Introduction

Diagnosis and Error

Imagine a medical consultant within a hospital's intensive/critical care unit. They are asking a colleague for advice about a particular patient. The patient has presented with a series of symptoms, including dizziness, breathing difficulties and eventual chest pain. She has been placed under continuous monitoring of her 'vital signs', including heart rate, body temperature, blood pressure, blood oxygen saturation and respiration rate. There has been a slow decrease in her blood pressure and blood oxygen saturation. The consultant is deciding what the most likely cause of this patient's symptoms and how this may inform her future care/treatment. It is possible that the patient is suffering from pulmonary oedema, whereby fluid is collected in the air sacs of the lungs, causing severe and sometimes fatal congestion. The symptoms could also be suggestive of a tension pneumothorax, when a lung collapses. Alternatively, there could be a cardiac cause of the patient's condition. The consultant must integrate the information they have so far, align their individual mental models of the patient with colleagues and decide the following:

1. Do they have enough information to diagnose the patient's condition?
2. If not, what extra information do they need? Are there further tests that need to be performed?
3. What actions should they start taking to treat the patient given the most likely diagnosis?

One of the difficulties within this scenario is that symptoms may be indicative of multiple underlying conditions, which is a common occurrence in medical making. This example is illustrative of why many medical decisions are 'ill-structured'

problems: they present several possible courses of action, and produce disagreements over both the current hypothesis for the patient's condition and desired end goal for that patient's care (Jonassen, 1997). During this thesis, we will investigate the cognitive mechanisms of medical diagnoses. Diagnosis is a core aspect of several medical subdisciplines and we choose it as an area of study for a few reasons. Firstly, accurate diagnosis is crucial to a patient's treatment. Secondly, from a psychological standpoint, it allows for an extension of previous research on information gathering and confidence to an ecologically valid, real-world setting. Finally, past work looking at diagnosis has not yet provided clarity on the causes of diagnostic errors.

Accurate medical diagnosis is crucial to safe, high quality patient care and is a core part of a doctor's job. Research on diagnosis has been grounded in the incidence of errors. Looking into medical errors allows healthcare systems to learn from mistakes and improve technical and safety processes for future patients. A report from the US Institute of Medicine (McGlynn, McDonald & Cassel, 2015) concluded that most patients will experience a diagnostic error within their lifetime. When looking at records of new diagnoses for spinal epidural abscess in the US Department of Veteran Affairs, Bhise et al. (2017) found that up to 55.5% of patients experienced diagnostic error. Other papers have estimated a lower incidence of diagnostic error. The Harvard Medical Practice Study found that diagnostic errors were responsible for 17% of adverse events (Kohn, Corrigan & Donaldson, 1999), whilst the Canadian Adverse Events Study found this value to be 10.5% (Baker, Norton & Flintoft, 2004). The Quality in Australian Health Care Study found that 20% of adverse events were due to delayed diagnosis (Wilson et al., 1999). Even using the most conservative of these estimates, the scale of the diagnostic error is substantial when extrapolated to the population of patients. Past work has hence attempted to find the sources of diagnostic error. Understanding the common sources of medical errors and adverse events can then be extremely valuable for improving healthcare in the future. Around 32% of clinical errors have been found to be caused by clinician assessment, particularly the clinician's failure to weigh up competing

diagnoses (Schiff et al., 2009). Diagnostic errors have also been found to lead to longer hospital stays and even increased patient mortality (Hautz et al., 2019). In addition to stays in hospital, errors also manifest in differences to treatment for patients. Unnecessary treatment (or 'overtreatment') was estimated to cost the US healthcare system 158-226 billion dollars in 2011 (Berwick & Hackbarth, 2012). There has been increasing emphasis on overtesting, such as requesting costly imaging scans when they may not be medically necessary (Carpenter, Raja & Brown, 2015). Salem-Schatz, Avorn and Soumerai (1990) also found that 61% of surveyed resident doctors had ordered unnecessary transfusions at least once a month due to a suggestion to do so by a more senior physician.

Diagnostic error is by no means the sole cause of medical incidents. There are a number of factors tied to the wider work environment, culture and technology that can contribute to incidents and errors. Individuals involved in clinical decision making have to frequently contend with an uncertain decision making environment, as well as time pressure and personal stresses (Orasanu & Connolly, 1993). However, by understanding the individual psychological factors of the diagnostic process, we better understand how sociotechnical and environmental factors interact with and amplify individual contributors to diagnostic error. Gaining a greater understanding of the causes of diagnostic error can have important implications for future interventions within healthcare settings.

Cognitive Biases in Diagnoses

Diagnostic error can stem from cognitive biases during the diagnostic decision making process, such as primacy (Frotvedt et al., 2020) or recency (Chapman, Bergus & Elstein, 1996) biases. These mean that individuals tend to overweight information or stimuli that arrive earlier or later before a decision respectively. While it seems intuitive that classical decision making biases affect those in healthcare too (Restrepo et al., 2020), the empirical evidence of impact for medical decision

making is scant, (van den Berge & Mamede, 2013). One example from dermatology looked found examples of satisficing bias (also known as premature closure, whereby clinicians arrive at an incorrect diagnosis too early and subsequently stop considering alternatives and seeking other information) and anchoring (whereby individuals are highly influenced by a reference point that considering responses to a decisions are judged against), but few examples of others such as availability bias (overweighting information or symptoms that correspond with a recent or memorable case from a clinician's experience) (Crowley et al., 2012). One type of bias that has manifested in more experimental findings is overconfidence (Berner & Graber, 2008, Meyer et al., 2013), in which clinicians report them as being more confident in their decisions than they should be given their objective accuracy. We shall now explore how confidence has been studied within cognitive psychology and why it is very relevant to medical diagnoses.

Confidence and Miscalibration

Confidence can be viewed as one's "subjective probability of a decision being correct" (Fleming & Daw, 2017). Confident individuals tend to be more influential on others in a group (Zarnoth & Sniezek, 1997) and can even causally increase the confidence of other observers (Cheng et al., 2021). This behaviour has been observed in mock jury trials, during which participants hear eyewitness testimonies presented with high confidence and then perceive those testimonies as more credible than testimonies provided with low confidence (Cutler, Penrod & Dexter, 1989, Roediger, Wixted & DeSoto, 2012). Confidence is a commonly used predictor of another person's accuracy, especially when feedback is not readily available of an individual's true accuracy. Confidence also varies across individuals with what may be considered a 'subjective fingerprint' (Ais et al., 2016), and individuals may be systematically underconfident or overconfident. Confidence has been explained computationally as the difference in the strength of evidence for a decision alternative compared to other alternatives (Vickers & Packer, 1982). After a decision is made, we continue to process evidence, i.e. we continue to think about a decision after it has been

made and having 'second thoughts' or changes of mind are more likely with a lower level of confidence (Charles & Yeung, 2019).

We refer to confidence as being 'calibrated' if it closely predicts objective accuracy (i.e., such that the individual is neither overconfident nor underconfident, see Figure 1 below). In experimental studies, confidence sometimes exhibits impressive calibration to objective accuracy (Boldt & Yeung, 2015), which is thought to reflect people's ability to evaluate the quality of evidence on which they base their decisions (Shekhar & Rahnev, 2023). But calibration is rarely perfect because confidence also depends on factors that do not directly correlate with accuracy, such as the time spent deliberating and the total amount of evidence considered (independent of the quality or consistency of this evidence) (Kiani, Corthell & Shadlen, 2014, Ko et al., 2022), as well as the mood (Rouault et al., 2018), personality (Schaefer et al, 2004), gender (Syzmanowicz & Furnham, 2011) and status (See et al., 2011) of the decision maker. Miscalibration of confidence come from a lack of deliberation over one's decisions and confidence, leading to an overreliance on intuitive decision making (Mata, Ferreira & Sherman, 2012). The resulting under- and overconfidence matters: overconfident decision makers leap to premature conclusions and ignore useful information or advice, while underconfident decision makers waste time collecting evidence that will not improve their decisions (Desender, Boldt & Yeung, 2018). Effective decision making in groups likewise depends on team members sharing calibrated information about their uncertainty: Confident team members tend to be listened to more, which can lead others astray if they are overconfident (Zarnoth & Sniezek, 1997). Conversely, underconfident team members may be ignored or may fail to share potentially useful information (Silver, Mellers & Tetlock, 2021).

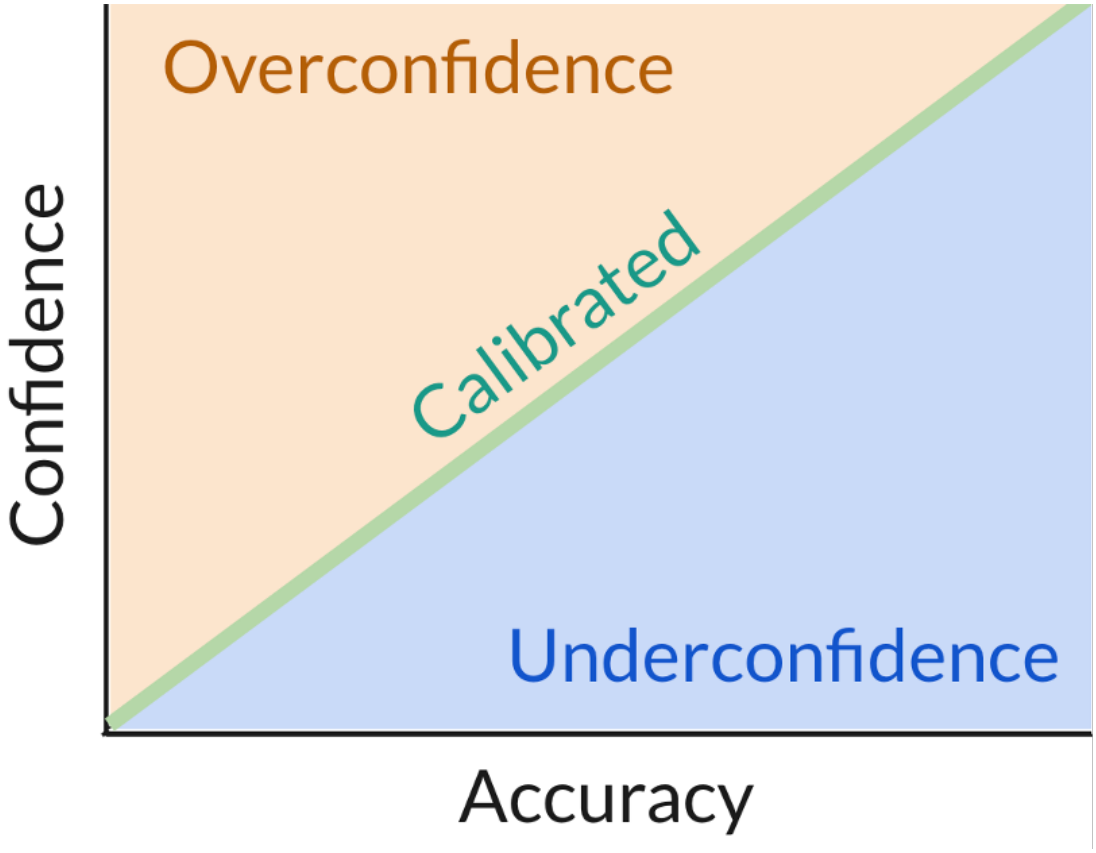

Figure 1: Visual representation of confidence calibration when comparing objective accuracy (x-axis) to subjective confidence (y-axis). Confidence is said to be calibrated when the two are relative equivalent (green line). Individuals are considered underconfident when their confidence is lower than their true accuracy (blue area) and overconfidence when their confidence is higher than their true accuracy (orange area).

These features of confidence highlight its potential importance in healthcare, as overconfidence can lead to insufficient consideration of diagnostic alternatives and inadequate care (Kovacs, Lagarde & Cairns, 2020). In the absence of objective feedback, confidence can be used as a marker of how likely someone is to be correct (Price & Stone, 2004). In medicine, a lack of clearly communicated feedback can cause clinicians to proceed as if they have received positive feedback. This means that they may not adequately update their internal model of the patient and then increase their confidence inappropriately, whether working individually or in teams (Jaspen et al, 2022).

Building on the current research landscape of diagnostic confidence is important. If there is an assumption that others will calibrate their confidence to their true accuracy, this would mean that heeding high confidence advice or judgements would be an optimal strategy for maximising accuracy. However, this can be a serious issue when high confidence errors lead others astray. This is important, as in addition to seniority and speciality experience, a clinician's confidence is one of the only markers available for other clinicians and for patients when making key medical decisions.

Some past work has explored instances of miscalibrated confidence specifically within the context of diagnoses. Meyer et al. (2013) found that instances of overconfidence in physicians, even with the receipt of further information, were especially stark for difficult cases when objective accuracy was very low. In a task that involved diagnosing ultrasound scans, it was found that overconfidence was inversely associated with the amount of clinical experience that the clinicians/participants had (Schoenherr, Waechter & Millington, 2018). However, it has also been found that underconfidence can be more prevalent than overconfidence, especially when comparing medical students to residents (Friedman et al., 2005). Similarly, Yang and Thompson (2010) found that experienced nurses exhibited similar performance to nursing students, but were more confident in their judgements, showing differences in confidence calibration across experience levels. More broadly, highly confident members within a group could unknowingly reduce the chance of less confident members speaking up about potential errors, which is a common problem within healthcare (Hémon et al., 2020). Overconfidence has also been linked to a lower likelihood of sufficient patient management and clinical effort as per a field study in Senegal (Kovacs, Lagarde & Cairns, 2019). The fact these papers cover different medical subdisiciplines and experience levels indicate the broad relevance of confidence calibration across medical decisions.

To summarise, confidence can be thought of as a readout of the evidence/information received in favour or against a particular decision. Past work has not only shown

evidence for miscalibration of confidence (as overconfidence or underconfidence), but also that such miscalibration has an impact on patient treatment. When taking these together, one can infer that the decoupling between confidence and accuracy is linked to the manner in which evidence/information is sought or received. This is pertinent within medicine where doctors have to synthesise an array of patient information (e.g. documents, test results, examinations etc.). In the next section, we review the extant literature on the relationship between information seekind and confidence.

Information Seeking and its link to Confidence

The link between confidence and information seeking has been previously investigated in cognitive psychology research. This association can be studied in two directions, looking at either how the evidence collected subsequently informs confidence or how confidence informs subsequent information seeking. In relation to the former of these, confidence is primarily determined by the accumulated strength of evidence in favour of a particular decision alternative relative to others (Vickers & Packer, 1982). However, it has also been shown that the mere quantity of information, even if some information favours the non-preferred option, increases confidence in of itself (Ko, Feuerriegel, et al., 2022). One potential reason that has been suggested for this is that individuals experience 'decision inertia', whereby they tend to maintain their previous choices regardless of the evidence presented against those choices (Akaishi et al, 2014). Alternatively, past decisions made by an individual may bias any subsequent information sampling, as per a phenomenon commonly referred to as 'confirmation bias' (Nickerson, 1998). Individuals have been to shown to have a tendency to sample information that corresponds with a previous choice, with confidence increasing the extent to which information sampling is biased (Kaanders et al, 2021). The qualities of the information received matters for confidence too however, with more variability in information being associated with lower confidence (Desender, Boldt & Yeung, 2018). Choosing when to stop gathering information has also been found to produce a 'boost' in confidence when compared to being forced to stop gathering information at a certain time (Wei,

2022). In summary, the process of seeking evidence is implicated in the confidence that individuals have in their decisions.

In relation to the latter directionality, such that confidence informs subsequent information seeking, confidence has been found to affect the amount of time spent viewing (Rausch et al., 2018) and deliberating over evidence (Kiani et al., 2014). Confidence also affects the type of evidence that is sought, as it has been found to be associated with a tendency to seek confirmatory evidence (Rollwage et al., 2020) and how likely incoming evidence will be to change one's mind (Hauperich et al., 2021). After a decision is made, we continue to process evidence, meaning that we continue to think about a decision after the decision is made. This means that having 'second thoughts' or changes of mind are more likely with a lower level of initial confidence (and hence a lower strength of evidence). When taken together, confidence and information seeking are deeply connected during the decisional process. Given that the confidence that individuals have in their decision is based on evidence evaluated either in favour or against that decision, systematic differences in information seeking and evaluation could be responsible for differences in confidence calibration. Hence, studying information seeking in the context of medical decisions can elucidate how clinicians' confidence can become decoupled from their objective accuracy. We now look at research on information seeking and confidence during medical decisions.

One of the earliest papers to find evidence of the link between overconfidence and information seeking in clinical settings was by Oskamp (1965). This study focused specifically on clinical psychology and tasked participants with answering questions about a patient who may have been displaying signs of post-traumatic stress disorder caused by the patient's army service. Participants received some information about this former soldier named Joseph Kidd and were asked to answer 25 multiple choice questions about Kidd's past and predicted future behaviour. They finally reported their 'confidence' by estimating the percentage of questions they answered correctly, ranging from 20% (at chance) to 100% (all correct).

Participants then received more information about Kidd in three subsequent stages, focusing on Kidd's childhood, his time in school and his time in the army. After receiving each set of new information, participants could revise their answers to all questions and report their new confidence. Oskamp found that with each new set of information, participants increased their confidence but did not significantly improve their accuracy. In fact, participants were less likely to change their answers as more information was provided (see Figure 2 below). This demonstrated that confidence could be linked to mere receipt of information and that participants were more confident than they should have been.

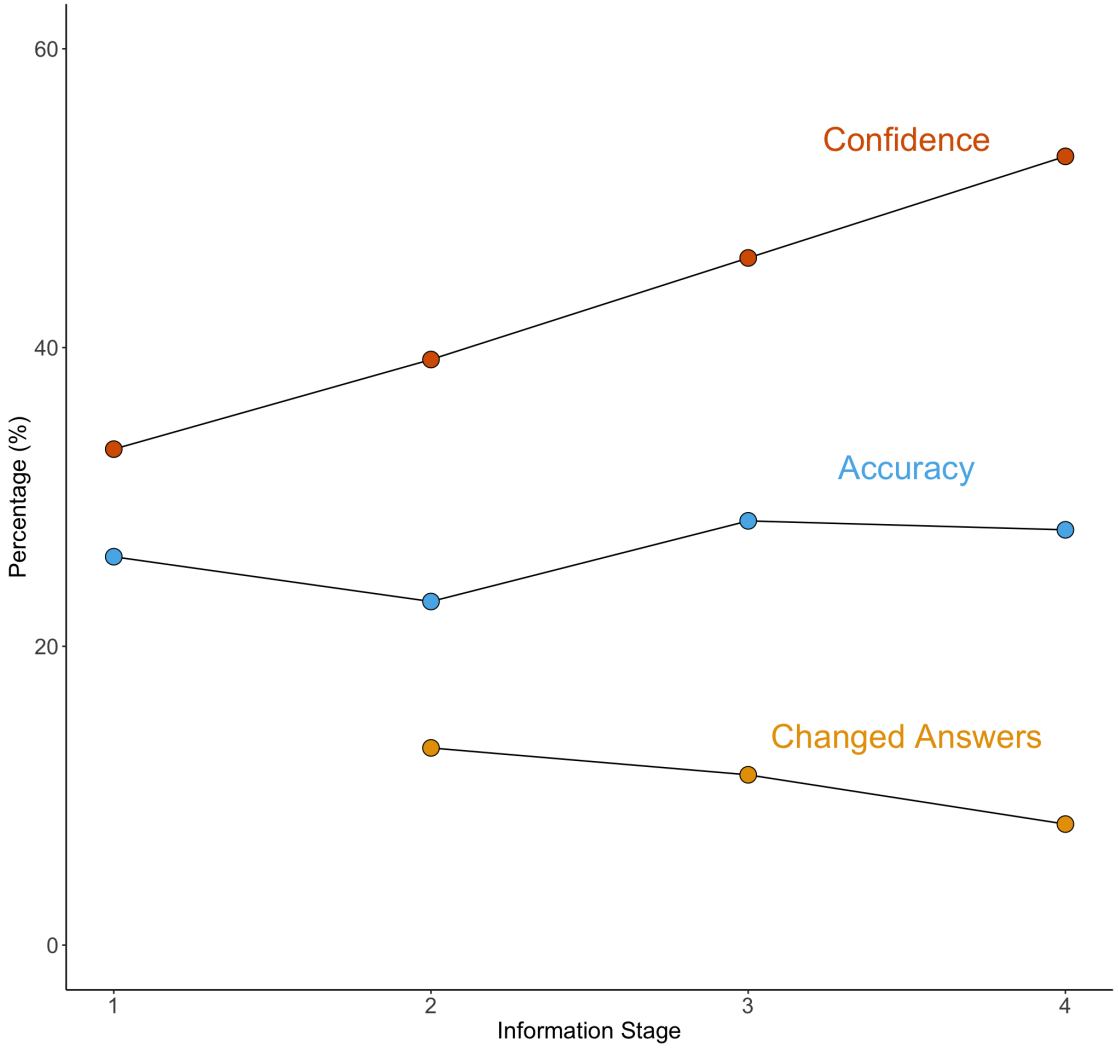

Figure 2: Graph representing the results from Oskamp (1965), which was plotted using the reported data (the original paper did not have such a figure). The graph

shows at each stage of information being provided about the patient (x-axis), the average confidence (y-axis, red), accuracy (y-axis, blue) and proportion of answers that were changed compared to responses during the previous stage (y-axis, blue). This graph, showing that confidence increased with more information whilst accuracy remained at a similar level, is representative of findings from other papers (e.g. Meyer et al., 2013) of overconfidence even with the receipt of further information.

Studying investigating confidence and information seeking in medical diagnosis have revealed similar interactions and imperfections. Notably, Gruppen, Wolf & Billi (1991) found that clinicians were less confident when they had to seek relevant information for themselves compared to all information was already provided, indicating that information seeking as a task is contributory to formulating diagnostic confidence. While this shows the relationship in one direction, past work has also viewed confidence as contributory to further information seeking. Pathologists with more calibrated confidence were found to request more information, such as second opinions or ancillary tests, when unconfident in their judgements (Clayton et al., 2022). In a sample of 118 physicians presented with patient vignettes, it was found that higher confidence was associated with a decreased amount of diagnostic tests being ordered, even if confidence and accuracy were larger decoupled/miscalibrated (Meyer et al., 2013).

It has also been observed previously that physicians may 'distort' neutral or inconclusive evidence to be interpreted as supporting prior beliefs (Kostopolou et al., 2012). Similarly, it has been found that a patient's case history that suggests a particular diagnosis prompts selective interpretation of clinical features that favour the initial diagnosis (Leblanc, Brooks & Norman, 2002). Together, these findings have implications for how clinicians may seek and integrate evidence when making decisions and how patterns of receiving information could affect decision confidence and in turn confidence calibration.

Information seeking has a clear goal during medical diagnoses however: weighing up possible diagnoses. Clinicians generate hypotheses and then gather information to reduce the space of hypotheses. They should ideally eliminate hypotheses from consideration only when it makes sense given the incoming evidence. By the same token, they should also not continue attaching themselves to a hypothesis when there is overwhelming evidence to the contrary. One conclusion of Wason (1960) was that individuals struggle to remove a hypothesis from consideration even if they receive evidence against it. Individuals may look to sample positive evidence for their internally generated hypotheses and then reject alternatives in a decision that were relatively unsampled (Hunt et al, 2016). When taken together with previous papers, hypotheses are potentially evaluated based on the quantity of evidence, rather than the balance of evidence for and against. Understanding how individuals generally reason about a possible space of hypotheses is interesting for understanding how the reasoning process works differently for novices and experts, especially in a specialised domain such as medicine. We would argue then that the information seeking process feeds into hypothesis generation during diagnoses.

There are interesting questions here around how individuals generate hypotheses and then gather information to reduce the space of hypotheses. One should ideally want to eliminate hypotheses from consideration only when it makes sense given incoming evidence. By the same token, they should also not continue attaching themselves to a hypothesis when there is overwhelming evidence to the contrary. Understanding how individuals generally reason (and vary from each other in their approach) about a possible space of hypotheses is interesting for understanding how the clinical reasoning process works. We are particularly interested in how one's 'process of elimination' affects confidence. Contrary to experiments with two or three set alternatives, a lot of real world decisions have a large set of potential options (which depends on the individual's task-specific knowledge in order to generate plausible hypotheses). In theory, individuals gather information in order

to reduce the initial set of potential alternatives to a more manageable list (or in some cases, deciding on a single option). On the one hand, individuals with more domain knowledge will be able to generate a larger set of plausible alternatives (including more 'obscure' or lesser known options), but their knowledge may also allow them to eliminate hypotheses earlier in a decision process based on less information. A question here is how the amount of hypothesis elimination affects information seeking patterns and hence subsequent confidence. If an individual has a larger set of initial hypotheses, this means that the problem space is more complex and potentially harder to whittle down.

Diagnostic decisions have been thought of as 'ideal' when using the hypothetico-deductive process (Kuipers & Kassirer, 1984), whereby hypotheses are formulated based on specific features of a patient and are then linked to established criteria for a diagnosis, with further information gathering to test these hypotheses (Higgs et al., 2008) or eliminate others. This account was challenged by Coderre et al. (2003), who found, via analysis of clinicians' explanations as they worked through diagnostic cases, that more accurate diagnoses were based more on pattern recognition (matching observed information/symptoms to the most likely diagnosis), especially for more experienced clinicians. Either way, the bridge between confidence and information seeking is the reasoning strategy utilised by clinicians. Diagnostic reasoning is currently taught using cognitive frameworks such as the surgical sieve and the ABCDE mneumonic for patient assessment (Airway, Breathing, Circulation, Disability, Exposure). However, current education does not account for differences in reasoning strategies, whether strategies may meaningful vary by case and by clinician and how these strategies have a downstream influence on the diagnostic process in terms of seeking information, generating differentials and formulating confidence.

Current Work

Based on the literature previously presented, we can surmise that there has been considerable work on understanding the sources of diagnostic error. This has led to key insights, such as cognitive biases being implicated in diagnostic decisions and the role of information seeking in diagnostic errors. However, past research has often used simplified diagnostic tasks that do not capture the complex interplay between information seeking and decision making that are evident in real-world decisions, both medical and otherwise. There is evidence from both psychology and medicine that this rich interplay is important to study. As we have explained, lab experiments within the field of psychology have found imperfections and biases in the link between information seeking and confidence, and there is evidence of similar biases in clinicians. There is also evidence that clinicians vary in the strategies they use to seek and integrate information within their consideration of diagnostic hypotheses. Taken together, this implies gaps in our understanding of how clinicians make diagnostic decisions, which has implications for the development of cognitive interventions or educational/training resources on how to make accurate diagnostic decisions. We aim to fill these gaps with research that better represents the complexities of real diagnostic decisions, whilst also studying the cognitive mechanisms of how the decision making evolves over time and as more information is afforded to the clinician.

There is a need for the teaching and assessment of non-technical skills and human factors in healthcare (Higham et al., 2019), which is currently not addressed in a widespread standardised manner in speciality curricula (Grieg, Higham & Vaux, 2015). Curricula within medicine also place little emphasis on how uncertainty is communicated and approached in medical decision making (Hall, 2002). This is especially important to note given that doctors can be reticent to express their uncertainty (Katz, 1984). Clinical experience may also be connected to risk aversion and further information seeking behaviour (Lawton et al., 2019), which offers an

important avenue for future medical education. Hence, this research aims to inform medical education on non-technical skills such as diagnostic reasoning, especially around evaluating diagnostic differentials and seeking information during the diagnosis process. This work also allows for the application of cognitive psychology to a type of important real-world decision, in turn testing the ecological validity of the field's findings and using the field to better understand diagnosis as a complex decision making process. This has been cited as a necessary avenue for research:

"Problems in diagnosis have... been heavily dominated by physicians with little input from the cognitive sciences. What is missing... is foundational work aimed at understanding how clinicians in actual situations take a complex, tangled stream of phenomena... to create an understanding of them as a problem." (Wears, 2014)

Thesis Structure

We now present the structure for the subsequent chapters in this thesis.

In Chapter 2, we present a systematic scoping review of the medical and psychological literature in which confidence or certainty has been studied within diagnostic studies. Whilst some of the extant literature has been presented here, we use this review to describe the existing literature more thoroughly. Our main aim is to look at whether past work has found evidence for calibration of confidence judgements provided during diagnostic decisions. We also look at whether confidence predicts aspects of the patients' care pathway. Alongside these research aims, the review is used to map out the themes of the extant literature and types of methodologies used. We use the reviewed literature to propose a conceptual model for how diagnostic decisions are affected by decisional, clinician-based and contextual factors.

In Chapter 3, we present an online behavioural study with medical students where participants freely sought information and provided diagnostic differentials at

different stages during a series of patient vignettes. This study allows us to study how diagnostic differentials and confidence are affected by patterns of information seeking. In particular, we introduce and investigate different aspects of information seeking, namely how much, how valuable and how variable the information that medical students seek is, with these aspects differentially informing diagnostic confidence and accuracy.

In Chapter 4, we present in-person study using a similar paradigm where medical students think aloud as they are making these diagnoses, with the aim to use these think aloud utterances to classify different diagnostic reasoning strategies. These different strategies are used to investigate how reasoning strategies affect confidence and information seeking. This study also allows for a qualitative analysis of medical students' thought process whilst they are making diagnoses, including their main considerations with regard to evaluating diagnostic hypotheses.

In Chapter 5, we present the third empirical study, where we investigate diagnostic decisions in a more naturalistic manner by using virtual reality paediatric scenarios to investigate differences in information seeking and confidence. This paradigm allows for study of information seeking in an open-ended manner that is more akin to real medical practice. Given the higher fidelity of this paradigm, we also study how the ongoing treatment of patients is informed by confidence and information seeking.

In Chapter 6, we present a reflective chapter based on observations in both Intensive Care and Emergency Medicine, whereby the findings from this DPhil are contextualised within the decisions made during actual medical practice. This chapter is used to better characterise the strengths and limitations of the studies within this DPhil by evaluating how well they relate to aspects of everyday medical decision making, as well as discussing real examples of where diagnostic uncertainty arise.

In Chapter 7, we present an overall discussion of the studies conducted in the context of the current literature on diagnostic decision making. We aim to demonstrate how confidence and information seeking and their cognitive mechanisms have an impact on medical diagnoses. We discuss the implications of this work for both the field of cognitive psychology and medicine (in particular, how our findings can inform future medical education given our focus on medical students). We finally evaluate the generalisability of our work and suggest some future directions for future research and medical practice.