT	158	x10 ⁹ /L	(150 - 450)
WCC	7.2	x10 ⁹ /L	(4.0 - 11.0)
Neut	42%	3.0 x10 ⁹ /L	(2.0 - 7.5)
Sodium	142	mmol/L	(135 - 145)
Potassium	3.6	mmol/L	(3.5 - 5.0)
Creatinine	87	umol/L	(59 - 104)
Urea	3.6	mmol/L	(2.0 - 7.0)
Albumin	38	g/L	(35 - 50)
Total Bilirubin	9	umol/L	(<17)
ALT	33	IU/L	(<45)
Alk. Phos.	45	IU/L	(40 - 130)
Gamma GT	31	IU/L	(<60)
Creatinine Kinase	740	IU/L	(40 – 320)

UR. TOXICOLOGY SCREEN

Paracetamol (Urine)	POSITIVE
Amphetamines (Urine)	Negative
Metamphetamines (Urine)	Negative
Barbiturates (Urine)	Negative
Benzodiazepines (Urine)	Negative
Cocaine (urine)	Negative
Methadone (Urine)	Negative
Opiates (Urine)	Negative
Phencyclidine (Urine)	Negative
THC (Urine)	Negative

Chest X-Ray: normal heart and lungs

Patient B

Chief Complaint: This 28 year old primary school teacher, in long term relationship, presented to the Emergency Department with low mood and suicidal ideation.

History of Presenting Complaint: She describes a very low mood for the past three weeks, down for the entirety of the day with little variation. It is associated with feelings of fatigue and generalised weakness. She has felt unable to go to work for the past week, citing poor concentration and poor energy levels. She has begun feeling increasingly hopeless about the future and has experienced fleeting suicidal ideation over the past two days. She denies any suicidal intent or plans. She has poor sleep with early morning wakening. She says her appetite is markedly diminished. She also describes a history of intermittent nausea, abdominal pain and constipation over the six months. The abdominal pain is usually nonspecific, and can disappear for weeks at a time. The only other symptom of which she complains is an increase in urinary frequency. She denies any current stressors or worries, apart from the death of her grandmother four months ago, to whom she had been very close.

Past Psychiatric History: She hasn't had any contact with psychiatric services in the past. Today was the first time she ever went to her GP with low mood. The GP advised presenting to A/E.

Past Medical History: There is nothing of significance in her past medical history.

Medication: Yasmin (Ethinylestradiol with Drospirenone). Ibuprofen PRN. She has no known drug allergies.

Family History: Her family history is non-contributory.

Social History: She lives with her boyfriend of three years. She works as a primary school teacher. She drinks a glass of wine every evening.

Premorbid and Collateral History: Her boyfriend describes her premorbid personality as generally happy and sociable. 'She isn't the sort of person to mope or stay down. She's practical.' He says that she was sad after the death of her grandmother, and spent a lot of time at home with her family, but had gotten back to normal after the first month. He confirms the change in mood over the past three weeks and that she is no longer eating properly. He says she drinks a lot of Lucozade and 7up.

Physical Examination: Her vital signs are all within normal range. Her abdomen is soft and non-tender. On mental state examination, she presents as

overweight with signs of poor self-care: her hair unwashed, her clothes grubby, and not wearing make-up. Her eye contact is poor. She is calm. Her speech is monotone, but otherwise normal. Objectively her mood is depressed and her affect is flat. She has a passive death wish and fleeting suicidal ideation. There were no psychotic symptoms elicited.

Laboratory Information

HB	13.6	g/dl	(12.0- 15.5)
PLT	170	x10 ⁹ /L	(150 - 450)
WCC	6.2	x10 ⁹ /L	(4.0 - 11.0)
Sodium	142	mmol/L	(135 - 145)
Potassium	3.6	mmol/L	(3.5 - 5.0)
Creatinine	90	umol/L	(59 - 104)
Urea	3.5	mmol/L	(2.0 - 7.0)
Albumin	25	g/L	(35 - 50)
Calcium	2.55	mmol/L	(2.10-2.55)
Phosphate	0.75	mmol/L	(0.74-1.52 mmol/L)
ALT	33	IU/L	(<45)
Alk. Phos.	45	IU/L	(40 - 130)
Gamma GT	31	IU/L	(<60)
Fasting Glucose	4.5	mmol/L	(3.8-6.0 mmol/L)

Urine dipstick: Positive for blood and protein

Patient C

Chief Complaint: This is a 68 year old male, retired paramedic, referred by GP to Psychiatric services with low mood and memory difficulties.

History of Presenting Complaint: He describes a deterioration in mood over the last two to three months, generally down at the beginning of the week and then returning to normal by the weekend. His energy levels are unchanged. His sleep and appetite are unaffected. He denies anhedonia, saying that he enjoys spending time with his children and grandchildren. He no longer engages in many of his usual activities, such as bowling, and walking to the pub for two pints every evening. He blames a difficulty in walking as the reason for these changes in habit.

The difficulties walking began one year ago, and were first noticed as a general slowing – hesitancy in initiating steps and changing direction and a tendency to fall over – especially after drinking. He began using a stick and avoiding all but the shortest distances on foot. He has noticed that it takes longer to complete simple household tasks, like washing, cooking and dressing.

One of the biggest problems, he says, is getting to the toilet on time. He experiences urinary incontinence on an almost daily basis, and says that a lot of his day revolves around washing bedsheets and clothes. In addition to these problems, he has noticed problems with his memory over the past half a year. He has difficulty in finding the right words; he forgets recent events; and it has been pointed out to him that he often repeats himself.

Past Psychiatric History: He has not had contact with psychiatric services prior to this assessment.

Past Medical History: Hypertension. Hypercholesterolemia.

Medication: Amlodipine. Rosuvastatin. He has no known drug allergies.

Family History: Family history is non-contributory.

Social History: His wife died two years ago. He lives alone. He used to drink two pints of Guinness every night in the pub, but stopped drinking five months ago. He has never smoked.

Collateral: His daughter is very worried about him. She says that he has changed drastically over the last six months. She says that he doesn't seem to be able to look after himself properly anymore. 'It's like he has given up caring.' She says that he used to be proud of the tidiness and orderliness of his house, but now he doesn't bother looking after it anymore. She says he isn't eating properly. 'He's well able to cook, but all he lives on now is bread and cheese and tea.' It is his apathetic attitude that she finds disturbing. She says that his memory seemed to deteriorate when he gave up drinking five months ago.

Physical Examination: Vital signs were within normal limits, apart from blood pressure, which measured 145/85. He scored 24/30 on MMSE, losing points on recall and attention and calculation. Frontal lobe tests were normal. He was found to have normal tone and power in both upper and lower limbs. Deep tendon reflexes were present. He was found to have a magnetic gait – his feet not rising much off the ground – with steps that were short and shuffling, and a wide based stance. There was no evidence of nystagmus or dysdiadokokinesis. Cranial nerves 2 -12 were intact. On mental state examination, he presented as calm and pleasant, with evidence of poor self-care. His clothes smelled of stale urine. His behaviour was appropriate. He maintained eye-contact. The rate, rhythm and volume of his speech were normal. His mood was euthymic and affect reactive. He denied suicidal ideation. He did not display abnormalities in thought form or communicate abnormalities in thought content. He denied perceptual abnormality in any of the sensory modalities.

Patient D (pilot study only)

Chief Complaint: This 34 year old male presented to Accident and Emergency Department with his wife, after communicating to her the belief that he controls the weather.

History of Presenting Complaint: For the past year, he says that he has ‘been keeping the world in balance, so that night follows day and summer follows spring.’ He says that his main focus is on the balancing of ‘energy forces’ through the control of the weather. He believes that he decides whether it will be sunny or if it will rain. He also believes that he is the cause of natural disasters in the world, and cites last month’s Hurricane Sarah and the earthquake in Central America as recent products of his work. He says that ‘natural disasters are sometimes the only way of offloading gigantic amounts of energy.’ This control of weather and overall ‘balancing of the world’ is achieved through ‘meditation and intense visualisation.’

His beliefs are unshakeable by argument. He denies any perceptual abnormality and says that the ideas began to form in his head after he noticed he could control clouds one day last year. His understanding of his ‘purpose and powers’ developed over a period of five months. He describes being scared in the beginning and wary of exercising these powers, but has now accepted ‘the task’ he has been assigned, and has spent at least two hours a day in meditation – ‘working full time for the last seven months or so.’

He says that he continues to function normally, and spends the rest of his day walking the dog, tidying the house and watching TV boxsets. He has been unemployed since losing his job as a civil servant a year and a half ago. He describes a period of low mood after losing his job, which lasted for roughly three months. He denies any recent low mood. His concentration levels are good, but are generally worn out after his meditation. He concedes that he spends a lot of time on his own, but that he continues to engage in his usual social activities, playing pitch and putt every weekend with friends and an active member of his community’s Tidy Towns committee.

Past Psychiatric History: He was treated by his GP for a depressive episode in his early twenties. The episode resolved after nine months’ treatment with Escitalopram 20mg.

Family History: His family history is non-contributory.

Social History: He is unemployed. His wife works as a nurse, usually doing night shifts. He spends his time in meditation, gardening, walking the dog and watching TV boxsets. He stopped drinking alcohol a year ago. He continues to play pitch and putt with his friends once a week, but no longer goes out to the pub with them on weekend nights. He is an active member of his community’s Tidy Towns committee, organising the flowers and taking part in fundraising events. He denies ever taking illicit drugs.

Collateral History: His wife is very surprised by her husband's disclosures. She says that he has never mentioned anything like this before, and that he has not been acting noticeably different over the last year. She admits to not spending very much time with him.

Mental State Examination: He presented as a calm, alert man, dressed casually in a polo shirt and jeans. He showed evidence of good self-care. His behaviour throughout the review was appropriate. His eye-contact was good and rapport was established. The volume of his speech was slightly low, otherwise normal. He said that his mood was fine, and objectively it was euthymic. His affect was reactive. He denied any suicidal ideation. There were no obvious abnormalities in thought form. Thought content included systematised delusions, described above. He denied any perceptual abnormalities. He did not have any cognitive deficits. He was convinced in the veracity of his beliefs and refused to entertain the possibility that they might be delusions.

Patient E (pilot study only)

Chief Complaint: This is a 72 year old woman, divorced, living alone, with no previous contact with psychiatric services, referred by her GP with persecutory delusions for several months.

History of Presenting Complaint: She does not believe that there is a problem with her mental health, and says this referral is all part of a conspiracy to have her 'labelled as disabled.' It is being orchestrated by her neighbours, whom she admits were once her friends but have become envious of her in recent months and want to 'bring me down.' The problems began after she returned from hospital fourteen months ago. She suffered a stroke. There were no focal neurological deficits but she remained weak for a number of months. She converted her sitting room into a bedroom and spent the majority of her time recuperating there.

'The sitting room windows are frosted. And while I was lying there I could see the shapes of people snooping around outside. It was them – the neighbours. I was very frightened.' She says she has now returned to full fitness, but has continued to notice a number of suspicious things. She believes her house has been entered many times. She has never caught anyone yet. Recently she had the people from the alarm company come out and reset her code.

She says that the worst thing that has happened is the disappearance of important papers from her safe. When pressed she will not be specific but says that these papers include her medical records. She also believes that her phone has been hacked. She gets cut off in phone calls and sometimes her messages fail to deliver. She is convinced that her neighbours are conspiring against her. These beliefs are unshakeable by rational argument.

She does not believe that her thoughts are being interfered with or that her body is being controlled in any way. She denies hearing voices, or experiencing hallucinations in any other sensory modalities. She says that her sleep, appetite and energy levels are the same as they have always been. She says that her mood is good; she just wants to be left alone. When pressed on whether she has done anything or will do anything about her perceived neighbours' actions, she replied in the negative: she had done nothing and hopes that by ignoring them they will leave her alone soon. She reports no physical problems apart from an ongoing backache. She denies any problem with her memory.

Past Psychiatric History: She has never had contact with psychiatric services in the past.

Past Medical History: Stroke 14 months ago. Hypertension. Hypcholesteraemia. Right hip replacement 2 years ago.

Medication: Clopidogrel. Esomeprazole. Amlodipine. Atorvastatin.

Family History: One of her sons was diagnosed with schizophrenia in his teens. Her other son is healthy and well.

Social History: She lives on her own. She is a semi-retired dressmaker. She continues to work from home. She spends her time working, reading, cleaning, gardening and listening to the radio. She plays bridge twice a week. She doesn't drink or smoke.

Collateral History: One of her friends, her bridge partner, described her as 'the kind of woman who has her quirks.' But also said she had become noticeably more paranoid over the last five months, constantly talking about her neighbours. Her memory, mood and normal level of functioning seem unaffected.

Mental State Examination: She is a thin woman, appropriately and well-dressed, neat and tidy. She was very animated in conversation, and played constantly with a necklace. Eye contact and rapport were good. She displayed no abnormal movements and she was not distracted. The rate, rhythm and volume of her speech was normal. She described her mood as good; objectively it was euthymic. Her affect appeared reactive and appropriate, with no signs of irritability or incongruous reactions. She denied any thoughts of self-harm. She has a number of paranoid delusions. She believes that people have been in her house. She believes that papers have been stolen from her safe. She believes that her phone has been hacked. She believes that her neighbours are envious of her and want to interfere with her life. She believes they are trying to have her declared insane. These beliefs are unshakeable and she will not entertain any rational explanations. She denies any thought insertion, withdrawal or broadcasting. She denies any hallucinations or illusions. She was alert and orientated. Both her short and long term memory are good. Her score on MMSE was 29/30. Clock drawing and frontal lobe tests normal.

Physical Examination: Vital signs are normal. Apart from signs of her stroke 14 months ago, there are no abnormal findings on physical examination or routine blood testing.

Appendix I: Correct / acceptable diagnoses (Think-aloud study, Chapter 4)

Patient A

Correct diagnosis: Mania / Bipolar affective disorder; Acute dystonic reaction

Acceptable diagnosis: Schizophrenia; Psychosis; Extrapyrarnidal side-effects

Scoring: 0.5 awarded for each element of correct diagnosis, 0.25 for each element of acceptable diagnosis

Patient B

Correct diagnosis: Depression

Acceptable diagnosis: Hypercalcaemia

Scoring: 1 awarded for depression, 0.5 awarded for hypercalcaemia

Patient C

Correct diagnosis: Normal pressure hydrocephalus

Acceptable diagnosis: Dementia

Scoring: 1 awarded for normal pressure hydrocephalus, 0.5 awarded for dementia

Patient D (pilot study only)

Correct diagnosis: Schizoaffective disorder

Acceptable diagnosis: Schizophrenia; Delusional disorder

Patient E (pilot study only)

Correct diagnosis: Delusional disorder

Appendix J: Semi-structured interview schedule (Think-aloud study, Chapter 4)

Participant number _____
Date _____
Year graduated _____
Year of study _____

After each case

- Reflect:** Can you tell me how you went about your diagnosis?
- Confidence:** How confident are you in your final diagnosis?
- Guided reflection:** How did you find using the guided reflection process for this case?
You chose to use guided reflection <at this point>. Can you tell me about that?
What aspects of the guided reflection process did you use?
- Anything else:** Anything else you'd like to say about this case?
- Reminder:** Remember to speak out loud as you're diagnosing the next case. The guided reflection process is there for you to use if you wish.

General questions

- Aspects:** How did you find using the guided reflection process for these cases?
Are some aspects more helpful or appealing than others?
- Real-world use:** Is this a process you could see yourself using in real life?
What would encourage you to use it?
What would discourage you from using it?
- Error applications:** How closely does this process align with your own natural diagnostic process?
Helping doctors with their diagnostic process is ultimately about preventing diagnostic error. What are your thoughts on what can be done to prevent errors?
Do you think a process like this could be helpful in preventing errors?
- Anything else:** Anything else you'd like to say?

Appendix K: Experimenter script (Think-aloud study, Chapter 4)

Brief introduction:

For this study, I'm going to give you some clinical cases to diagnose. I'm going to show you a method for diagnosing cases that you can choose to follow if you think it's helpful. I'd like you to diagnose the case 'out loud', speaking your thoughts, and I'll record what you say so that I can transcribe and analyse it later.

Guided reflection:

The method I'm going to show you is called guided reflection, and it's based on reflective practice. Studies show that it may be helpful when diagnosing complex cases. We want to understand how doctors use it.

The first step is to think of your first diagnostic hypothesis – the first thing that comes to mind. Then reread the case. Then gather evidence that supports the hypothesis, refutes the hypothesis, and evidence that you would expect to be there if you're correct, but that's actually missing. The second step is to generate alternatives. And do the same thing – for each alternative, gather evidence for and against, and evidence that you would expect to be there if you're correct, but that's actually missing. You can generate as many alternatives as you like. Finally, you rank your alternatives. And state your final hypothesis. That's there to use if you want to for some or all of the cases, it's up to you.

Cases:

There are three fictional scenarios, based on real-life cases, and they're just there to stimulate your thinking.

There's pen and paper there if you'd like to use them, and feel free to scribble on the cases or take notes.

We can take about 45 minutes for the cases, and then the last 15 I'll ask your general impressions about the cases and this process.

So as you're going through the case, remember to speak out loud, or you can read out loud if you like, and let me know what you're thinking about, what you're focusing on and what's coming to mind for you.

After each vignette:

Remember to think out loud as you're going through the case, and remember that the guided reflection process is there for you to use if you choose to do so.

After one minute's silence:

What are you thinking about at the moment?

Appendix L: Guided reflection written instructions (Think-aloud study, Chapter 4)

Initial hypothesis

State **the first diagnosis that comes to mind** (your diagnostic hypothesis)

Reread the clinical scenario.

Recall any details from the scenario that **support** your first diagnostic hypothesis

Recall any details from the scenario that **refute** this hypothesis

Recall any details that would be **expected** if the hypothesis were true, but that are **missing** from the scenario

Alternatives

State one **alternative possible diagnosis**

Recall any details from the scenario that **support** this diagnostic hypothesis

Recall any details from the scenario that **refute** this hypothesis

Recall any details that would be **expected** if the hypothesis were true, but that are **missing** from the scenario

Repeat with additional alternative possible diagnoses

Diagnose

Rank your alternatives in order of likelihood

State your **final** diagnosis

Appendix M: Coding framework (Think-aloud study, Chapter 4)

Diagnostic process codes

Codes based on elements of the guided reflection process and other observed diagnostic reasoning behaviours

D1: Hypotheses

Participant names diagnostic hypotheses

Definition:

- (a) Initial hypothesis: Participant names the first possible diagnosis for the case
- (b) Alternative hypothesis: Participant names other possible diagnoses for the case
- (c) Refinement: Participant refines a hypothesis to make it more specific, indicate a subtype or to add specific detail

Examples:

- (a) Initial hypothesis: “So the initial thing that’s first coming to mind before reading anything else would be em, I’m wondering whether it’s an acute psychotic illness.”
- (b) Alternative hypothesis: “This now is appearing to me as more manic type behaviours, so that’s also in my mind now ... differential of a manic episode.”
- (c) Refinement: “A psychotic element is also a feature of his mania, so maybe it’s a dual presentation.” “That’s pushing towards more severity.” “So it would be bipolar one if it was with manic symptoms.”

D2: Identifying evidence

Participant comments on features of the case and identifies them as types of evidence in relation to a mentioned hypothesis

Definition:

- (a) Supporting evidence: Participant comments on a feature of the case that supports a mentioned hypothesis
- (b) Refuting evidence: Participant comments on a feature of the case that refutes a mentioned hypothesis
- (c) Missing evidence: Participant comments on a feature of a mentioned hypothesis that they expect to see, but that is missing from the case

Examples:

- (a) Supporting evidence: “Increase in urinal frequency, that’s in keeping with anxiety.”
- (b) Refuting evidence: “Normal tone, so that’s going against the Parkinsonian theory.”
- (c) Missing evidence: “You’d expect cog wheel rigidity in that em situation.” “[One alternative is] drug-induced, but I would expect there to be something showing up in the toxicology screen.”

D3: Gathering evidence

Participant reviews the case and gathers evidence around a mentioned hypothesis

Definition:

- (a) Gathering supporting evidence: Participant reviews the case and gathers supporting evidence for a mentioned hypothesis
- (b) Gathering refuting evidence: Participant reviews the case and gathers refuting evidence for a mentioned hypothesis
- (c) Gathering missing evidence: Participant gathers features of a mentioned hypothesis that they expect to see, but are missing from the case

Examples:

- (a) Gathering supporting evidence: “Things that support it are the presence of low mood for more than two weeks as meeting criteria for the entirety of the day, little variation. The biological symptoms are really convincing. So she’s got the fatigue, eh, poor motivation, poor concentration, poor energy, hopelessness, suicidal ideation, early morning wakening, decreased appetite, that’s all in keeping with em depression.”
- (b) Gathering refuting evidence: “What refutes it? This abdominal pain is eh, is not in keeping with depression, so that refutes it. Em, and death of her grandmother four months ago is eh also not in keeping with depression, that would be more so eh, an abnormal kind of grief reaction or bereavement.”
- (c) Gathering missing evidence: “What would you expect otherwise? Em. You might expect his MMSE to be a little bit lower, given the kind of the collateral saying that he’s really bad from the collateral, he’s not cooking and his memory’s really bad, but em the MMSE is twenty-four, so it could be worse.”

D4: Selection

Participant chooses among hypotheses

Definition:

- (a) Ranking: Participant ranks their potential diagnoses
- (b) Final diagnosis: Participant names a final diagnosis, distinct from ranking

Examples:

- (a) Ranking: “If I ranked them, maybe I’d go with Wernicke’s one, vascular dementia two, em, Alzheimer’s three, Parkinson’s four.”
- (b) Final diagnosis: “Final diagnosis would probably be a first episode psychosis here with a dystonic reaction.”

D5: Explicit use of guided reflection

Participant clearly begins to use the guided reflection framework

Definition:

- (a) Explicit use of guided reflection: Participant clearly begins to use the guided reflection framework, either explicitly mentioning it or using language to indicate this

Examples:

- (a) Explicit use of guided reflection: “So em to go back to the guided reflection instructions...” “So initial hypothesis...”

D6: Other diagnostic behaviours

Participant engages in an identifiable strategy or diagnostic behaviour

Definition:

- (a) Criteria: Participant makes reference to DSM or ICD criteria for psychiatric diagnosis
- (b) Statistical reasoning: Participant makes reference to statistical information concerning diagnosis
- (c) Risk factors: Participant makes reference to risk factors for diagnoses mentioned
- (d) Seeking information: Participant expresses a question or desire for particular further information
- (e) Querying patient report: Participant expresses doubt in truth or validity of patient report of behaviour or symptom
- (f) Urgency / prioritisation: Participant makes reference to a particular symptom as requiring immediate attention, or to a diagnosis that must be ruled out as a matter of urgency
- (g) Personal experience: Participant makes reference to their personal experience with a given diagnosis
- (h) Seeking support: Participant makes reference to seeking a second opinion or aid from colleagues

Examples:

- (a) Criteria: "So at this point he's kind of meeting a few of the criteria for a manic episode."
- (b) Statistical reasoning: "Alzheimer's is the most common form of dementia, so that has to be in the list."
- (c) Risk factors: "Past medical history is positive for hypertension and hypercholesterolemia, I guess that's important because they're both risk factors for vascular dementia." "I might be a bit more worried about this gentleman immediately because he's sixty-eight, male, and he's retired, so he's kind of at risk, he's in a high-risk category already before we even start."
- (d) Seeking information: "Might be nice to see what the paracetamol level is and just explore that more."
- (e) Querying patient report: "Maybe he's feeling low about it and there might be an attempt to harm himself there too that he's not em disclosing." "Adamant that he's unlikely to have ever tried alcohol. You can take the word for that but obviously you need to rule that out by toxic screen."
- (f) Urgency / prioritisation: "So what I'd act on immediately would be em management of neuroleptic malignant syndrome, secondary to anti-psychotic use, em, because that's the most life-threatening thing, that's what could kill him, em."
- (g) Personal experience: "Wouldn't be as confident in saying about the NMS because of the fact that I haven't actually seen one in real, I haven't actually seen it in real life."
- (h) Seeking support: "I would probably look for em help from colleagues with regards to the, to the rest, what was going on."

Reflection codes

Codes based on topics that have arisen during reflection on the diagnostic process and interview

C1: Implementation of guided reflection

Participant discusses supports, barriers or triggers for the use of guided reflection in vignettes or real-life cases

Definition:

- (a) Triggers: Participant names personal, environmental or case-specific factors that would prompt the use of guided reflection for vignettes or real-life cases
- (b) Supports: Participant names personal or environmental factors that would support the participant in the use of guided reflection in a real-life case
- (c) Barriers: Participant names personal or environmental factors that would hinder the participant in the use of guided reflection in a real-life case

Examples:

- (a) Triggers: “So useful in a, more, as I said, in the more complicated patients where you can take a minute and maybe try and save yourself time by breaking it down, but in the simple cases, on a day-to-day basis, like in a practice or in ED or even on the wards, I think time doesn’t allow for it, most of the time, unless you’re really stuck and it’s really tricky then, breaking it down might actually be useful, yeah.”
- (b) Supports: “[regarding pro-forma admission paperwork] ““Maybe if there was a section beside it, supportive features and refuting features and then alternatives? Maybe that might be a way to prompt or reminder of, of, of thinking about it in that way.”
- (c) Barriers: “Busy clinical workload and you just mightn’t remember to.”

C2: Impact of guided reflection on diagnostic process

Participant describes how guided reflection impacted their diagnostic process

- (a) Slowing down: Participant describes guided reflection as prompting them to slow down or not jump to conclusions
- (b) Expanding reasoning: Participant describes guided reflection as promoting broad differential diagnosis, lateral thinking, and diagnoses / questions the participant would not otherwise have considered
- (c) Reminder: Participant describes guided reflection as reminding or prompting them to think through a case more completely
- (d) Visible structure: Participant describes guided reflection as a framework or tool to structure diagnostic thinking

Examples:

- (a) Slowing down: “It would encourage further eh reading and yeah, exploring the alternatives more so than just picking one and going with that.”

- (b) Expanding reasoning: “Mightn’t even have considered em Wernicke’s until you have to pick the supporting features and the refuting.”
- (c) Reminder: “It’s kind of a reminder to consider alternatives.” ”You kind of have your list in the back of your mind but to put it into context, sometimes, unless you’re reminded, em, you mightn’t think about it that way.”
- (d) Visible structure: “It’s nice to have it kind of put out in front of you like that and to put a framework on it to direct you, so you’re not, to make you think about those things specifi--, specifically.”

C1: Experience of guided reflection

Participant makes reference to how they experienced the use of guided reflection

Definition:

- (a) Impressions: Participant comments on their general impressions of guided reflection
- (b) Helpful components: Participant describes particular components or steps of guided reflection as helpful
- (c) Unhelpful / difficult components: Participant describes particular components or steps of guided reflection as unhelpful or difficult to use
- (d) Confidence: Participant describes guided reflection as increasing or decreasing diagnostic confidence, encouraging second-guessing
- (e) Natural process: Participant refers to natural diagnostic process and how guided reflection aligns or does not align with this
- (f) Case complexity: Participant refers to the use of guided reflection with more or less complex vignettes and real-life cases

Examples:

- (a) Impressions: “Generally helpful, absolutely.” “I suppose it’s straightforward and easy to use.”
- (b) Helpful components: “So I think maybe for me the most helpful was when you see more things that support your diagnostic hypothesis , then you’re more reassured that you have chosen the right diagnosis.”
- (c) Unhelpful / difficult components: “I suppose unless you have lots of experience or, you know, you have stuff in your hand or in front of you, sometimes you can actually just forget what’s missing. So that part is probably less useful kind of on a day-to-day basis, em, but you can always see what’s there in front of you and what, what supports this and that.”
- (d) Confidence: “I think you can be more confident when you’ve used that because you’re not just plucking it from the sky, you’re saying well, here’s A, B, C, and D reasons why I think this is, so em it’s not just a hunch or a feeling.”
- (e) Natural process: “I think this is how I usually think anyway, to be honest.” “I think, I think we do a certain amount of it subconsciously anyway.”
- (f) Case complexity: “I think again because it was quite clear-cut that it was just easy to see like each line was just filling the criteria, very easy to say what supported and what refutes it.” “It was maybe more useful in the last one, where there were more things going on and a couple of different factors there.”

C4: Diagnostic error

Participant discusses diagnostic error in general

Definition:

- (a) Scope of error: Participant comments on the scope and scale of diagnostic error
- (b) System factors: Participant refers to features of the healthcare system or working conditions that contribute to error
- (c) Personal factors: Participant refers to personal factors that contribute to error
- (d) Shame: Participant refers to shame or lack of openness around diagnostic difficulties

Examples:

- (a) Scope of error: "I think there's a lot of errors happening all the time that aren't kind of picked up on or reported on."
- (b) System factors: "I think if we had more time, we definitely would be making less mistakes."
- (c) Personal factors: "Definitely your own kind of mind state at the time as well, if, if you're feeling under pressure or you're tired, you know, kind of dealing with twenty-four-hour, thirty-six-hour shifts where at the end of it your, you know, your thought processes aren't clear."
- (d) Shame: "I think possibly a huge part of it is em maybe a feeling of, of em incompetence or shame, when you can't accurately make the diagnosis."

C5: Diagnosis in psychiatry

Participant discusses diagnosis in psychiatry as compared to other specialties in medicine

Definition:

- (a) Emerging over time: Participant refers to psychiatric diagnoses emerging over time
- (b) Secondary priority: Participant refers to psychiatric diagnosis being relatively less important than other aspects of care

Examples:

- (a) Emerging over time: "I think there would be a lot of uncertainty in the definite diagnosis at this point, seeing as it's his first presentation."
- (b) Secondary priority: "Obviously at three o'clock in the morning you just, you're not, you're not very concerned about em diagnosis, your, your main concern is about em risk." "Because there is a lot of crossover even in the treatment regimes with, you know, if someone comes in with a psychosis, they're going to treat that the same no matter what the cause initially."

C6: Methodology

Participant comments on the materials or experience of participating

Definition:

- (a) Vignettes vs real cases: Participant comments on the differences between vignettes and real cases

Examples:

- (a) Vignettes vs real cases: “When you’re trying to gather a history from a patient, it, it doesn’t come out in a nice kind of package.”

Appendix N: Ethical approval (Experimental study, Chapter 5)



Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

F.A.O. Kathryn Lambe

School of Psychology Research Ethics Committee

11th December 2017

Dear Kathryn,

The School of Psychology Research Ethics Committee has reviewed your application entitled "Guided reflection intervention for diagnostic reasoning" and I am pleased to inform you that it was approved.

Please note that you will be required to submit a completed **Project Annual Report Form** on each anniversary of this approval, until such time as the research is complete and the thesis is submitted. The form is available for download from the Ethics section of the School website.

Adverse events associated with the conduct of this research must be reported immediately to the Chair of the Ethics Committee.

Please ensure that you keep this letter safe for inclusion with final project/dissertation/thesis.

Yours sincerely,

A handwritten signature in blue ink, appearing to read 'Clare Kelly'.

Dr Clare Kelly
Acting Chair
School of Psychology Research Ethics Committee

Scoil na Siceolaíochta

Dámh na nEolaíochtaí Sóisialta agus Daonna,
Áras an Phiarsaigh, Coláiste na Tríonóide,
Baile Átha Cliath 2, Éire.

School of Psychology

Faculty of Arts, Humanities and Social Sciences,
Áras an Phiarsaigh, Trinity College,
Dublin 2, Ireland.

T +353 1 896 1886
F +353 1 671 2006
psychology@tcd.ie
www.tcd.ie/psychology

Appendix O: Recruitment message (Experimental study, Chapter 5)

Dear student,

I am a postgraduate researcher from the School of Psychology at Trinity College Dublin, working with the Discipline of Psychiatry and the School of Medicine, and I am conducting a study that explores the way in which medical students and doctors make decisions. I would like to invite you to take part in my research.

Participation involves reading some medical scenarios and offering your opinions on the situations they describe.

It is hoped that the results of the study will help to advance our understanding of the factors that contribute to medical decision-making for students and doctors and to design useful educational materials to improve decision-making.

The study will take place at <date and time> and will take under one hour to complete. Participation is completely voluntary and so you are not obliged to take part.

If you have any questions about the research, feel free to contact me at lambeka@tcd.ie.

Kind regards,

Kathryn Lambe

Appendix P: Information sheet (Experimental study, Chapter 5)

Title of study: Impact of guided reflection interventions on diagnostic accuracy

Introduction:

Existing research suggests that the majority of medical errors arise from errors of thinking, rather than errors of technique, and that it may be possible to reduce such error through cognitive strategies. This study aims to identify how these strategies can help or hinder students and doctors in their diagnostic decisions.

Your participation should take no longer than one hour. Your participation is entirely voluntary, and your participation or non-participation will have no implications for your studies, grades or future job opportunities.

Procedures:

You are eligible to take part in this study if you are:

- Aged 18 years or older, and
- Enrolled in an undergraduate or graduate entry medical course at an Irish university.

If you decide to take part, you will be asked to complete a short demographic questionnaire. You will then be asked to read a series of fictional patient cases and to make diagnoses based on the information contained therein.

Benefits:

The study offers no direct benefits to participants. However, it is hoped that the findings will help to advance our understanding of how doctors make decisions and how we may further educational efforts to improve medical decision-making. The implications of this research may help educators to improve curricula and teaching methods, thereby contributing to the quality of care experienced by patients in the future.

The study may also offer you the opportunity to reflect on your own decision-making processes, which participants in similar studies often report as being an interesting experience that benefits their practice.

Risks:

It is not anticipated that you will be exposed to any physical, psychological or social risks through this study that you would not encounter in everyday life. As the study asks you to make diagnoses based on some fictional patient cases, you might find this portion of the study to be exam-like or stressful. At the end of the study, you will be provided with the details of some agencies that may offer support if you feel upset or distressed. Again, your participation is entirely voluntary, and your participation or non-

participation will have no implications for your studies, grades or future job opportunities.

Exclusion from participation:

No exclusion criteria apply.

Confidentiality:

The information you give us will be anonymous; we will not ask for your name or any information that could reasonably identify you. Therefore, your name cannot be linked in any way to your answers to our questions.

Insurance:

This study is covered by insurance policies organised by Trinity College Dublin.

Voluntary Participation:

You have volunteered to participate in this study. You may withdraw at any time. If you decide not to participate, or if you withdraw, you will not be penalised and will not give up any benefits which you had before entering the study. Your participation or non-participation will have no implications for your studies, grades or future job opportunities. As your data will not be identifiable, it may not be withdrawn from the study once you have submitted it.

Stopping the study:

You understand that your Principal Investigator may stop your participation in the study at any time without your consent.

Permission:

The study has received approval from the Trinity College Dublin School of Medicine Research Ethics Committee.

Further information:

You can get more information or answers to your questions about the study, your participation in the study, and your rights, from any of the research team listed below.

If your Principal Investigator learns of important new information that might affect your desire to remain in the study, you will be informed of this.

Kathryn Lambe
PhD Candidate, TCD School of Psychology
Email lambeka@tcd.ie

Prof Brendan Kelly
TCD School of Medicine
Email brendan.kelly@tcd.ie
Tel 01 896 3799

Dr David Hevey
TCD School of Psychology
Email heveydt@tcd.ie
Tel 01 896 2406

Appendix Q: Debriefing sheet (Experimental study, Chapter 5)

Title of study: Impact of guided reflection interventions on diagnostic accuracy

Thank you very much for taking part in this research.

Existing research suggests that the majority of medical errors arise from errors of thinking, rather than errors of technique, and that it may be possible to reduce such error through acknowledgement of cognitive biases and deliberate implementation of cognitive strategies by individual doctors. This study aims to identify how these strategies can help or hinder students and doctors in their diagnostic decisions.

Your answers to our questions will not be identifiable and cannot be traced back to you.

Other students may have the opportunity to participate in this study or one like it in the near future. In the interests of preserving the integrity of the experiment, we kindly ask that you do not discuss the fictional cases with students who have not taken part in the study.

If you have found participating in this study stressful or feel in any way upset or distressed following participation, the following contacts may be able to assist you:

Samaritans
www.samaritans.org
Tel 116 123

Niteline
www.niteline.ie
Tel 1800 793 793

TCD Student Counselling Service
www.tcd.ie/Student_Counselling
Tel 01 896 1407

If you have any further questions about the study, please feel free to contact the research team.

Kathryn Lambe
PhD Candidate, TCD School of Psychology
Email lambeka@tcd.ie

Prof Brendan Kelly
TCD School of Medicine
Email brendan.kelly@tcd.ie

Tel 01 896 3799

Dr David Hevey
TCD School of Psychology
Email heveydt@tcd.ie
Tel 01 896 2406

Appendix R: Instructions and diagnostic tables (Experimental study, Chapter 5)

(1) Control condition

Instructions

Many thanks for taking part in this study.

This booklet contains a number of clinical scenarios. Each clinical scenario is followed by a table.

Please read each case carefully and enter your diagnosis in the accompanying table.

We ask that you do not consult with your fellow participants as you complete this task.

When you are finished, please return the booklet to the researcher.

Diagnostic table

	Write down the first diagnosis that comes to mind	
	Rate your confidence in this diagnosis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty	

(2) Short condition

Instructions

Many thanks for taking part in this study.

This booklet contains a number of clinical scenarios.

Each clinical scenario is followed by a diagnostic table, containing a series of steps that will help you to reach a diagnosis.

Please read each case carefully and complete the table as fully as possible in order to reach a diagnosis.

We ask that you do not consult with your fellow participants as you complete this task.

When you are finished, please return the booklet to the researcher.

Diagnostic table

PART 1			
Write down the first diagnosis that comes to mind (your original diagnostic hypothesis)			
Rate your confidence in this diagnosis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty			
Reread the clinical scenario and tick this box once you've done this			
Write down any details from the scenario that support your first diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario	
Rate your confidence in this diagnosis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty			

PART 2			
A1	Write down one alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario
A2	Write down a second alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario

PART 3	
Rate your confidence in your original diagnostic hypothesis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty	
Rank your alternatives in order of likelihood	1 2 3
Rate your confidence in your original diagnostic hypothesis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty	
Write down a final diagnosis	
Rate your confidence in your final diagnosis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty	

(3) Long condition

Instructions

Many thanks for taking part in this study.

This booklet contains a number of clinical scenarios.

Each clinical scenario is followed by a diagnostic table, containing a series of steps that will help you to reach a diagnosis.

Please read each case carefully and complete the table as fully as possible in order to reach a diagnosis.

We ask that you do not consult with your fellow participants as you complete this task.

When you are finished, please return the booklet to the researcher.

Diagnostic table

PART 1		
Write down the first diagnosis that comes to mind (your original diagnostic hypothesis)		
Rate your confidence in this diagnosis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty		
Reread the clinical scenario and tick this box once you've done this		
Write down any details from the scenario that support your first diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario
Rate your confidence in this diagnosis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty		

PART 2			
A1	Write down one alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario
A2	Write down a second alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario
A3	Write down a third alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario

A4	Write down a fourth alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario
A5	Write down a fifth alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario
A6	Write down a sixth alternative possible diagnosis		
	Write down any details from the scenario that support this diagnostic hypothesis	Write down any details from the scenario that refute this diagnostic hypothesis	Write down any details that would be expected if the hypothesis were true, but that are missing from the scenario

PART 3	
Rate your confidence in your original diagnostic hypothesis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty	
Rank your alternatives in order of likelihood	1 2 3 4 5 6 7
Rate your confidence in your original diagnostic hypothesis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty	
Write down a final diagnosis	
Rate your confidence in your final diagnosis on a scale of 1-6, where 6 indicates complete certainty and 1 indicates complete uncertainty	

Appendix S: Correct / acceptable diagnoses (Experimental study, Chapter 5)

Case 1

Gold standard diagnosis: Appendicitis

Acceptable diagnoses: Appendix mass; atypical appendicitis

Case 2

Gold standard diagnosis: Amoebic liver

Acceptable diagnoses: Acanthamoebiasis; amoeba infection; amoebiasis; amoebiasis cysts; amoebic dysentery; amoebic liver cyst; entamoeba histolytica; hepatic abscess; hepatic abscess of unknown origin; hepatic amoebiasis; hepatic cyst; hepatitis; infectious disease of the liver, tropical parasite; liver amoebiasis (cysts); liver cyst with pleural involvement; liver cystic lesion; liver cystic rupture; liver cysts secondary to tropical parasite; liver parasite; parasitic cystic lesions; parasitic infestation; pyogenic liver abscess

Case 3

Gold standard diagnosis: Colon cancer

Acceptable diagnoses: Ascending colon carcinoma; bleeding colonic polyps; cancer of the large bowel; colon cancer with liver metastasis; colorectal cancer; colorectal carcinoma; GI bleeding secondary to malignancy (CRC); GI malignancy; HCC / other cancer; lower GI bleed / malignancy; lower GI bleed secondary to metastatic colorectal carcinoma; malignancy; metastatic colorectal cancer; microcytic anaemia possibly due to GI bleed, query colon cancer; perforation of the bowel secondary to CRC; rectal cancer

Case 4

Gold standard diagnosis: Guillaine Barré Syndrome

Acceptable diagnoses: n/a