Information Seeking and Confidence in Medical Decision Making

Sriraj Aiyer

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

For my paatis and my thathas

Acknowledgements

Having a PhD thesis written is a weird position to be in. On one hand, this is very much a new beginning in my career and my professional life. On the other hand, reaching this stage feels like the culmination of many years of work and life experiences. I still have a lot to learn, which is very humbling. But getting to the end of something like this really forces me to take stock of where I am at and where I want to be at in future. I'm grateful to have experienced this kind of personal growth over the course of this DPhil, and will take it with me going forward.

It took a village to make this thesis a reality. I would firstly like to thank my amazing supervisors, Nick and Helen, for your insights, enthusiasm and knowledge that helped shape this thesis into what it is. I have learnt so much from you both, and I will be forever grateful for that. Thank you for your patience and for being willing to help me on this journey and for just being good people. I could not have asked for better, more supportive supervisors.

I also want to thank everyone who was from the OxSTaR group, it was wonderful to be part of such a supportive academic and clinical community. Thank you to Wendy, Rosie, James, Olivia, Will and Aditi for all your support. I want to particularly highlight within the group who directly helped with my research, including Nathan, Kapil, Archie, Anil, Kaushila and Laura. Your inputs, feedback and help are so appreciated! And I would like to thank Mo for his hard work. I have also been lucky to be a part of the ACCLab, and to have met so many brilliant, smart folks. There are a lot of names to mention: Maja, Sarah, Aaron, the two Matts, Rebekka, Jasmine, Joshua. Thank you for your candour, your stimulating conversations and your friendship.

I am so grateful to have had friends in so many places I can rely on, especially for the (very needed) emotional support and for laughs! I could not have got this stage without such amazing support systems around me. In Oxford, I am thankful to have made friends for life. Thank you to The Crisps (Riddhi, Nilgoun and Samuel), Soyoung, Olly, Georgie, Malin. It's incredible to have such brilliant, funny and supportive people around me during this time. Outside of Oxford, I want

to thank the Bath Lads (John, Matt, Robert, Gary) for your years of friendship and laughs, you mean a great deal to me.

And to the Foresters and Forester-adjacents! That means you Dom, Alessio, Dan, Ellen, Jesse, Connor, Will, Phil, Edwin, James, Nick and Tim! What a group to be a part of. It may not seem like it, but it's the little things that helped me a lot over the course of this PhD.

Last, and by no mean least, I want to thank my family. My Akka, my Amma and my Appa. I do not have words to express and go through everything that you done for me (that would be another whole thesis to write, and I think that the one thesis is enough). For now, I love you all so very much and you share in this achievement with me, alongside my thathas and paatis, who I am here because of. I miss them every day.

<div align="right">

Sriraj Aiyer
Wolfson College, Oxford
30 September 2024

</div>

Abstract

Decisions within healthcare are unique within the wider realm of decision making. They are often made within high-pressure situations and have severe consequences if done so incorrectly. Hence, they require intensive training and a wide knowledge base for clinical staff to draw from. What is remarkable is that despite the intimidating amount of material for medical students to learn and the pressures that can befall them in their everyday line of work, as well as an ever-expanding understanding of medical conditions, treatment methods and technology to maintain, clinicians frequently make swift and accurate decisions that can have a profound impact on patients' lives. When seeking to apply past research within decision making to an applied context, medicine is an interesting domain to study decision making, especially if findings can inform the training of the newer medical students. In particular, 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). Similarly, curricula within medicine place little emphasis on how uncertainty is communicated and approached in medical decision making (Hall, 2002). Hence, this research looks into non-technical skills such as communication of confidence, management of uncertainty and mental model alignment. Over the course of this thesis, we will look at confidence and information seeking in general decision making and then apply insights from cognitive psychology to the realm of medicine.

Contents

List of Figures

List of Tables

List of Abbreviations

AD . . . . Aortic Dissection

DKA . . . . . . Diabetic Ketoacidosis

GBS . . . . . . Guillain-Barre Syndrome

HD . . . . Hypothetico-Deductive Reasoning

ICU . . . . . . Intensive Care Unit

MTB . . . . . Miliary Tuberculosis

OMS . . . . . . Oxford Medical Simulation

OSF . . . . . . Open Science Framework

PhEx . . . . . Physical Examination

PaHi . . . . . . Patient History

PR . . . . Pattern Recognition

SI . . . . Scheme-Induced Reasoning

TA . . . . Temporal Arteritis

Te . . . . Testing

TTP . . . . . . Thrombotic Thrombocytopenic Purpura

UC . . . . Ulcerative Colitis

VR . . . . Virtual Reality

Chapter 1 - 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 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 causes of this patient's symptoms are and how they 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 becomes severely compressed. 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's condition (i.e. the root causes and contributing factors of a patient's symptoms and risks associated with them) with colleagues (e.g. nurses, specialists) 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?

Introduction

One of the difficulties within this scenario is that the symptoms are indicative of multiple underlying conditions, which is a common occurrence in medical decision making. This example is illustrative of why many medical decisions are 'ill-structured' problems, in that they present several possible courses of action, and can produce disagreements between clinicians over both the current hypothesis for the patient's condition and desired end goal (be they or short-term or long-term) for that patient's care (Jonassen 1997). During this thesis, we will investigate and aim to elucidate 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 care and treatment. Secondly, from a psychological standpoint, it allows for an extension of previous research on the relationship between 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 (van den Berge and Mamede 2013; Norman et al. 2014; Restrepo et al. 2020).

Diagnosis in medicine has been defined as "the science…to distinguish one disease from another and trace symptoms to causes from which they spring" (Fenwick 1891). Identifying the condition that a patient has is important for guiding subsequent treatment. 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 then been grounded in the incidence of errors in order to better ensure safer patient care. Looking into medical errors allows healthcare systems to learn from past mistakes and improve both technical and safety processes for future patients. For instance, the Healthcare Safety Investigation Branch (HSIB) has reviewed patient case studies to guide future diagnosis of aortic dissection (*Investigation report* 2021) and lung cancer (*Investigation report* 2022) among others, due to these case studies exemplifying delayed recognition of these conditions and their negative consequences for patients.

Past work has attempted to quantify the scale of diagnostic errors within healthcare.

Introduction

A report from the US Institute of Medicine 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, Meyer, et al. 2017) found that up to 55.5% of patients experienced a 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 (injuries/harm that were caused by medical management, rather than the underlying disease) (Kohn et al. 2000), whilst the Canadian Adverse Events Study found this value to be 10.5% (Baker et al. 2004). The Quality in Australian Health Care Study found that 20% of adverse events were due to delayed diagnosis (Wilson et al. 1999). Even when 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 in order to identify how to reduce their prevalence. All in all, understanding the common sources of medical errors and adverse events can be extremely valuable for improving healthcare in the future.

With this scale of diagnostic errors in mind, another subset of past work has connected diagnostic errors to clinicians' diagnostic thinking, including how they weigh up diagnostic hypotheses (also known as differentials) against each other and how they seek information to support or rule out these hypotheses. Around 32% of clinical errors have been found to be caused by clinician suboptimal assessment, particularly the clinician's failure to weigh up competing diagnoses (Schiff et al. 2009). Another study estimated that 75% of diagnostic failures could be attributed to failures in clinicians' diagnostic thinking (Thammasitboon and Cutrer 2013), including having inadequate knowledge, faulty gathering of information, and not thoroughly verifying information. In terms of their downstream consequences, diagnostic errors have also been found to lead to longer hospital stays and increased patient mortality (Hautz, Kämmer, et al. 2019). In addition to longer 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 and Hackbarth 2012). There has been increasing emphasis in the research literature on overtesting, such as requesting costly imaging scans when they may not be medically necessary (Carpenter et al. 2015). Salem-Schatz et al. (1990) 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. Taken together, this set of literature illustrates the broad impact of diagnostic error and the amount of work that has gone into understanding its impact on patients.

Diagnostic error is by no means the sole cause of medical incidents. There are several factors tied to the wider work environment, culture and technology that can contribute to incidents and errors. Individuals involved in clinical decision making must frequently contend with an uncertain decision making environment, as well as time pressure and personal stresses (Yates 2020). However, by understanding the individual psychological factors that affect the diagnostic process, we better understand how sociotechnical and environmental factors may interact with and amplify individual contributing factors to diagnostic error. Gaining a greater understanding of the causes of diagnostic error can have important implications for future interventions within healthcare settings and improvement of patient care. Studying diagnosis also has added insights for the field of cognitive psychology, within which decision making has been studied in the past. Such insights can then be applied to the real-world context of medicine with the aim of improving diagnostic decisions. In the next two sections, we describe the extant psychology literature on decision making and how it has been previously applied to medical diagnoses.

Cognitive Biases in Diagnoses

One potential account of diagnostic errors is that they stem from cognitive biases during the diagnostic decision making process. Cognitive biases have been investi-

gated substantially in past psychology research on judgement and decision making. Studies of biases have aimed to elucidate the ways in which the decisions we make are reliant on heuristics that can often contribute to suboptimal or erroneous decisions. These heuristics are believed to be used as 'shortcuts' to make quicker and more automatic decisions. This was originally posited by Kahneman (2011) as part of the dual-system theory of thinking, such that decisions are either made on a fast, intuitive level (System 1) or on a slower, systematic basis (System 2). Biases can include weighting information differently depending on whether they arrive earlier (primacy bias, Sætrevik et al. (2020)) or later (recency bias, Chapman et al. (1996)) in a decision making process. They can also include judging a decision as better when it results in a success rather than failure ceteris paribus (outcome bias, (Baron and Hershey 1988; Aiyer et al. 2023)) and judging members of an outgroup less favourably when compared to an ingroup (ingroup bias, (Tarrant et al. 2012)). The study of decision making biases has been broad and considered a large number of ways in which decision making deviates from what we might consider as optimal or rational.

The dual system theory of reasoning has been proposed as being applicable for diagnostic decisions (Croskerry 2009). According to one review, 77% of studies in the allied health professions found evidence of a cognitive bias (Featherston et al. 2020). Whilst the biases in the previous paragraph have tended to be implicated within System 1 thinking, there are also error-prone approaches that arise from the more analytical, deliberate mode of System 2 thinking: reasoning from a false premise, having inadequate contextual knowledge and being paralysed with indecision are a few notable examples (Croskerry et al. 2014). Making a simple 'debiasing' intervention to aid accurate diagnoses is not a trivial task given that in many cases, decision makers are not aware of their biases (Croskerry 2013). It has been proposed that integrating education on cognitive biases within medical education would reduce diagnostic errors and improve patient safety (Royce et al. 2019). What complicates this picture however is the argument that the use of

heuristics, that may seem biased or suboptimal, can actually be beneficial in many situations where environmental and cognitive constraints are placed upon the decision maker (Gigerenzer 2008), which is common within healthcare. For example, a clinician who has less time with a patient due to other commitments may employ certain heuristics to expedite their decision making process. Hence, it is important to understand the situations in which medical decisions are biased in a manner that increases the risk of errors and situations in which these biases instead reflect good decision making.

While it seems intuitive that classical decision making biases affect those in healthcare too (Restrepo et al. 2020), the empirical evidence of the impact for medical decision making is scant (van den Berge and Mamede 2013). One example from dermatology found examples of satisficing bias (also known as premature closure, whereby clinicians arrive at an incorrect diagnosis too early and subsequently stop considering alternatives or seeking other information) and anchoring bias (whereby individuals are highly influenced by a reference point that other options for a decision are judged against), but few examples of other biases, such as availability bias (overweighting information or symptoms that correspond with a recent or memorable case from a clinician's experience) (Crowley et al. 2013). These biases have been considered to be most applicable to medical decision making based on case studies and compelling anecdotal evidence (Groopman 2010). For many of these biases, it can be challenging to establish a causal empirical relationship such that errors can be shown to be a result of a bias. For example, investigating availability bias requires researchers to simulate a patient situation that clinicians have experienced recently and then show that relating the current situation to that recent case is what caused a diagnostic error (if indeed the clinician does make an error).

The literature on cognitive biases is vast, with a lot of different biases named and defined. In an attempt to condense the large number of defined biases, Oeberst

and Imhoff (2023) categorised several biases from psychology research under a number of groups of beliefs about how individuals process information. For instance, a belief that one's opinion is shared by many others (false consensus effect, (Nickerson 1999)) and the tendency to judge others as similar to oneself (social projection, (Robbins and Krueger 2005)) are both heuristics/biases driven by a common underlining belief: that one's own experience is a reasonable reference to extrapolate others. Similarly, tendencies to believe that successes come from one's self rather than external factors (self-serving bias, (Mullen and Riordan 1988)) and that one's performance is better than others (better-than-average effect, (Alicke and Govorun 2005)) are both underscored by a belief that we are good or high performing at what we do. This latter belief is of particular interest to our present work, as it is related to the estimation of one's own ability. In other words, it can be considered as one's subjective 'confidence'. Confidence is important to study given that it can often not be matched to objective performance and underscores a large amount of medical practice, as we shall explore during the remainder of this section.

Confidence is important to consider within healthcare given the scarcity of clear feedback available to clinicians. In addition, in line with the implications of cognitive biases in diagnoses, one type of bias that has more consistently been revealed in experimental findings within medical decision making (when compared to other aforementioned decision making biases) is overconfidence (Berner and Graber 2008; Meyer et al. 2013). These findings revealed clinicians to report themselves as being more confident in their decisions than they should be given their objective accuracy. When making a diagnosis for a patient, clinicians likely do not receive a lot of feedback about the correctness of their diagnosis. Generally, doctors gather information through tests, patient documentation and other means to generate a model of the patient's condition, through which they can surmise a hypothesis for what could be the underlying cause of a patient's symptoms. Some may view diagnostic tests (eg blood tests) as a form of feedback: doctors use these test results to either reinforce or re-evaluate their prior beliefs. However, tests

are not objective markers of feedback, as they have differing levels of sensitivity and specificity rates, leading to false positives, false negatives or even inconclusive results. Clearer feedback may be available to the clinician based on how a patient's condition changes. A patient's reaction to treatment, and their rate of recovery, can be seen as a form of feedback. This in itself is imperfect however, as patients can deteriorate or improve due to circumstances outside of the doctor's control or awareness. This also requires a patient's condition to develop over time and cannot be known in the moment of a diagnosis. In the absence of clear feedback then, confidence is used as one of the only markers available for how likely someone is to be correct. 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 defined as one's "subjective probability of their decision being correct" (Fleming and Daw 2017), and has been viewed as a way for humans to communicate their thought process to others (as per Kahneman's System 2 mode of thinking) (Shea et al. 2014). Individuals have been shown to be able to evaluate their own decisions without any objective feedback via judgements of confidence (Henmon 1911) and evaluations of their own accuracy (Rabbitt 1968) that correlated well with objective performance. Confident individuals tend to be more influential with others in a group (Zarnoth and Sniezek 1997) and can even causally increase the confidence of other observers (Cheng et al. 2021). This effect has been observed in mock jury trials, during which participants heard eyewitness testimonies presented with high confidence and then perceived those testimonies as more credible than testimonies provided with lower confidence (Cutler et al. 1989; Roediger Iii et al. 2012). Confidence is a commonly used predictor of another person's accuracy, especially when feedback is not readily available on an individual's true accuracy. Confidence also varies across individuals with what may be considered a 'subjective fingerprint' (Ais et al. 2016), meaning that 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 and 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). Having 'second thoughts' or changes of mind are more likely with a lower level of confidence (Charles and Yeung 2019). If an individual is systematically overconfident, they would be less likely to change their mind in the face of evidence that is contradictory to their beliefs (i.e. confirmation bias, Nickerson (1998)).

What can be surmised here is that confidence is important for decision making, as it allows for an indicator of how likely a decision is to be correct. This is especially useful in contexts where objective feedback is scarce, such as in medicine. In the absence of objective feedback, confidence can be used as a marker of how likely someone is to be correct (Price and Stone 2004). Confidence is also important for deciding when a individual commits to a decision or whether more information is needed first before committing. Building on the current research landscape of confidence is then important within medical decision making. If there is an assumption that others will calibrate their confidence to their true accuracy, this would mean that heeding high confidence advice/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 specialty experience, a clinician's confidence is one of the only markers available for other clinicians and for patients during key medical decisions.

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 1.1 below). In experimental studies, confidence sometimes exhibits impressive calibration to objective accuracy (Boldt and Yeung 2015), which is thought to reflect people's ability to evaluate the quality of evidence on which they base their

decisions (Xue et al. 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 et al. 2014), as well as the mood (Rouault, Seow, et al. 2018), personality (Schaefer et al. 2004), gender (Syzmanowicz and 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 et al. 2013). Overconfidence has also been associated with insufficient consideration of reasons to choose alternative options/decisions (Koriat et al. 1980; Scherer et al. 2015). The resulting under- and overconfidence matters: overconfident decision makers leap to premature conclusions and ignore useful information or advice, whilst underconfident decision makers waste time collecting evidence that will not improve their decisions (Desender et al. 2018). Especially in the absence of feedback, decision makers may develop systematically incorrect evaluations of their general performance (Rouault, Dayan, et al. 2019) and their performance in comparison to their peers (Kruger and Dunning 1999). Individual miscalibrations of confidence can also become amplified within groups. Effective decision making in groups 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 and Sniezek 1997). Conversely, underconfident team members may be ignored or may fail to share potentially useful information (Silver et al. 2021).
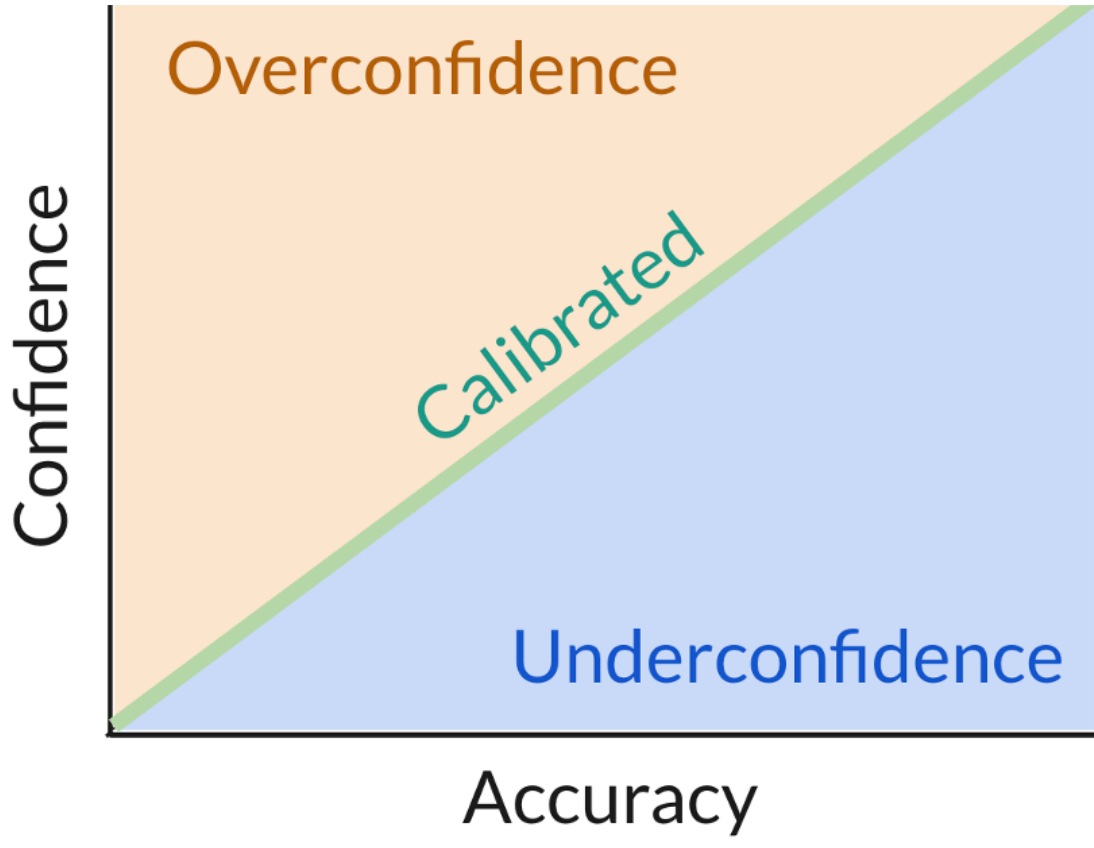
Figure 1.1: Visual representation of confidence calibration when comparing objective accuracy (x-axis) to subjective confidence (y-axis). Confidence is said to be calibrated when the two are relative equivalent (green line). Individuals are considered underconfident when their confidence is lower than their true accuracy (blue area) and overconfident when their confidence is higher than their true accuracy (orange area).

These findings of miscalibrated confidence are important to highlight specifically within healthcare, as overconfidence can lead to insufficient consideration of diagnostic alternatives and inadequate care in terms of seeking appropriate tests and treatment for patients (Kovacs et al. 2020). In medicine, a lack of clearly communicated feedback can cause clinicians to proceed as if they have received positive feedback (also known as diagnostic momentum bias, Aron et al. (2024)). Without clear feedback on whether their decisions are correct, clinicians may not adequately update their internal model of the patient and then increase their

confidence inappropriately, whether working individually or in teams (Jaspan et al. 2022). As we shall explore in the next chapter, the link between confidence and eventual patient care/treatment has been explored in past work, demonstrating the importance of confidence calibration of studying in 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 et al. 2018). However, it has also been found that underconfidence can be more prevalent than overconfidence, especially when comparing medical students to residents (Schoenherr et al. 2018). Similarly, (Yang and Thompson 2010) found that experienced nurses exhibited similar performance to nursing students, but were more confident in their judgements, resulting in differences in confidence calibration across experience levels. However, (Brezis et al. 2019) found that compared to students, experienced physicians were both more confident and less accurate at making a diagnosis for a paediatric case. Similarly, (Friedman, Gatti, Franz, et al. 2005) found that residents were overconfident in their diagnoses on 41% of cases, whilst students were overconfident on 25% of cases. As can be observed from this set of research, past work has tended to focus on drawing out the link between experience and confidence calibration. The overall finding that additional experience as a clinician does not lessen (and could even exacerbate) miscalibration of confidence shows that studying the mechanisms of diagnostic confidence would have benefits for clinicians across all levels of experience. This is especially pertinent in healthcare environments where more experienced clinicians tend to be listened to more. Highly confident members within a group could unknowingly reduce the chance of less confident (or less experienced) members speaking up about potential errors (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 et al. 2020).

To summarise, confidence can be thought of as a readout of the evidence/information received in favour of a particular decision relative to the evidence against that decision. Past work has not only shown evidence for miscalibration of confidence (i.e. overconfidence or underconfidence), but also that such miscalibration has an impact on patient treatment. In sum, one can infer that the decoupling between confidence and accuracy is linked to the way in which evidence/information is sought or received. This is pertinent within medicine where doctors must synthesise a vast array of patient information (e.g. documents, test results, examinations etc.). In addition, the fact the papers covered in this section span different medical subdisiciplines and experience levels indicate the broad relevance of and interest in confidence calibration across the field of medical decision making. In the next section, we review the extant literature on the relationship between information seeking and confidence and how it may provide additional insight into how confidence can become miscalibrated during diagnoses.

Information Seeking and its link to Confidence

The way that individuals seek information is important, as it affects what information they are using to make their decisions. Information seeking is also thought to signal the importance of a task and quality of the source of the information among factors (Xu et al. 2006). Information seeking is an aspect of real-life medicine that should be considered: two clinicians confronted with the same patient case are likely to not use the same information to make a diagnosis if they seek different investigations/examinations. In addition, considering information seeking allow us to conceptualise decision making as an active, ongoing process where information is sought in response to previously seen information. For instance, an individual may be more likely to seek further information when they receive information that

is contrary to, as opposed to being supportive of, their prior beliefs (Adams 1961). Similarly, individuals with staunchly held beliefs have been found to be less likely to seek new information to refine their beliefs (Schulz et al. 2020). Information seeking patterns are also a signal of certainty or uncertainty, especially when a clinician has to decide whether the information they have is adequate (Gehlbach et al. 2024) to make a diagnosis or whether they need more information before coming to a decision. Seeking confirmatory information has been thought of as indicative of calibrated judgements of confidence when information is not processed in a biased way (Rollwage et al. 2020), such as weighting confirmatory information higher than corrective information (Schulz-Hardt et al. 2000). Taken together, information seeking as a research area from cognitive psychology has potential applications for a deeper study within medical diagnoses.

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 and 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 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 (i.e. 'confirmation bias', (Nickerson 1998)). The utilisation of confirmatory information has been found to be especially higher when presenting information sequentially rather than simultaneously (Jonas et al. 2001), which can be the case in healthcare when it comes to requesting tests or examining the patient. Individuals have been 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. 2022). The qualities of the information received matters for confidence too however, with more variability in information being associated with lower confidence (Desender et al. 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 2021). 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 accumulation of perceptual evidence Balsdon et al. (2020), as well as the amount of time spent viewing (Rausch et al. 2020) 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 is to change one's mind (Pescetelli 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 relative 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 the 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 investigate 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 had 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 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 1.2 below). This demonstrated that confidence could be linked to mere receipt of information and that participants were more confident than they should have been given their objective accuracy.
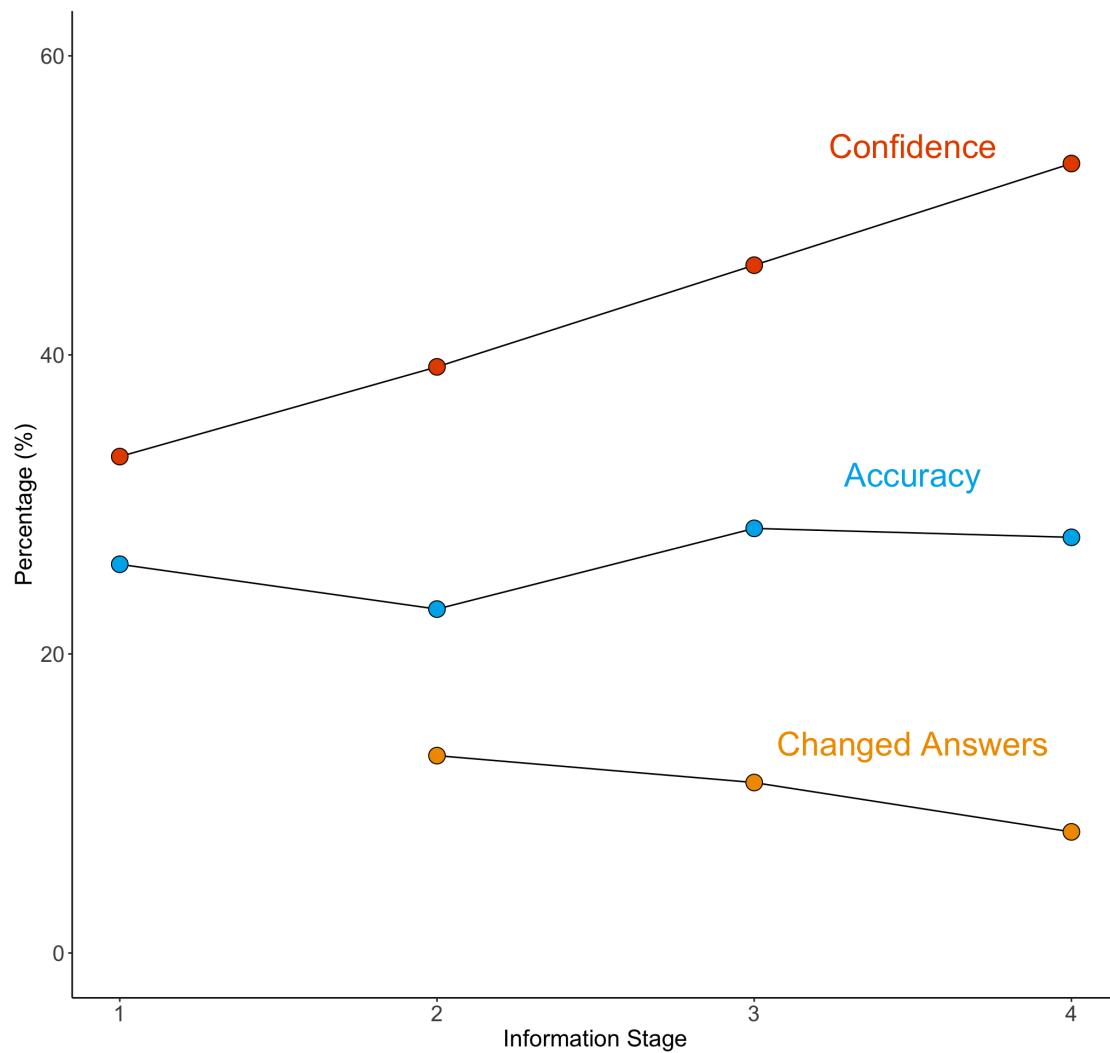
Figure 1.2: Graph representing the results from Oskamp (1965), *which was plotted using the reported data (the original paper did not have such a figure). The graph shows at each stage of information being provided about the patient (x-axis), the average confidence (y-axis, red), accuracy (y-axis, blue) and proportion of answers that were changed from the answers provided during the previous stage (y-axis, orange). 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 with the receipt of further information.*

Studying the interaction between confidence and information seeking in medical

diagnosis has revealed similar interactions and imperfections. Notably, Gruppen et al. (1991) found that clinicians were less confident when they had to seek relevant information for themselves compared to when 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 (and hence less accurate) in their judgements (Clayton et al. 2023). In a sample of 118 physicians who were presented with patient vignettes, it was found that higher confidence was associated with a decreased number of diagnostic tests being ordered, even if confidence and accuracy were largely 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 (Kostopoulou, Russo, 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 this initial diagnosis (Leblanc et al. 2002). Together, these findings have implications for how clinicians may seek and integrate evidence when making diagnostic decisions and how patterns of receiving information could affect decision confidence and in turn confidence calibration.

Evaluation of Diagnostic Hypotheses

Information seeking has a clear goal during medical diagnoses: weighing up possible diagnoses. When making a diagnosis, clinicians generate hypotheses and then gather information to evaluate the likelihood of these 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 be attached to a hypothesis when there is overwhelming evidence to the contrary. One conclusion of

Introduction

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 generated hypotheses and then reject alternative hypotheses that were relatively undersampled (i.e. such that they did not seek as much information to support them) (Hunt et al. 2016). Past work on positive evidence bias predict confidence to be biased towards confirmatory evidence in favour of a chosen decision (Peters, Thesen, et al. 2017). When taken together with the previous papers on information seeking, 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 important 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 consider and maintain multiple hypotheses at once. Past work looking at hypothesis consideration (Robinson and Hastie 1985; Van Wallendael and Hastie 1990) has tended to show that when participants assign probabilities to each hypothesis in their 'list' that they are considering, probabilities are considered for each hypothesis in turn. This leads to situations where the sum of probabilities of all hypotheses exceeds 1, indicating a poor grasp of probability theory but perhaps a more realistic depiction of how individuals consider a set of hypotheses. Individuals may not be maintaining a set of hypotheses in their mind that they modify as they receive information, but instead focus on one at a time. Hypothesis generation and information seeking are linked together, as has been found in one study, in which individuals were more willing to integrate information that conflicted with a prior belief when they had already considered an alternative hypothesis that could explain said information (Vallee-Tourangeau et al. 2000).

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' (if clinicians do practice this) affects confidence. Contrary to experiments that prompt a set of two or three alternatives (Meyer et al. 2013; Chartan et al. 2019; Küper et al. 2024), 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 set (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 also allows them to eliminate hypotheses earlier in a decision process based on less information. A question here is how the amount of hypothesis elimination is related to information seeking patterns and 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 and 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. 2019). This account was challenged by Coderre et al. (2003), who found, via analysis of clinicians' verbal 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. (Gilhooly 1990) proposed that both novice and expert clinicians used a mixture of reasoning strategies to arrive at a diagnosis. Either way, the bridge between confidence and information seeking could be considered as a function of the reasoning strategy utilised by clinicians. The reasoning strategy used impacts how many diagnoses are considered at once, how much information

is needed to increase or decrease the number of diagnoses being considered and, in turn, how confidence changes as a function of the information received. For example, a clinician may consider many diagnoses to start with, require more information to eliminate the diagnoses being considered to form a more manageable set, and by eliminating more diagnoses, they increase their confidence to signal how much the information has refined their thinking. Diagnostic reasoning is currently taught using cognitive frameworks such as the surgical sieve (which prompts clinicians to individually consider each pathophysiological system in turn as a possible cause of the patient's condition) and the ABCDE mneumonic for patient assessment (Airway, Breathing, Circulation, Disability, Exposure). However, current education does not teach differences in reasoning strategies, whether strategies may vary meaningfully by case or by clinician and how these strategies have a downstream influence on the diagnostic process in terms of seeking information, generating differentials and formulating confidence. Making clinicians and medical students more aware of their own reasoning processes can be useful for addressing cognitive errors in seeking and integrating information (Nendaz and Perrier 2012).

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 on a cognitive level, 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. In particular, we aim to study diagnosis not as a single decision but as a more continuous 'practice' of information interpretation, hypothesis evaluation and choosing when to commit to a particular course of action (Alby et al. 2015).

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 medical curricula (Greig et al. 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. Uncertainty within medicine can stem from many sources, be they cognitive, emotional or behavioural (Han et al. 2011; Lee et al. 2021), and the ongoing tolerance of uncertainty is an important skill that medical students develop but also is a source of stress for them (Hancock and Mattick 2020; Stephens et al. 2021). If we can understand the mechanisms by which uncertainty arises during medical decision making, this can be beneficial both for patients and for clinicians.

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 important real-world decisions, 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 considered 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)

Over the course of thesis, we aim to elucidate the cognitive mechanisms that underpin medical diagnoses. In particular, we aim to gain a better understanding of how clinicians seek and interpret the information they receive pertaining to patients and translate their understanding into a set of diagnoses and their subjective confidence. Clinical reasoning is taught as a non-technical skill to medical students, but the field is currently lacking the input from cognitive sciences on how to foster accurate decision making and appropriate expressions of uncertainty. By emulating the diagnostic decision process, we aim to formulate some practical implications for medical education and clinical practice. We also aim to evaluate the methods from cognitive psychology and their applicability to everyday decisions that have significant real-world impact.

Thesis Structure

In this chapter, we presented an overview of past work studying medical diagnoses. In particular, we underscored the need to study their cognitive mechanisms due to the larger incidence of diagnostic errors and their impact on patient care. We laid out research that sought to draw a causal link between cognitive biases and diagnostic errors, with a particular focus on overconfidence/underconfidence. We outlined the importance of confidence from a cognitive psychology perspective, particularly in the field of medicine where objective feedback that could correct biases in confidence is scarce. Finally, we explored the link between information seeking and confidence, which has been investigated in cognitive psychology but only to a limited degree in the context of medical diagnoses. To this end, we orient our current work on further elucidating this link in diagnostic decisions. The overarching hypothesis of our work is that miscalibrations of confidence (when compared to objective accuracy) are caused by clinicians' suboptimal information seeking. For instance, a clinician may cease seeking information and decide on a diagnosis too early (known as 'premature closure'). Alternatively, a clinician may preferentially seek more information to confirm their beliefs, inflating their subjective confidence. We also aim to understand how information seeking relates to the breadth of diagnoses being considered by clinicians. By developing our understanding of how information seeking relates to diagnostic uncertainty, we can present implications for future work on how to prompt appropriate information seeking, and in turn appropriate diagnostic confidence.

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 decisions. Whilst some of the extant literature has been presented here, we use this

review to describe the existing literature more thoroughly. This is especially pertinent given the broad relevance of diagnostic confidence across medicine alongside the siloed nature of literature within the various medical specialties, necessitating a more systematic approach to capturing the relevant literature. The main aim of this review is to look at whether past work has found evidence for calibration of confidence judgements provided during diagnostic decisions. We also look at how 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 where participants (medical students) freely sought information and provided diagnostic differentials at different stages during a series of patient vignettes. This study allows us to look at 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 an in-person study using a similar vignette-based paradigm where medical students verbalised their thought process as they were making these diagnoses, with the aim to use these think-aloud utterances to classify different diagnostic reasoning strategies. We use this paradigm 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 regarding the evaluation of 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 the 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 Adult 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 arises. This section acts as a precursor to our General Discussion section, such that evaluations of this DPhil's ecological validity and generalisability can be grounded in everyday medical practice.

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 the cognitive mechanisms of confidence and information seeking impact 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 directions for future research (within both cognitive psychology and medical decision making) and medical practice.

Chapter 2 - Systematic Scoping Review on Confidence and Certainty in Diagnoses

Introduction

There is extensive evidence of diagnostic error in most healthcare specialities. As we discussed in the Introduction section, overconfidence is an important source of medical error. It has been suggested that cognitive biases, such as overconfidence, are causally linked with these errors. To our knowledge, there has not been existing work to synthesise past research on diagnostic confidence. Such a review would enable us to understand the factors that contribute to clinicians' confidence in their diagnoses, as well as how diagnostic confidence affects treatment and patient care.

In this chapter, we present a systematic scoping review to collate and synthesise the existing literature studying diagnosis as a cognitive process. To our knowledge, this is the first scoping review with such a remit to include studies of confidence across medical subdisciplines, given its broad importance across medicine. The use of scoping review is suitable given this remit. We aimed to identify key determinants of confidence and characterise how judgements of confidence affect the wider medical decision making process. For the purposes of this review, and in common with practice in the reviewed articles, we treat "confidence" and "certainty" as interchangeable terms, while noting that the psychological literature discusses whether they are subtly different concepts (Pouget et al. 2016).

We systematically searched SCOPUS, MEDLINE, PsycINFO and Global Health.

Scoping Review

Articles were categorised according to methodology and clinical speciality. Findings were analysed thematically. Our review methodology adheres to the JBI's PRISMA-ScR Checklist for Scoping Reviews. We then performed citation tracking within these papers' references to identify additional articles. Papers were included if they reported quantitative results from an empirical study in which participants reported their confidence or certainty during a diagnostic decision. Studies comprised several medical subdisciplines. 79 articles met the inclusion criteria.

We organised articles under two main themes: the determinants of confidence and the uses of confidence during the patient's care pathway. Firstly, we find little evidence that clinicians' confidence is aligned to their diagnostic accuracy, even when using certain cognitive interventions or aids. Confidence is found to be affected by several factors including case complexity, early diagnostic differentials, and the healthcare environment. Factors that affect confidence, but not accuracy, demonstrate how the two can become decoupled, resulting in overconfidence/underconfidence. Secondly, confidence is predictive of actions in many parts of the patient care process, such as ordering investigations, referrals to specialists or prescribing, which may be suboptimal if confidence is miscalibrated.

Based on the literature from this review, we propose a theoretical model of factors that affect diagnostic confidence/certainty and accuracy throughout various stages of the patient care process. The proposed conceptual model highlights our current understanding of diagnostic confidence and how future research might focus on underexplored areas, particularly on group decisions, individual differences in confidence, and on the link between information seeking and confidence. Such a model can inform future work on how appropriate diagnostic confidence can be prompted and communicated amongst clinicians. Improving the calibration of confidence should hence be a priority for medical education and clinical practice (e.g., via decision aids), with this model acting as a starting point for future work to target specific factors that contribute to misclibrated confidence.

Methods

Search Strategy

Our review protocol was preregistered on the Open Science Framework: `https://osf.io/wz5se`. We conducted a systematic scoping review of empirical studies on confidence and certainty in medical diagnosis using JBI's PRISMA-ScR Checklist for Scoping Reviews (Peters, Godfrey, et al. 2024). The search strategy was designed in cooperation with a subject specialist librarian at the University of Oxford's Bodleian Libraries group. The search string comprised keywords that captured the intersection of four elements: confidence/certainty, medical diagnoses, decision making and a study population of medical staff/students (i.e., clinicians, physicians, doctors and medics). The full search terms were as follows:

(clinicians OR physicians OR doctors OR medics) AND (confiden* OR uncertain* OR certain*) *AND* (diagnosis AND medical) *AND (decision OR (decision AND making) OR decision-making)*

The databases SCOPUS, MEDLINE, PsycINFO and Global Health were searched during February 2024. Finally, we hand-searched the citations of the included articles from these databases for further relevant articles via backward and forward scanning (Webster and Watson 2002; Tranfield et al. 2003).

Study Selection

The inclusion criteria for screened studies were as follows:

1. Studies must report original empirical work with quantitative results.
2. Studies must be written in the English language.
3. Studies must use an experimental paradigm with medical diagnostic decisions
4. Confidence or certainty in diagnoses must be measured as a dependent variable

5. Studies can be from any medical discipline.

6. Editorials, review papers and opinion papers are all excluded

Identified articles were uploaded onto Rayyan (https://rayyan.ai/) to detect duplicate papers for manual checking and removal. This is the only part of the review process in which an automation tool is used. Data from the papers was collected using manual review. Research characteristics were derived iteratively and can be found on OSF (https://osf.io/4g6s3/).

Research Synthesis

Papers selected for review were first categorised by their broad research methodology (e.g., patient vignettes, in situ questionnaires, etc.) and their medical population of study (e.g., medical students, general practitioners/hospital physicians etc.). We reviewed the experimental procedures to extract their key manipulations and independent variables (e.g., case complexity, use of a cognitive intervention, level of medical experience). We also extracted dependent variables as they pertain to confidence/certainty and, where relevant, recording of both diagnostic differentials and information seeking. Each of the paper's key findings were summarised and then all findings were categorised under recurring themes.

Results

Findings of Scoping Review

The initial search returned a total of 3,332 articles. Applying the inclusion criteria identified 50 eligible articles. 439 further articles were retrieved for review from the included articles' citations. After applying both exclusions of duplicates and our inclusion criteria, 29 further articles were identified. This produced a total of 79 articles for inclusion and synthesis (see Figure 2.1 below for PRISMA diagram). The full set of papers can be found in Table S1 of the Supplemental Materials.
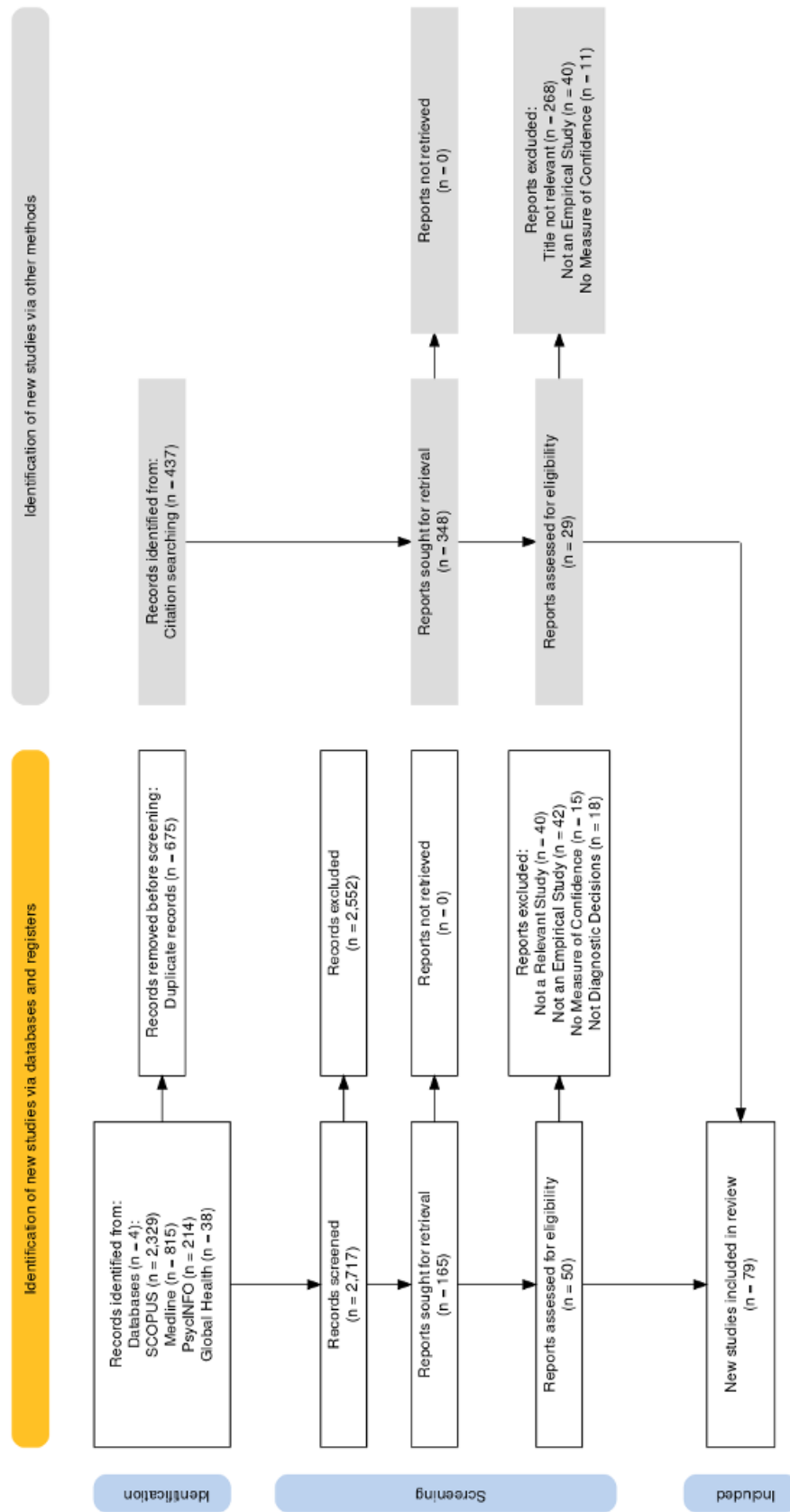
Figure 2.1: PRISMA Diagram of Literature Review.

Study Characteristics

Table 1.1 below summarises study characteristics and Figure 2.2 shows that 36 of the 79 studies (46%) were published since 2019, indicating a recent surge of research interest in this field and the timeliness of this scoping review. The studies appeared in 59 different publications, including both medical and psychological journals, with medical education journals being most common (19 studies). Clinical areas most represented were Primary Care/General Practice, Emergency Medicine and Nursing.

Publication Year		Subdiscipline / Population	
1991-2000	10	Primary Care / General Practice	26
2001-2010	11	Medical Students	15
2011-2020	29	Emergency Medicine	10
2021-	29	Nursing	6
		Pathology	4
Study Environment/Context		Radiology	4
Text Vignette	44	Other	14
Imaging Interpretation (e.g. ECG)	20		
In Situ Questionnaires/Surveys	13	**Study Population Sample Size**	
High-Fidelity Simulation	2	< 100	44
		101-200	20
Participant Experience Levels		201-300	9
Fixed Across Participants	50	> 300	6
Multiple Experience Groups	29	**Total**	**79**

Table 1.1: Characteristics of Included Studies, including year of publication, study environment used and medical population (recruiting single or multiple levels of participant experience, medical subdiscipline, sample size). A full list of all included papers can be found in the Supplemental Materials (Table S1).
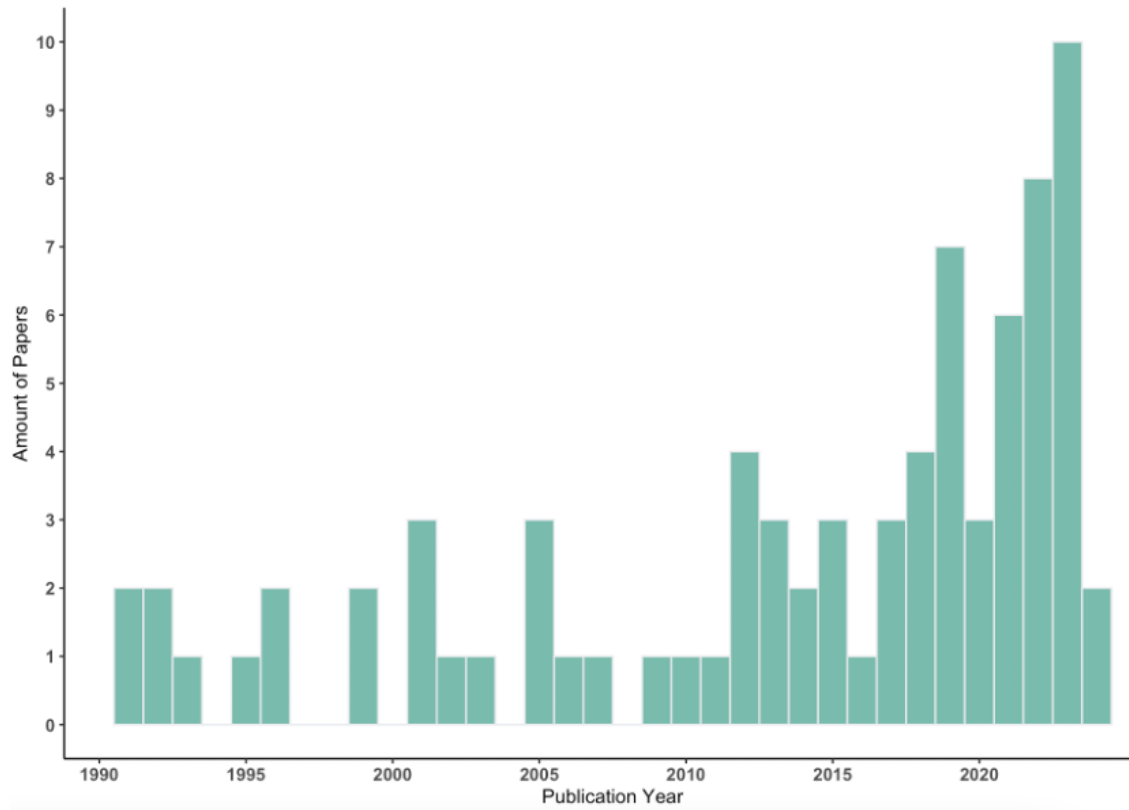
Figure 2.2: Distribution of Papers by Publication Year.

Study designs were split roughly evenly between focusing on how confidence varies across individuals (35 studies) and on how confidence varies according to features of the patient case (31 studies), with remaining (13 studies) studying both. Thirty-four studies (43%) looked at the level of medical experience or training's effect on confidence, either measured as a dependent variable or by recruiting participants in 'novice' versus 'experienced' group. Nineteen studies (24%) manipulated the complexity or difficulty of the patient case. Finally, ten studies (13%) investigated how diagnostic confidence varies with the information presented or the opportunity to seek information.

Most of the studies (44 (56%)) used clinical patient text vignettes. For vignettes, there is an established ground truth in each case (which may not be possible for in situ studies involving real patients) to compare the participants' confidence to

their true accuracy in order to gauge calibration. Because vignettes are quick and simple to administer, participants can complete several diagnoses during a single study such that both their confidence and accuracy can be averaged across cases. Other experimental methodologies include the use of imaging (e.g., ECG, X-Rays, MRI) for diagnosis, high-fidelity simulations (either using extended reality tools or a patient mannequin), or questionnaires administered in situ to measure confidence during real patient cases as they are happening (Table 1). The preponderance of vignette studies is noteworthy given the finding from one study that nurses were both less accurate and less confident in a high-fidelity simulation compared to a paper-based vignette (Yang, Thompson, and Bland 2012), suggesting the need for caution when generalising experimental findings to how medical professionals behave in their everyday practice.

Studies varied in how confidence and diagnostic accuracy were assessed. Studies mostly used a self-reported scale for confidence (usually 1-10 or 1-100) as opposed to verbal expressions of confidence (e.g., "not sure" to "certain") or visual analogue scales. The use of self-report numerical scales is common within cognitive psychology, where measured confidence values predict other behavioural indices of uncertainty, such as the tendency to seek further information or to opt out of making a decision (Gherman and Philiastides 2015). Twenty-four studies (30%) allowed participants to input multiple diagnostic differentials rather than a single diagnosis. Confidence is then either measured for each differential or in the set of differentials as a whole.

In terms of accuracy, most studies prompt clinicians for a single diagnosis that is marked as correct or incorrect. However, clinicians may consider multiple possible diagnoses in their everyday practice. Hence, 24 studies (30%) allowed participants to record multiple differentials in their diagnosis. This complicates scoring accuracy and confidence: If accuracy is operationalised as whether a correct diagnosis is included in this set of multiple differentials, clinicians are more likely to be

correct with more differentials, and it may remain unclear how clinicians weigh up competing differentials. Hence, the operationalised calibration of confidence judgements is heavily contingent on how diagnoses are recorded.

Emerging Research Themes

Miscalibration of Confidence and Certainty Judgements to Objective Accuracy Calibration is assessed by comparing confidence ratings with objective diagnostic accuracy: When clinicians rate 100% (or 50% or 60%, etc.) certainty in their diagnosis, are they in fact correct 100% (or 50%, 60%, etc.) of the time? Calibration is then an indirect measure that is calculated by comparing two other observed measures: confidence and accuracy. In our study sample, there was limited evidence of calibrated confidence judgements, with some studies reporting underconfidence (Mann 1993; Yang, Thompson, and Bland 2012; Brezis et al. 2019)and others overconfidence (Friedman, Gatti, Elstein, et al. 2001; Fernández-Aguilar et al. 2022; Garbayo et al. 2023). To examine these findings in more detail, we considered factors that impact and promote calibration in diagnoses.

The Impact of Experience on Calibration The first major theme of interest is whether calibration improves with experience. This was not always observed in the results (Yang and Thompson 2010; Clayton et al. 2023). However, experienced clinicians seem better able to identify when a case is more complex and adjust their confidence accordingly (Tabak et al. 1996; Brannon and Carson 2003). Looking at the link between calibration and experience alone may be too simplistic, and there are other aspects of experience that influence diagnoses. Experienced clinicians were found to be less likely to 'distort' neutral information to be in support of their reported diagnoses, indicating a lower tendency toward confirmation bias (Kostopoulou, Russo, et al. 2012). Past work has also suggested a distinction between experience (operationalised as years of experience or role seniority) and knowledge (measured using standardised tests of medical knowledge). In medical students, the calibration of confidence judgements were found to improve with years

of education but not with medical knowledge (Hautz, Kämmer, et al. 2019). Information 'distortion' was found to affect novice clinicians more (Kostopoulou, Russo, et al. 2012) and lower knowledge was found to be related to higher susceptibility to irrelevant, distracting features of a patient (Mamede et al. 2024), However, the latter study from Mamede et al. (2024) found that medical knowledge on the part of resident physicians was not directly associated with calibration.

The Impact of Contextual Factors on Calibration   The second major theme relates to contextual and environmental factors. Studies have found that calibration is affected by the complexity or difficulty of the presented case (Meyer et al. 2013; Hausmann et al. 2019; Li et al. 2023). When confidence judgements are not sensitive to the difficulty or complexity of the case, confidence stays fairly constant for difficult cases whilst accuracy decreases, leading to increased overconfidence (and decreased calibration) (Meyer et al. 2013). In past studies, complexity is manipulated by either presenting patient cases with more comorbid conditions (Hausmann et al. 2019) or by showing conflicting information about the patient to indicate multiple possible conditions (Yang and Thompson 2010). Calibration can be improved by the presence of feedback during a training period (Kuhn, van den Berg, et al. 2022; Staal, Katarya, et al. 2024).

Contextual factors that pertain to the situated medical environment can also affect overall levels of confidence (rather than calibration), as found using naturalistic paradigms. For example, clinicians may be constantly interrupted by other tasks (Soares et al. 2019), especially during busier shifts where they have to manage more patients (Gupta et al. 2023) and may not be present for the sharing of information during handovers (Bergl et al. 2024). Studies that simulated these situations found they resulted in lower diagnostic confidence. However, such studies cannot assess the effect of contextual factors on confidence calibration with diagnostic accuracy because they were conducted in-situ, meaning that researchers do not yet have a ground truth of the patient's condition. At this stage, we can only determine how these contextual factors affect confidence, rather than calibration.

Imaging and Decision Support Systems   A subset of papers found evidence for an increase in confidence when providing clinicians with specialised imaging for a patient to assist diagnoses, be they MRI scans (Mackenzie et al. 1996; Albrechtsen et al. 2022), CT scans (Abujudeh et al. 2011), evacuation proctography (Harvey et al. 1999) or photos of wounds (Sanger et al. 2017). Another subset of papers used various forms of computer-aided decision support systems with the goal of increasing confidence (Hillson et al. 1995; Berner and Maisiak 1999; Dreiseitl and Binder 2005; Neugebauer et al. 2020). Hillson et al. (1995) found that the adoption of diagnoses that were recommended by a computer-aided decision support system was not associated with an increase in confidence. Neugebauer et al. (2020) did find evidence for such an association, however, with use of decision support leading to both increased confidence and increased diagnostic accuracy when compared to diagnoses made without using the system. On the other hand, both Berner and Maisiak (1999) and Dreiseitl and Binder (2005) found that usage of decision support recommendations were associated with lower confidence when compared with decisions in which such recommendations were not utilised. Taken as a whole, whilst useful imaging increases confidence when available to clinicians, the efficacy of decision support systems at increasing confidence is likely dependent on other factors that require future work to elucidate.

Interventions at the Point of Generating Differentials   The process of generating diagnostic differentials has been subject to experimental manipulations and interventions (such as early diagnostic suggestions) to investigate their effect on accuracy and confidence. This work is applicable, for instance, to understanding how a clinician transitions care of a patient to another clinician and gives a handover of relevant information. A general theme of this work is that there is a tendency toward higher weighting of early information. Early diagnostic suggestions have been found to be highly influential in the subsequent decision process where clinicians find these suggestions difficult to ignore and have more confidence in them (Kämmer, Schauber, et al. 2021; Kourtidis et al. 2022). This also affects

the breadth of differentials considered, with fewer differentials considered when provided with early suggestions (Kourtidis et al. 2022; Staal, Speelman, et al. 2022) and an underweighting of differentials if they were considered later in the diagnostic process (Eva 2001). Interventions aimed at mitigating this tendency by asking clinicians to explicitly consider alternatives, increased their accuracy and calibration (Feyzi-Behnagh et al. 2014), or prompting the consideration of the patient's 'red flags' in diagnoses, which increased confidence on simpler cases but not accuracy (Chartan et al. 2019). These interventions seem to require explicit instructions: Simply asking clinicians to reflect on their decision without guidance (Lambe et al. 2018; Costa Filho et al. 2019) or participate in an educational training course (Benvenuto-Andrade et al. 2006; Kuhn, Mamede, et al. 2023) does not seem to improve diagnostic accuracy and calibration.

Studies have also investigated how confidence is affected by the manner in which information is presented to clinicians during the diagnostic process. Higher confidence was found when clinicians were presented with additional patient information even when this did not carry diagnostic value (Heller et al. 1992) and when given all available patient information rather than having to gather information themselves (Gruppen et al. 1991). Clinicians were also found to be more confident and more accurate when presented with an Electronic Health Record of the patient alongside other information (Ben-Assuli et al. 2015) and when presented with the patient history first rather than out of order (Tio et al. 2022). This finding indicates that complete patient history available early on has a positive impact on confidence. However, an erroneous patient history has also been found to cue both novice and experienced clinicians to incorrect diagnoses whilst confidence remained relatively high, resulting in overconfidence (Fawver et al. 2020).

Uses of Confidence   With more naturalistic studies, it is possible to isolate ways in which confidence is utilised within the wider diagnostic process, especially where healthcare involves transitions of care between multiple clinicians and departments. Past work has attempted to establish a link between confidence and further seeking

of patient information and tests, with mixed results. US hospitalists (medical staff who provide care for patients specifically within US hospitals) with lower confidence were found to order more tests (Gupta et al. 2023) whilst pathologists who were better calibrated (i.e., who tended to report confidence judgements that were closer to their true accuracy) were found to be more likely to request further tests when they were unsure (Clayton et al. 2023). Confidence has also been linked to prescribing medication, though overtreatment with unnecessary medications has been linked to both underconfidence (Levin et al. 2012) and overconfidence (Garbayo et al. 2023). Higher confidence has also been linked to more referrals to specialists in other departments (Calman et al. 1992) and to a lower willingness to admit mistakes (Brezis et al. 2019). One study found that whilst experienced clinicians were not more accurate in their initial diagnoses, they were more willing to change diagnoses and request more information (Krupat et al. 2017). Lower confidence has been found to result in less specific diagnoses for patients in situ (Hageman et al. 2013). Although psychology research on confidence has examined its role within groups (as discussed in the Introduction), only one article looked at confidence in group decisions in medicine. This study found that a multidisciplinary panel was more confident and better calibrated than a single clinician (Thorlacius-Ussing et al. 2021).

Conceptual Model for Diagnostic Decisions

We synthesised the reviewed findings into a theoretical model (Figure 2.3) that illustrates how various factors distinctly impact diagnostic confidence and accuracy. This model aims to clarify existing research and identify directions for future work. The model starts by mapping out the stages of the diagnostic process (Figure 2.3, bottom panel). Based on initial patient presentation, clinicians gather and interpret patient information (e.g. history, examinations, tests) to inform their diagnosis of the patient's condition. The clinician's confidence in their diagnosis guides their judgment on when they have enough information to begin treatment versus whether further tests or additional information are needed. Once a diagnosis

is reached, this guides patient treatment and care, the success of which is evident in the outcome for the patient.

The middle panel of Figure 2.3 characterises the cognitive processes of the clinician that determine the accuracy of the diagnosis and confidence with which the diagnosis is made. A key feature is that many factors have dissociable effects on accuracy vs. confidence. Diagnostic accuracy depends more on the level of medical knowledge and the quality of information gathering and interpretation; confidence depends more on years of medical experience and the quantity of information gathered (Kostopoulou, Russo, et al. 2012; Hautz, Schubert, et al. 2019; Mamede et al. 2024). Knowledge is improved through feedback on how a patient case was handled and its outcome, which in turn improves future diagnostic accuracy (though this feedback loop's impact on later confidence is yet to be explored).

The top panel of Figure 2.3 highlights factors pertaining to the medical environment/context. Separately from the patient case, confidence is reduced by time pressures (Yang, Thompson, and Bland 2012), interruptions to work (Soares et al. 2019), busy shifts (Gupta et al. 2023) and complex patient cases (either due to conflicting information or comorbidities) (Meyer et al. 2013; Hausmann et al. 2019; Li et al. 2023).

Figure 2.3 highlights three primary directions for future research. First, given the focus of research to date on diagnosis by individual clinicians, we recommend that future work also studies diagnoses in groups, given that diagnoses are often made by teams rather than individuals, particularly in secondary care settings. This is especially pertinent given the social influence that experience/seniority can have within a group: junior clinicians may be less likely to speak up about potential errors in the presence of more experienced clinicians (Hémon et al. 2020). Second, future work should study individual differences on the part of clinicians to characterise how personality and trait level factors impact diagnostic confidence. In particular, factors such as personality (Schaefer et al. 2004), gender (Syzmanowicz and Furnham 2011) and status (See et al. 2011) may impact a clinician's confidence in their diagnoses. Finally, we recommend future work investigate the association

between the ongoing receipt of information and confidence. We recommend future work on the implications of diagnosis as a dynamic process where confidence and information seeking interact. Past work has tended to frame information seeking as a further action after diagnosis, rather than information seeking as a process that forms the diagnosis in the first place. Future work should prioritise examining how to prompt appropriate information seeking (i.e., neither overtesting nor undertesting) via educational tools or cognitive interventions.
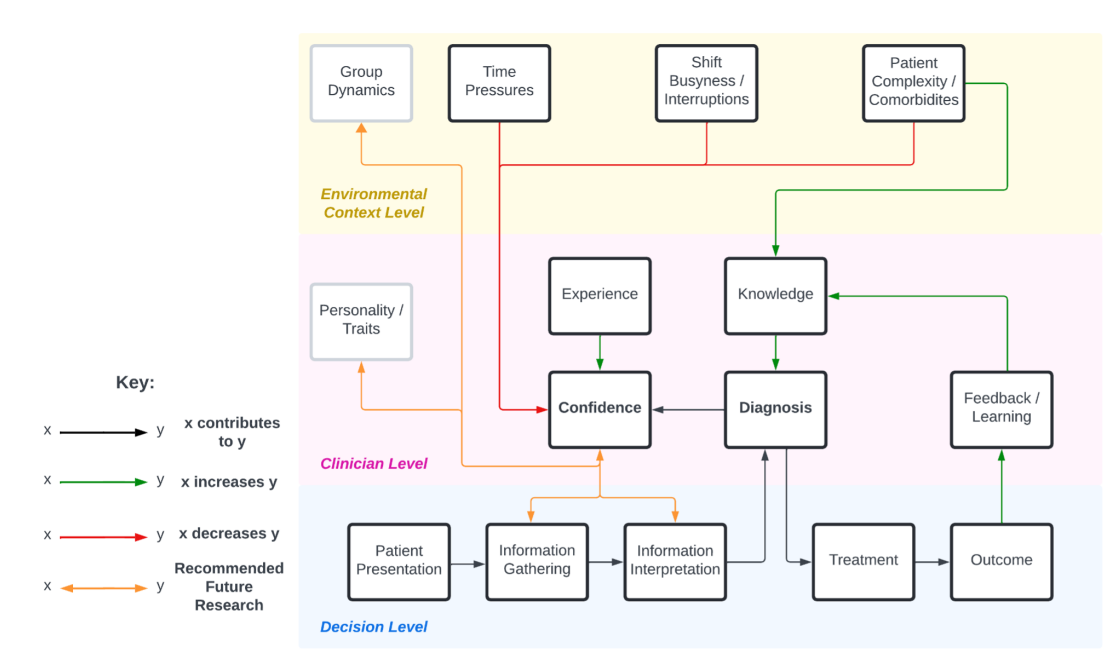
Figure 2.3: Conceptual model that depicts the various factors that impact the course of a diagnostic process, with links established between concepts based on findings from this systematic scoping review. Factors are categorised in three levels: the level of the diagnostic decision process (bottom box, blue, where the course of the decision proceeds from left to right), the level of the clinician (middle box, pink) and the level of the environmental context within which the clinician operates (top box, yellow). Black arrows represent a progression from one concept to another. Green arrows indicate positive impacts between concepts; red arrows indicate the opposite (i.e. a negative relationship). Orange arrows represent links between concepts that are areas for future research. Light grey boxes represent factors that are known

*to affect decisions and confidence within the psychology literature but are currently
less understood in the context of medical decisions.*

Discussion

Review Findings

The present work comprehensively maps out the literature on confidence in medical diagnoses, thus extending previous work exploring how cognitive biases contribute to medical error (Kostopoulou, Delaney, et al. 2008; Graber et al. 2012; Saposnik et al. 2016) and characterising medical uncertainty more broadly (Hall 2002; Bhise, Rajan, et al. 2018). This scoping review shows the importance of, and the recent surge in interest in, diagnostic confidence. Although confidence has been linked to diagnostic error in the past (Berner and Graber 2008), full understanding will benefit from leveraging insights from cognitive psychology to inform medical education and practice (Wears 2014). How clinicians evaluate their decisions contributes to their effectiveness: An overconfident clinician may overlook diagnostic possibilities, delay treatment or ignore crucial information. Conversely, an underconfident clinician may be less likely to speak up in a group about potential errors (Hémon et al. 2020).

Our review finds that confidence and accuracy are rarely aligned during diagnoses. Notably, miscalibration of confidence is not only a function of social and environmental factors: Miscalibration is consistently observed in vignette studies performed by individual participants, where decision makers tend to be overconfident particularly when dealing with complex cases. Nevertheless, social and environmental factors may amplify systematic tendencies toward misaligned confidence/certainty. Overconfidence is associated with overlooking differentials, ignoring important patient information and being less willing to admit mistakes. Hence, mitigating overconfidence is an important direction for future research. Underconfidence has received less attention, but is observed in medical trainees

(Mann 1993; Yang, Thompson, and Bland 2012; Brezis et al. 2019) and can lead to negative outcomes such as delayed treatment (Thompson et al. 2004) and ordering of unnecessary tests (Gupta et al. 2023). Interventions have been tested to improve confidence calibration (such as considering alternative diagnoses and guided reflection), but these have not proven fully successful (Lambe et al. 2018; Costa Filho et al. 2019). More work is therefore needed to design interventions to improve calibration, as one can surmise here a link between miscalibrated confidence and suboptimal patient care. Findings from metacognition are already being used to inform educational practices outside medicine to improve students' memory retention (Dunlosky 2013; Putnam et al. 2016). Although cognitive interventions such as considering alternative diagnoses and guided reflections have been tested, there is yet to be a standardised cognitive framework to teach non-technical skills such as constructive confrontation or expressions of uncertainty.

Our conceptual model of the diagnostic decision process reflects how different levels of factors (related to clinician and environmental context) differently impact accuracy and confidence. This model demonstrates the importance of both behavioural and work system factors within healthcare and how environmental aspects can inform an individual's decision process. Considering the work environment is important given our findings of lower confidence due to environmental factors such as shift busyness and time pressures. This corresponds with other findings of stress being associated with decreased confidence for intermediate levels of uncertainty (Heereman and Walla 2011) and this stress could be contributed to by the healthcare environment that the clinician operates in. Future interventions on diagnosis can refer to this model to understand the part and context of the decision process at which they are administering the intervention.

Beyond these key research themes, two further points emerge from the scope and variety of work identified in our review. First, our review highlights the broad relevance of confidence across different medical subdisciplines, suggesting

the value of focusing on confidence calibration within medical education as a generally applicable approach to improve diagnostic decisions. Second, in terms of methodology, we find that confidence and certainty have been studied in a variety of ways (e.g. using 'assessments' or 'interpretations' as well as diagnostic decisions), but can sometimes be seen a primary outcome measure to improve. Increasing the confidence of clinicians without considering their objective accuracy may exacerbate instances of overconfidence. Future work should focus instead on prompting calibrated rather than increased confidence given the aforementioned impacts on patient care. As objective feedback is often unattainable in medical settings, confidence could be studied using methodologies like virtual reality and high-fidelity simulations that better emulate real medical settings when compared to vignette methodologies whilst also having markers of objective accuracy.

Implications for Future Clinical Research

The current healthcare context presents additional challenges to clinicians with substantial increases in clinical workload in the aftermath of the pandemic, and workforce stress and burnout at their highest in the NHS since recording began (*The state of medical education and practice in the UK. Workplace experiences 2024* 2024). Support to enhance clinical decision making through improved confidence/accuracy calibration could help to relieve pressure on the frontline.

Through our conceptual model of the diagnostic process, we identify three primary areas for future work that have been relatively underexplored in the extant literature. First, the majority of the studies we reviewed have studied diagnosis as a linear process in which information is presented sequentially prior to a final diagnosis, at which point confidence is assessed. This is a helpful idealisation of a process that is often more complex and dynamic in practice, with diagnosis evolving via back-and-forth between seeking information and evaluating that information in the context of currently considered diagnostic possibilities. Moreover, patients' unexpected responses to initial treatment can lead to revision of an initial diagnosis

and prompt further information seeking. Effects of (miscalibrated) confidence are likely to be amplified within these dynamics, such as an overconfident clinician paying too much attention to evidence supporting their diagnosis and neglecting opposing evidence, resulting in greater overconfidence. These dynamics could be studied in more naturalistic, in situ methodologies that are closer to everyday medical practice. Interrupting clinicians in real time to report their diagnostic thinking can be a distraction and potentially a patient safety risk. However, some methodologies permit capturing diagnostic reasoning as it evolves with time and the receipt of new information, such as asking clinicians to think aloud as they make diagnoses (Arocha and Patel 1995; Coderre et al. 2003) or using a visual representation of clinicians' thought processes to capture paths and sources of diagnoses (Feyzi-Behnagh et al. 2014). The use of high-fidelity or virtual reality simulations may also be useful for emulating the pressure and work environment of the clinician (which may affect decision making) (Schmidt et al. 2013; Jans et al. 2023), as well as providing an actual 'patient' to observe (unlike in textual vignettes). Use of such paradigms would also improve the generalisability of results.

Secondly, the vast majority of studies have investigated confidence from the perspective of individual clinicians. However, diagnosis and treatment decisions are often made by teams rather than individuals, particularly in secondary care settings. Evidence from organisational psychology indicates that group decisions depend critically on communicated confidence and uncertainty (Silver et al. 2021): Overconfident team members can anchor a group on an incorrect decision (Mahmoodi et al. 2015). Conversely, underconfident team members may fail to share important information that is unknown to the rest of the group, exacerbating the problems of 'hidden information' and 'shared information bias' (Stasser and Titus 1985). In addition, clinicians may modify how they communicate certainty with others, especially given the collaborative nature of healthcare and the social benefits of communicating opinions with confidence in order to be listened to in a group (Brezis et al. 2019). Situational awareness (SA) is also important in a group, and higher

stress may be associated with overconfidence in SA (Price, Tenan, et al. 2016). Taken together, group medical decisions are clearly an important and naturalistic area for future study.

Finally, the papers in this review have also not looked at individual differences in expressions of confidence, where past work from cognitive psychology has found individual systematic tendencies toward higher or lower confidence (Ais et al. 2016; Navajas et al. 2017). Hence, individual clinicians may have a consistent tendency toward underconfidence or overconfidence that impacts their clinical practice and that training or cognitive aids could address.

Empirical Studies of the Current Work

In terms of empirical work within this DPhil, we focus on the foremost of these underexplored avenues within the current research landscape. Namely, with the experimental studies presented in the remainder of this thesis, we aim to elucidate how diagnosis evolves over time, particularly with the back-and-forth between seeking information and evaluating that information in the context of currently considered diagnostic possibilities. Throughout our studies, we record diagnostic hypotheses whilst allowing participants to seek information they find most useful for determining a diagnosis for a patient and recording their confidence in their diagnosis. Clinicians have been found to be less confident in their diagnoses when they have to seek information themselves compared to when they are provided with all available patient information straightaway (Gruppen et al. 1991). Hence, the information seeking process during diagnostic decisions impacts confidence. Understanding in more detail how information seeking patterns impact confidence can help the future design of cognitive interventions for appropriately calibrated confidence. For example, clinicians could reflect on their confidence by considering what information has been made available to them on the patient and what further information they may need to determine a diagnosis. In addition to studying the interplay between information seeking and confidence, we also use a variety of

methodologies to study the cognitive mechanisms of medical diagnoses. Across the remainder of the thesis, we increase the naturalism of our methods with each chapter in order to address our point on the lack of high-fidelity methodologies in the extant literature. We start with a dynamic version of previous vignette studies that allow us to study information seeking patterns whilst maintaining the experimental control afforded to us by the use of vignette paradigms. We then use a think-aloud methodology to record the thought processes of clinicians as they make diagnoses. This is followed by the use of a virtual reality paradigm that allows for the simulation of more realistic clinical situations whilst still allowing for the recording of information seeking patterns. Finally, we use a rapid ethnography approach within real clinical settings to record instances of diagnostic uncertainty and information seeking within medical practice. Taken together, these approaches allow us to triangulate our findings and ground them in real medical practice. In the next chapter, we start with the first of our studies that uses a dynamic version of previous vignette paradigms where diagnostic hypotheses are recorded and updated over the course of a patient case, with participants able to freely seek information to inform their diagnoses.

Chapter 3 - Information Seeking and Confidence During Medical Diagnoses

Introduction

In the previous chapter, we presented a systematic scoping review of the extant literature on certainty and confidence during medical diagnoses. One output from this review was that we identified one gap in the literature in that past work has not studied the association between the ongoing receipt of information and confidence. Past work has tended to frame information seeking as a single action/choice taken after diagnosis, rather than an ongoing activity that causes regular reconsideration of a diagnosis and course of treatment. In this chapter, we aim to fill this gap by presenting results from an empirical study that investigates the interaction between confidence and information seeking during medical diagnoses.

In this study, we aim to retain the control and simplicity of vignette-based approaches while incorporating some of the complexities that characterise real diagnostic decision making. By doing this, we aim to study diagnostic confidence and accuracy as it develops over the course of a diagnostic decision. In this chapter, we first introduce our paradigm and its flexibility in allowing free information seeking and updating of diagnostic differentials over time. We then introduce the research questions that such a paradigm allows us to investigate.

Our systematic scoping review on confidence during diagnoses revealed two main findings. Firstly, past work that measured confidence and accuracy during diagnostic decisions found that confidence was rarely calibrated to objective accuracy,

leading to overconfidence (Friedman, Gatti, Franz, et al. 2005; Fernández-Aguilar et al. 2022; Garbayo et al. 2023) or underconfidence (Mann 1993; Yang, Thompson, and Bland 2012; Brezis et al. 2019). Secondly, confidence was associated with many aspects of the patient care process, such as prescriptions (Levin et al. 2012; Garbayo et al. 2023), referrals (Calman et al. 1992) and requesting investigations (Tabak et al. 1996; Gupta et al. 2023). The lattermost of these factors is of particular relevance to our research interests. As we identify in our scoping review, the link between information seeking and confidence is a current gap in the extant literature. This link currently made between confidence and information seeking in the literature is that confidence is a subjective judgement that then guides a clinician's subsequent testing and requests for information (Tabak et al. 1996; Gupta et al. 2023).

Crucially, this past work studies information seeking by asking participants/clinicians a single question of whether they would (hypothetically) seek further information or not. However, the link between confidence and information seeking can be expanded upon in three ways. Firstly, we can study how information seeking prior to the point at which confidence is reported affects this confidence. Secondly, we can look at specific aspects of information seeking that are linked to confidence aside from merely the intention to seek or not seek. This includes the amount of information sought, how relevant the information is to the patient case, and the degree to which clinicians vary their information seeking on a case-by-case basis. Finally, confidence and information seeking are likely to influence each other over time. Past work has tended to study diagnosis by asking clinicians to provide a single diagnosis/condition after delivering all available information. Whilst this is a useful simplification for the sake of empirical study, it leaves open the key aspects of how the diagnostic process unfolds in real clinical work. In everyday practice, clinicians engage with diagnosis as an active, ongoing decision process that develops with more time and as more information about the patient becomes available. With a more open-ended paradigm, we investigate how diagnoses evolve over time. For

instance, does a clinician reach an initial diagnosis and then change their mind when they received unexpected information (e.g. a test result)? And does a clinician tend to have a single diagnosis in mind or do they tend to keep an open mind by having considering several diagnostic possibilities at once? Then, with these questions in mind, how does a clinician seek information to further validate their diagnosis?

With these points in mind, we aimed to design a paradigm that better reflects the evolving nature of diagnosis and allows to us to study aspects of the information seeking process. Our paradigm is then a step towards more realistic diagnostic decisions, as it retains the simplicity and control of vignette-based diagnosis whilst allowing more flexibility in information seeking and committing to a diagnosis (or set of diagnoses). This allows us to investigate more fine-grained aspects of information seeking and how they impact diagnoses. Specifically, is clinician confidence informed by the quantity and quality of information sought during the diagnostic process?

For this study, we designed and implemented a novel vignette-based experimental paradigm where participants are asked to provide a list of all diagnostic differentials they are considering based on the information they have received. We ask clinicians to update this list and their confidence at each of a series of stages related to the information sought about the patient: Patient History, Physical Examinations and Testing. We then ask participants to update this list in light of new information by adding or removing differentials. This allows us to more comprehensively capture their thought process in terms of how differentials are being weighed up against each other. Participants report how severe and likely each of their differentials are to draw a more nuanced distinction between differentials. Whereas past work has tended to provide a preset amount of information to clinicians, we instead prompt participants to actively seek out information that they feel is useful for diagnosing the patient they are presented with. This is more analogous to real medical practice where all the required information is not immediately available to clinicians when

presented with a patient. We can then look at information seeking patterns within participants to study how they impact confidence.

Past work from cognitive psychology has shown a link to the quantity of information received and confidence, even if the information is disconfirmatory of one's beliefs (Ko et al. 2022). We can hence investigate in this experiment if this holds during medical diagnoses; if so, we would observe that higher amounts of information seeking would be associated with higher confidence. Information seeking could also be a marker of accuracy in addition to confidence though, as we can study whether clinicians who make more accurate diagnoses seek more appropriate information for the patient. This is important to study as some tests/information are less relevant than others for helping to reach a diagnosis for a patient, resulting in instances of overtesting. With all this in mind, allowing clinicians to freely seek information was then an important tenet for designing this experimental paradigm.

Another aspect of past work we aimed to expand on was on generating differentials (a term used in medicine to refer to hypotheses for diagnoses that a patient could have). Past work has tended to frame diagnosis as a single decision where a clinician responds either to a single diagnosis (Redelmeier and Shafir 2023) or a limited number of conditions that a patient could have (Meyer et al. 2013). In the latter case, clinicians may report multiple differentials when prompted to consider alternative differentials via a cognitive intervention that encourages clinicians not to miss other diagnoses (Feyzi-Behnagh et al. 2014). These experimental approaches do not necessarily represent the manner in which clinicians make diagnoses in their everyday medical practice. While clinicians may focus on a single differential at a time, they may also generate multiple diagnostic possibilities that past experimental paradigms do not capture. For instance, a clinician usually has to weigh up differentials (Schiff et al. 2009), based on their likelihood (taking into account the base rate of medical conditions within a given patient population) and severity (which may be less likely for a given patient, but would be more

dangerous if not considered by the clinician as a possibility). In this sense, a clinician may have, at least, a primary diagnosis that is most likely for the patient and a more serious diagnosis that is less likely but can be dangerous if missed. Our paradigm should then allow clinicians to report multiple differentials at a time without constraints, in order to capture both the primary differentials being considered and the differentials that clinicians keep 'in the back of their mind'. We can then use the breadth of differentials considered by clinicians as another marker of uncertainty that may guide their subsequent information seeking. By allowing participants to record a list of all differentials they are considering at each stage, we can capture their thought process as it pertains to the information they have received prior to that point.

As our paradigm is designed to capture the diagnosis process as evolving over time, we can also study confidence differently to past work. Rather than seeing confidence as a static quantity, confidence may shift to reflect the current relative strength of evidence in favour of a decision alternative (Vickers and Packer 1982). Our paradigm then records confidence alongside the participants' list of differentials as it is being updated. We can not only use this facet to link confidence to the breadth of diagnoses considered but also to examine how confidence changes over the course of a case. For instance, a clinician may receive a surprising or inconclusive test result for a patient, causing them to reduce their confidence and seek more information as a result to increase their confidence. Our measure of confidence is also distinct from measures used in past work as we aim to capture the diagnostic process as it pertains to subsequent treatment of patients. An ideal diagnostic process would involve a clinician seeking information to formulate a diagnosis of a patient and, in the process, create a treatment plan to address this diagnosis. We then capture confidence in this study specifically to measure how ready the clinician is to treat the patient, as opposed to past studies that have tended to ask clinicians how confident they are that their diagnosis is the correct one.

There are multiple ways we can define how calibrated participants' confidence is. To recap, measuring calibration requires a subjective judgement of confidence and an objective measure of accuracy to compare this confidence judgement against. For past work where a single differential is provided by clinicians when they are asked to make a diagnosis, accuracy is relatively easy to measure, as it simply requires marking the provided differential as either correct or incorrect. In our paradigm however, participants not only provide all possible differentials that they are considering but also provide assessments of how likely each differential is. We must then consider how to assess each set of differentials as being accurate or not. A lenient definition of accuracy is to simply mark a set of differentials as accurate if it includes a correct differential. Henceforth, we refer to this measure as Differential Accuracy. However, this measure does not take into account the likelihoods assigned to differentials, so it does not consider how clinicians weigh up differentials against each other. Participants are also more likely to be correct by simply including more differentials in their list. A stricter definition of accuracy would be to look at whether the most likely differential (as rated by the participant) is correct and use the likelihood value assigned to this. Henceforth, we refer to this measure as Highest Likelihood Accuracy. However, this penalises participants who consider the correct differential as likely but not as their primary diagnosis. We therefore use the following measure of accuracy as our primary measure: we look at the likelihood rating assigned to the correct differential if it is present in the participant's list. This provides a more nuanced measure of accuracy that takes into account how differentials are weighed up against each other, which marks a difference from accuracy as it is defined in past work. We should note however that assessing the calibration of participants' confidence judgements is potentially contingent on the accuracy measure used. We therefore measure calibration using our primary measure of accuracy (the likelihood of the correct diagnosis), but also provide results using the other two measures mentioned here (Differential Accuracy and Highest Likelihood Accuracy).

For our studies, we chose to focus on medical students who were relatively advanced in terms of their medical education but were still early in their clinical experience. Medical students are yet to settle on a particular medical subdiscipline to specialise in, which allows our vignettes to cover a variety of medical conditions and pathologies. We also focus on students as findings from our work could have implications for future medical education in terms of how clinical reasoning and cognitive psychology is taught. Finally, recruiting students allows us to collect a relatively large sample to facilitate detailed analysis of information seeking patterns.

Research Questions

With this study, we investigated the following research questions:

- **Is confidence calibrated to accuracy within medical students?** - Whilst past work has found disassociations between diagnostic confidence and accuracy, these were found in the context of simple tasks with limited flexibility in terms of information seeking and recording multiple diagnostic differentials. We therefore investigate if similar miscalibrations of confidence occur within a more flexible experimental paradigm.
- **How do medical students weigh up competing differentials during the diagnostic process** - Past work has considered that clinicians may have multiple differentials in mind when diagnosing a patient, but such research has not studied how the differentials being considered changes with the receipt of new information. Specifically, do medical students tend to narrow the differentials over time (i.e. akin to a process of elimination) or do they tend to broaden their thinking as new information on the patient is received?
- **How do confidence and information seeking interact in the diagnostic process?** - We expect that confidence would predict information seeking, such that confidence in diagnoses is predicted by both the quality and quantity of information sought.

- **Do differences in confidence and information seeking predict differences in diagnostic accuracy?** - We expect accuracy to associated with the quality/suitability of information seeking but not the quantity of information sought or by confidence (as per the aforementioned miscalibrations of confidence).

Methods

This study was designed to understand how information seeking, confidence and differential generation interact within the diagnosis process. Specifically, we investigated whether information seeking patterns were associated with diagnostic accuracy and confidence. We conducted a vignette-based diagnosis study with medical students to characterise their diagnostic process and potentially to inform future on how diagnostic reasoning is taught to students, especially when it comes to weighing up competing differentials. Data is openly available on OSF: `https://osf.io/kb54u/`.

Participants

We recruited final year medical students within the UK. 85 medical students completed the study, including 32 males, 52 females and 1 participant who identified as non-binary. Their ages ranged between 22-34 years (M = 24.2). Participants were recruited between July 11th 2022 and April 6th 2023 via emails sent to all UK medical students within a UK Medical Schools Council mailing list. Participants were emailed with a study information sheet and a link to access the experiment, where they first provided consent via an anonymous online form. After doing so, the participant provided demographic information (age, gender and years of medical experience). The study was conducted online, with participants able to run the experiment in a browser on a desktop computer or laptop (and not a phone or tablet) in a location of their choice. The experiment was coded using the JSPsych Javascript plugin. The code is publicly available on Github: `https:`

Online Study

`//github.com/raj925/DiagnosisParadigm`. Ethical approval was granted by the Oxford Medical Sciences Interdivisional Research Ethics Committee under reference R81158/RE001.

Materials

This study involved patient vignettes that we adapted from anonymised past cases developed by Friedman, Gatti, Elstein, et al. (2001). Six cases were chosen, each designed to indicate a specific underlying condition the patient had: Aortic Dissection (AD), Guillain-Barre Syndrome (GBS), Miliary TB (MTB), Temporal Arteritis (TA), Thrombotic Thrombocytopenic Purpura (TTP) and Ulcerative Colitis (UC). The order in which the cases were presented was randomised for each participant. We also included a practice case (Colon Cancer) to familiarise the participants with the experimental procedure and the interface. Cases were chosen to reflect a variety of affected pathophysiological systems and to test medical students on medical conditions that they were expected to know given their level of education/training.

A panel of 3 subject matter experts (practising doctors and researchers within the NHS and the OxSTaR centre: `www.oxstar.ox.ac.uk`) were recruited to design the vignettes used in this study. These medical professionals were at differing experience levels, with their medical roles at the time of this study as follows: Speciality trainee (ST7) in Anaesthetics, Foundation (F1) Doctor and Gastroenterology Consultant. The panel assisted with translating terms (e.g., medication names, tests etc.) from US to UK doctors' vernacular, updated patient details to be more current and provided input on the choice and complexity of the cases chosen.

Procedure

The goal of the task was to determine a diagnosis, or diagnoses, for each presented patient (see procedure in Figure 3.1 below). Information on the patient was split into a series of discrete stages to control what information the participants had

access to at any given point in the experiment. Each point of new information was termed an "information stage". Participants were able to seek information freely until they were ready to move on.
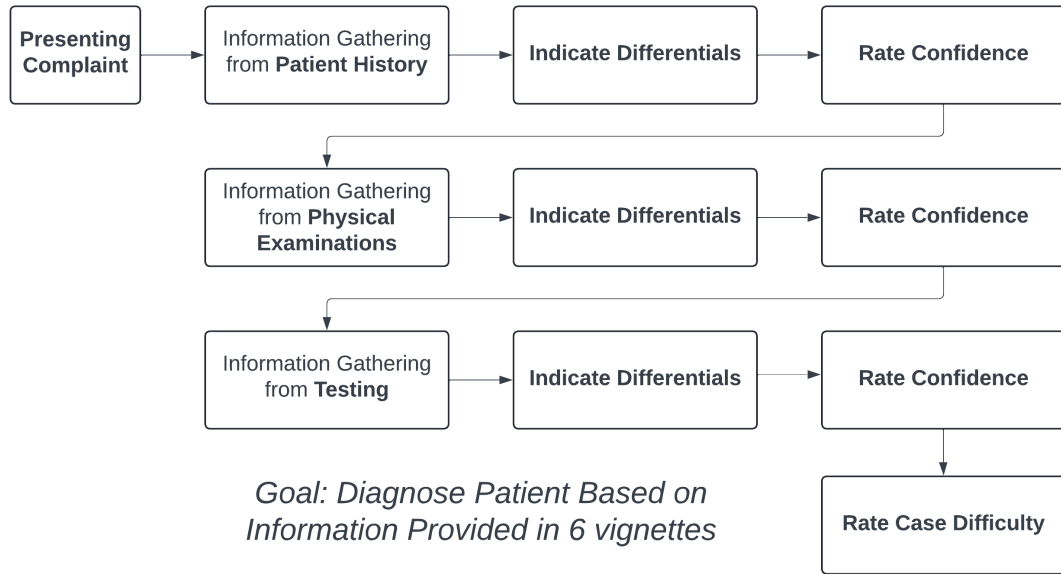
Figure 3.1: Paradigm of the online vignette study, showing the procedure for a single patient case.
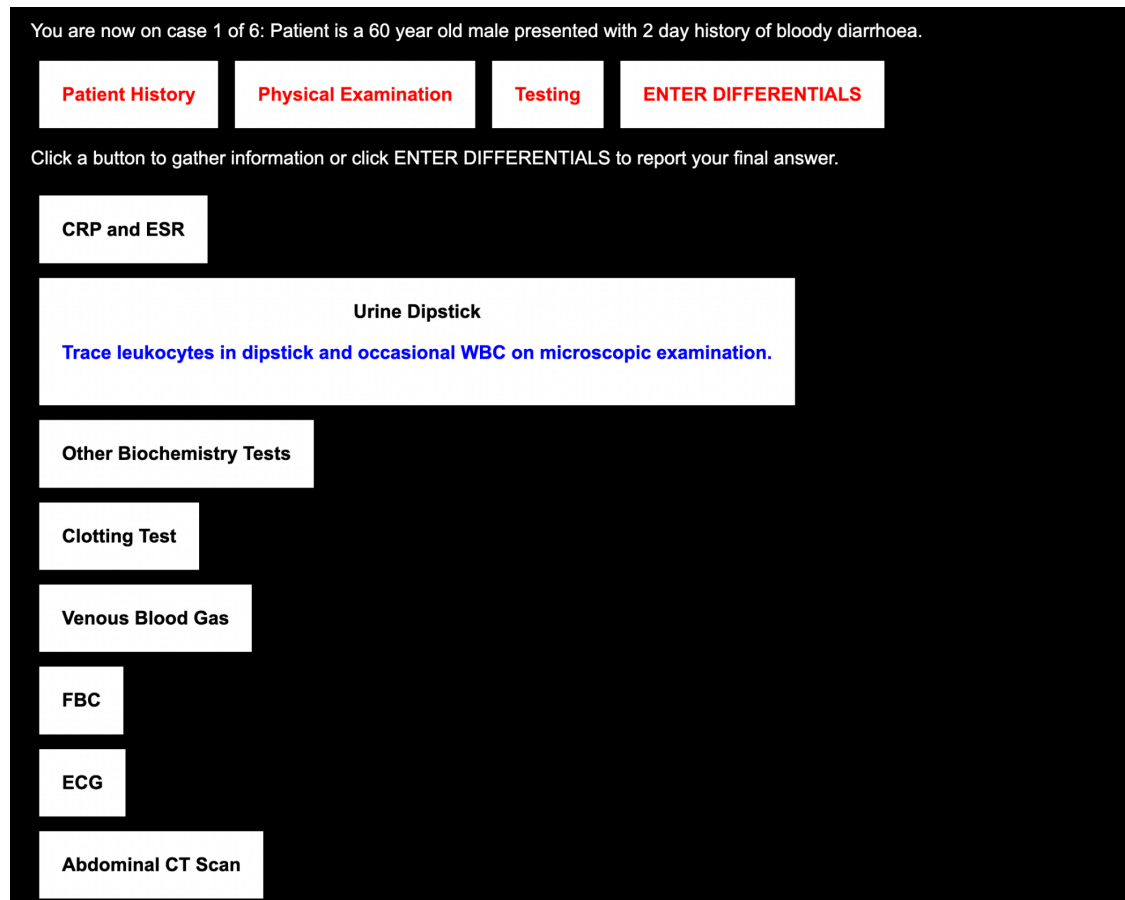
Figure 3.2: Screenshot of the interface. Shown here is the screen in which the participants seek information during the Testing stage.

The procedure of a single case is as follows. The participant was asked to imagine that they are working in a busy district hospital and they encounter patients in a similar way to how they would in their real medical practice. At the start of each case, the participant was shown a description of a patient, which includes the patient's gender, age and their presenting complaint. An example of this was: "Patient is a 68 year old male presenting with fever and arthralgia". Each case is split into three information stages: Patient History, Physical Examination and Testing (in this order). This structure has been utilised in past work as being prototypical for a diagnosis (Hampton et al. 1975; Peterson et al. 1992). The Patient History stage included information on "Allergies", "History of the Presenting Complaint", "Past Medical History" and "Family History". The Physical Examination stage included 'actions' that a doctor may take when examining a

patient, such as "auscultate the lungs", "abdominal examination", "take pulse" and "measure temperature". Finally, the Testing stage involved information on any bedside tests or tests they may request from another department. This includes "Chest X-Ray", "Venous Blood Gas", "Urine Dipstick" and "Clotting Test". In total, there were 29 possible information requests across the three stages, with the available set of information being the same for all patient cases.

Figure 3.3: Screenshot of the interface. This is the screen in which participants report their current list of differentials, including the name of each condition as well as the severity and likelihood ratings for each condition. Participants remove conditions by clicking the red cross on the right hand side of each differential. Participants add a new differential by clicking the plus icon below the list.

When a participant clicked on any of these requests, the information for that request was shown on screen after a 3 second delay. This delay was added after pilot testing (with 10 participants) revealed that participants tended to select most, if not all, of the information available to them. We mitigated this tendency by adding this delay and by emphasising to participants during the task instructions that they should only request information that they believe will help them with diagnosing the patient for that specific case. Participants were free to request the same piece of information multiple times, including information from a previous stage. At any

point, they could choose to stop gathering information for that stage. They were then taken to a new screen where they reported a list of all differential diagnoses that they were considering for that patient at that stage. For each differential, participants reported a likelihood rating, ranging from 1 (very unlikely) to 10 (certain), and a "level of concern" (which was how concerned they would be for that patient if this differential really was the patient's underlying condition) on a 4 point scale (labels of "Low", "Medium", "High" and "Emergency"). In subsequent stages, the list from the previous stages was available for participants to update concern/likelihood ratings, or to add/remove differentials from the list. Even at the last information stage, participants could report multiple differentials.

After recording their differentials, participants were then asked to report their confidence that they were "ready to start treating the patient" on a 100 point scale, ranging from not at all confident to fully confident. Participants also indicated using a checkbox whether they are ready to start treating the patient, at which point a text box appeared for them to report what further tests they would perform, any escalations they would make to other medical staff and treatments they would start administering for the patient. This allowed participants to express what actions they would take that were not covered by our set of available information requests. Once all three stages were completed, participants reported how difficult they found it to determine a diagnosis for that case, on a scale from 1 (trivial) to 10 (impossible). At the end of all six patient cases, participants were told the 'true' conditions for all the patients. The session took approximately 40-60 minutes to complete.

Data Analysis

During analysis, no sought information was recorded for three cases across participants (i.e. at all three stages during a case, the participant did not appear to seek any information). These cases were excluded from analysis. We now describe

the key dependent variables for this study. The first set of the measures (Case-Wise Measures) are calculated at each of the three information stages (except for Perceived Difficulty). When averaging these variables within a participant, we use the values obtained at the final stage (i.e. Testing). The second set of measures (Derived Information Seeking Measures) are based on information seeking by participants on each case across all three information stages.

Case-Wise Measures

- *Correct Differential Included*: This measure captures whether participants consider a correct diagnostic differential. Responses were coded for correctness manually with help from a medical consultant, who looked at all the information available for each case and determined which diagnoses could be valid answers. Each case is marked as correct if the list of differentials provided includes the correct condition or a differential considered correct as per our marking scheme in Table A1 of the Appendices. Otherwise, the case is considered incorrect if a 'correct' differential is not included.

- *Accuracy*: Our main measure of diagnostic accuracy is computed as the likelihood value assigned to the correct differential for the case (and scored as 0 if this differential is not listed). For a case to be considered 'correct', the participant should have reported the correct condition for that case within their list of differentials regardless of the number of differentials provided. Likelihoods range from 1-10 when a correct differential is included and has a value of 0 when a correct differential is not included. The value is then rescaled to range from 0 to 1, where 1 corresponds to a correct differential assigned maximum likelihood. If multiple differentials that are considered correct were provided, then the likelihood value of the closest differential (as per our marking criteria with help from a medical consultant) to the true condition was used.

- *Highest Likelihood Accuracy*: This stricter measure accuracy is computed as the likelihood value assigned to the differential with the highest likelihood (in comparison to other differentials provided in the participant's list) if this differential is considered correct. If not, a value of 0 is assigned. Again, likelihoods range from 1-10 for correct differentials, so this is rescaled to range from 0 and 1.

- *Confidence*: Participants reported their confidence that they are ready to start treatment at each information stage. Initial Confidence refers to the reported confidence after the first stage of information seeking (Patient History), whilst Final Confidence refers to the reported confidence after the third and last stage of information seeking (Testing). As with accuracy, confidence is rescaled to fall between 0 and 1 to allow for direct comparison between the two variables. We can then use these two variables to calculate Confidence Change, by subtracting the participants' Initial Confidence from their Final Confidence. Hence, a positive value for Confidence Change means that the participant has gained confidence over the course of the patient case.

- *Number of Differentials*: This measure captures the breadth of diagnoses considered by participants. The number of items in the list of differentials was recorded at each stage. Initial Differentials refer to the number of differentials after the first stage of information seeking (Patient History), whilst Final Differentials refer to the number of differentials after the third and last stage of information seeking (Testing).

- *Change in Differentials*: This measure captures how much participants change the differentials they consider over the course of the case. In other words, we can look at how much participants have narrowed or broadened their list of differentials as they receive more information. This is calculated by taking the absolute value of the difference between the number of Initial Differentials and the number of Final Differentials.

- *Perceived Difficulty*: The subjective rating by participants at the end of each case for how difficult they found it to determine a diagnosis for that patient case. This is reported subjectively by each participant on a scale from 1 (trivial) to 10 (impossible).

Derived Information Seeking Measures

- *Amount of Information Seeking*: This measure captures the amount of information that participants seek on cases relative to how much they could have sought if seeking all available information. We take the number of unique tests requested at a given information stage (i.e. not including any tests from a previous stage and excluding repeat tests) and divide this by the number of possible tests available.

- *Information Value*: We calculate a measure of information value to capture how appropriate the information sought for a case is for the patient's condition. We compute the average value of sought information across cases. To do this, we take each of the 29 pieces of information in turn by case and split all cases completed across participants into two groups: cases where that information was sought at any stage and cases where that information was not sought. For each group, we compute the proportion of trials where the students included a correct differential, and then take the difference between these two values. A positive value would indicate that students were more likely to identify the correct condition with that information rather than without that information. This difference can be considered that information's 'value'. We then calculate the sum of all information values for each case. This gives an overall measure of, on average, how useful the information was that participants sought on each case.

- *Information Seeking Variability*: We calculate a measure of how much, for a given set of cases, information seeking varies across cases. This is operationalised as the average dissimilarity between cases' information seeking (by

taking the average of all pairwise comparisons) using each piece of information as a binary variable (i.e. whether it was sought or not). This measure is calculated both within participants, to tell us how much each participant varied the information they sought across their cases, and between participants, to tell us how dissimilar participants are to each other in terms of the information sought for a given condition. We calculate this value using the Dice coefficient (Dice 1945), due to it being well suited specifically for binary data, as well as its increased weighting on discordant pairs (ie a piece of information being sought in one case but not the other). A higher value between two cases indicates that the information sought on those trials are more dissimilar to each other.

We used statistical analyses to consider differences in confidence, accuracy and information seeking. When looking at how our variables change over the three information stages, we used Analysis of Variance models with Bonfferoni-corrected pairwise T-tests on all pairwise comparisons. We test if there is a relationship between confidence and information seeking (Amount, Value Variability) and between accuracy and information seeking using Pearson's product moment correlation tests (an alpha value of less than 0.05 was regarded as statistically significant). These help us answer how confidence and information seeking interact during the diagnostic process and whether differences in diagnostic are predicted by information seeking and confidence. Our sample of 85 participants is calculated as having 80.4% power to detect a medium effect size of r = 0.3 (using an approximate arctangh transformation correlation power calculation). In addition to correlations for variables averaged (across cases, per participant), we also fit linear mixed effects models to predict information seeking, changes in differentials and changes in confidence (between the Patient History and Testing stages). In order to test if information seeking patterns are predictive of differences in accuracy, we used generalised logistic regression to classify cases as being performed by high or low

accuracy participants (via a median split). To test if information seeking patterns are predictive of the case (i.e., whether participants tailor their information seeking to each patient case), we use penalised multinomial regression to classify cases by their patient condition. Both models were trained on the information requests as binary variables (with a 1 signifying that the information was sought for that case and 0 when the information was not sought). We used Leave One Out Cross Validation for both models, such that each case is predicted by training the algorithm on all other cases.

Results

Overall Performance and Calibration

We first look at our research question as to whether confidence is calibrated within medical students. When comparing Accuracy (taking into account the likelihood assigned to correct differentials) to Confidence, we find, across stages, participants' Confidence was aligned to their Accuracy (see Figure 3.4 below). To determine whether there is any systematic discrepancy between subjective confidence and objective accuracy across stages, we compute a paired t-test between average Confidence and average Accuracy (across cases) at each stage. There was no evidence of a difference between the two at the Patient History ($t(84) = 0.29$, MDiff = 0.01, p = 0.77) and Physical Examination stages ($t(84) = 0.74$, MDiff = 0.01, p = 0.46), but there was a statistically significant difference between the two at the Testing stage ($t(84) = 2.35$, MDiff = 0.05, p = 0.02). This indicated well-calibrated confidence after Patient History and Physical Examination, but a slight overconfidence across participants after Testing.

To investigate the dynamics of confidence and accuracy further, we look at how both variables change over the course of the information seeking stages. Across cases, accuracy increased with each stage of information gathering as per our Accuracy measure, ($F(2, 252) = 21.6$, $\eta^2 G = 0.15$, p < .001). Participants had lower accuracy

at the Patient History stage (M = 0.31, SD = 0.14) than during the Physical Examination (M = 0.04, SD = 0.15) and Testing stages (M = 0.41, SD = 0.15). Pairwise comparisons between the History stage and each of the other two stages are significant (ps < .001). Table 3.2 shows overall accuracy (at the Testing stage) by case, indicating that there was variability in performance between cases.

Confidence also increased as participants received more information ($F(2, 252) = 21.6$, $\eta^2G = 0.15$, $p < .001$). Participants reported lower confidence during the Patient History stage (M = 0.3, SD = 0.15) than during the Physical Examination (M = 0.41, SD = 0.17) and Testing stages (M = 0.47, SD = 0.47). Pairwise comparisons between History and each of the other two stages are significant (ps < .001). We note here that confidence was on average below 50% even at the end of each case, which indicates that participants were not highly confident to start treatment. This is reflected in participants expressing their readiness to treat the patient in the vignette, which allows them to enter a treatment plan for the patient. In 38% of cases, participants reported they were ready to treat the patient and entered a treatment plan.

Case	Differential Accuracy	Accuracy	Highest Likelihood Accuracy	Final Confidence	Difficulty	Information Seeking
AD	0.60	0.28	0.12	0.49	5.9	0.59
GBS	0.75	0.41	0.30	0.37	6.9	0.63
MTB	0.43	0.24	0.10	0.45	6.7	0.64
TA	0.74	0.50	0.45	0.49	6.2	0.62
TTP	0.61	0.34	0.20	0.41	6.8	0.66
UC	1.00	0.73	0.69	0.62	5.2	0.55

Table 3.1: Average statistics across participants for each case (leftmost column, AD = Aortic Dissection, GBS = Guillain Barré Syndrome, MTB = Miliary Tuberculosis, TA = Temporal Arteritis, TTP = Thrombotic Thrombocytopenia Purpura,

Online Study

UC = Ulcerative Colitis). Differential Accuracy (0-1) refers to the proportion of participants who correctly included the correct condition or a condition considered correct for that case based on our marking criteria. Highest Likelihood Accuracy refers to the likelihood assigned to the differential with the highest likelihood if it is correct (1-10), otherwise the value for a given case is 0 if this differential is incorrect. This value is then rescaled to range between 0-1. Accuracy refers to the average likelihood (on a 1-10 scale, rescaled to range between 0-1) assigned to a correct differential if included. Confidence refers to the confidence provided by participants on their readiness to treat the patient at the Testing stage (on a scale of 0-100, rescaled to fall between 0-1). All these measures are calculated based on values observed at the final information stage of each case (i.e. the Testing stage). Difficulty refers to the subjective rating provided at the end of each case of how difficult participants found the case to be in terms of determining a diagnosis (on a scale of 1-10).
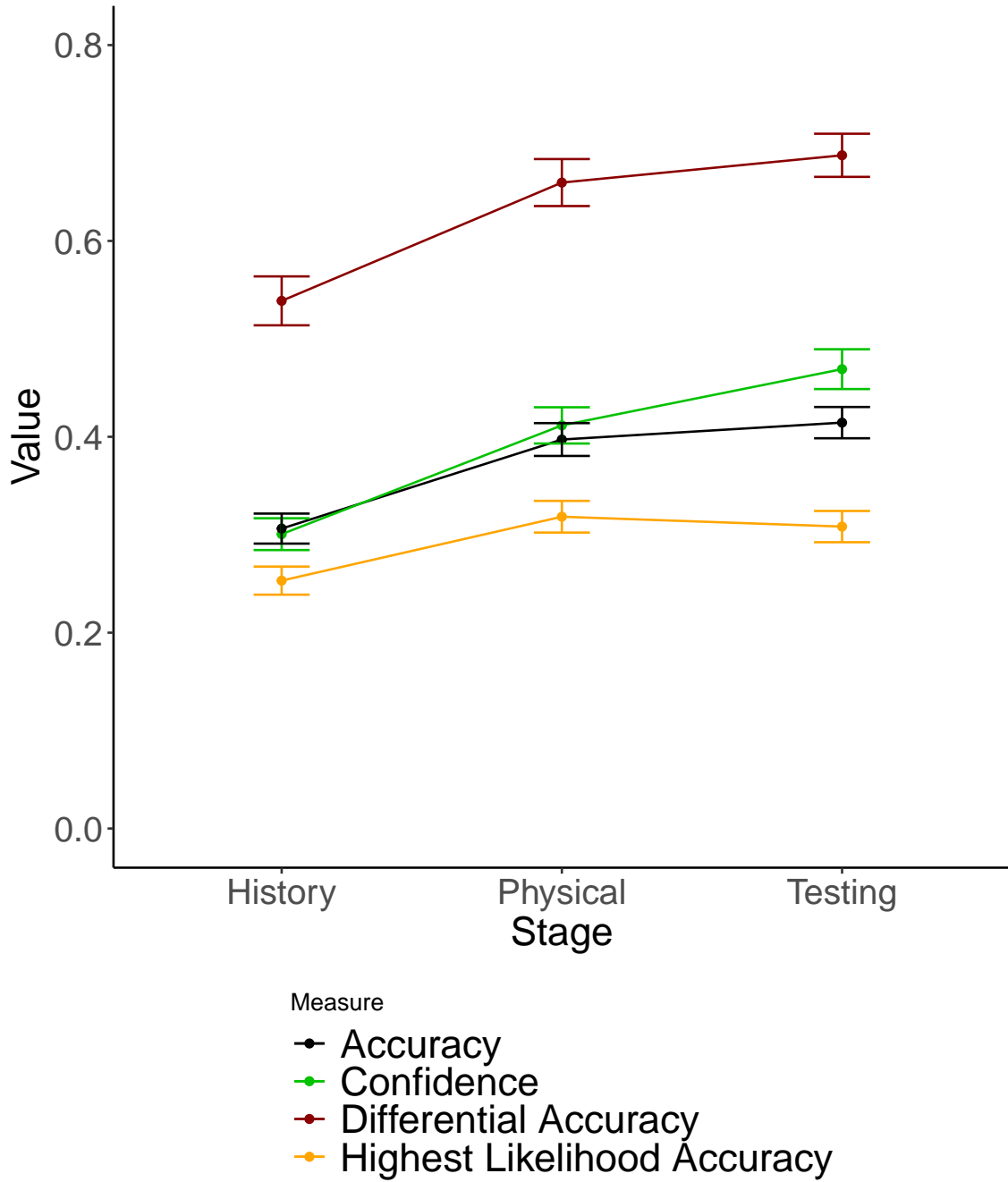
Figure 3.4: Graph showing Confidence (green) at each of the three information stages (History = Patient History, Physical = Physical Examinations, Testing = Testing) in comparison to our main accuracy measure (black, likelihood value assigned to the correct diagnosis), the more lenient measure of the proportion of trials where a correct differential was included (dark red) and the stricter measure of the value assigned to the highest likelihood differential if it is correct (orange). Values

shown are averaged across participants and cases, with the error bars representing standard error.

In order to examine the observed overconfidence in more granularity, we compare confidence and our primary accuracy measure by case (the mean values of which can be found in Table 3.1). We conducted paired t-tests for each condition's cases by comparing accuracy and confidence values (at the final Testing stage) to observe if they significantly differ from each other. A p value of less than .05 is interpreted as evidence for overconfidence or underconfidence (depending on the direction of the effect). We observed overconfidence for the AD case ($t(84) = 4.71$, MDiff = 0.21, p = < .001) and for the MTB case ($t(83) = 4.31$, MDiff = 0.21, p = < .001). We observe underconfidence for the UC case ($t(82) = -3.51$, MDiff = -0.12, p = < .001). The remaining cases did not yield a significant effect, indicating calibrated confidence judgements across participants. The overall overconfidence after Testing that we observe in Figure 3.4 is then driven by the AD and MTB cases, for which accuracy was lowest compared to other cases and confidence was not sufficiently adjusted to reflect this.

Differentials

We analysed the number of differentials to provide insights into the diagnostic decision process across stages, specifically the degree to which it follows a process of deductive narrowing (decreasing differentials) or open-minded broadening (increasing differentials). Analysis of the number of differentials considered by participants at each stage provides little evidence for an overall strategy of deductive narrowing towards a single differential. Instead, participants overall increased the number of the differentials they reported as they received more information ($F(2, 252) = 11.66$, $\eta^2G = 0.08$, p < .001). Participants reported fewer differentials during the Patient History stage (M = 3.2, SD = 1.12) than during the Physical Examination (M = 3.89, SD = 1.32) and Testing stages (M = 4.13, SD = 1.43). Pairwise comparisons between the History stage and each of the other two stages are significant (ps

$< .05$). The majority of participants (74/85) did not decrease the number of differentials between Patient History and Testing on any case, indicating a tendency to widen rather than narrow the set of considered diagnoses through the evolving decision process (even while, on average, growing increasingly certain of the correct diagnosis). As can be observed in Figure 3.5 below, there is general consistency in terms of participants broadening their differentials with more information despite some inter-participant variability, with a small minority of participants narrowing their differentials on average.
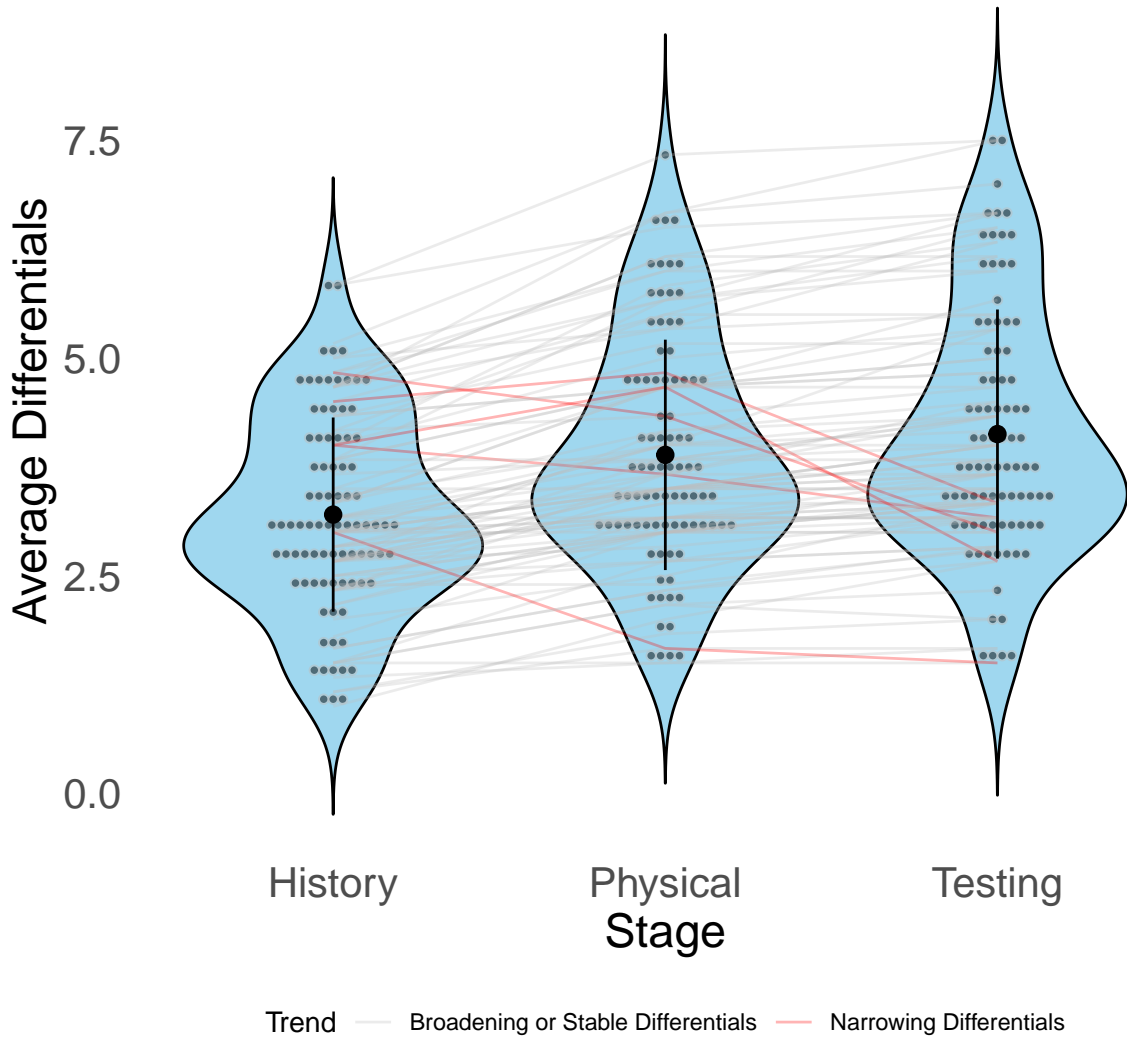
Figure 3.5: The average number of differentials after each stage of information seeking (x-axis, History = Patient History, Physical = Physical Examinations, Testing = Testing). The width of the blue area corresponds to the amount of data points that fall within that part of the y-axis, with a wider area meaning a higher concentration of data points. The larger black dots indicate the mean values, whilst the larger black vertical lines indicate standard deviations. The grey dots show individual values at each stage, with lines connecting the dots at each stage to represent individual participants' trend across the information seeking stages. The participants who show a narrowing of differentials (i.e. recording fewer differentials at the Testing stage compared to the Patient History stage) are marked with a red line, whilst the remainder of participants are marked with a grey line.

Online Study

As a first probe of the dynamics of the diagnostic process, we analysed whether participants who generated more differentials early in the diagnostic process go on to seek more information by conducting a Pearson's Correlation test on individual differences. We find a positive correlation (see Figure 3.6) between the average number of differentials generated from the Patient History and the average amount of information sought during cases ($r(83) = 0.3$, 95% CI = [0.09, 0.48], p = 0.005, Figure 3.6a). As previously discussed, participants rarely seem to remove differentials from consideration. Therefore, one can surmise here that higher information seeking is associated with the consideration of more diagnostic differentials. We also find evidence for a positive association between the number of initial differentials and the change in confidence (i.e. the difference in confidence reported during the Patient History stage and the Testing stage) ($r(83) = 0.23$, 95% CI = [0.02, 0.42], p = 0.04, Figure 3.6b).

Given that we observe an broadening (increasing number) of differentials across participants, we ask how this change in differentials related to information seeking and changes in confidence. As well the initial diagnostic breadth of participants, we are also interested in whether information seeking and changes in confidence relate to how much participants change the number of differentials they consider over the course of the case. This allows us to capture how much their diagnostic differentials have changed based on the information received. We find the amount of Differential Change was associated with both the amount of information sought ($r(83) = 0.3$, 95% CI = [0.09, 0.48], p = 0.005, Figure 3.6c) and change in confidence ($r(83) = 0.39$, 95% CI = [0.19, 0.56], p = < .001, Figure 3.6d). These results indicate that participants who tended to increase differentials also tended to seek more information and increase their confidence to a greater extent. If broadening of differentials was a reflection of diagnostic uncertainty, we may have expected a decrease in confidence, but this does not appear to be the case. We also fit a linear mixed model to predict Differential Change on individual cases, using the amount of information seeking (after Patient History), the number of initial differentials

and Confidence Change as fixed effects, as well as both the patient case/condition and participant as random effects. We find that the amount of information seeking was positively predictive of Differential Change ($\beta = 0.86$, SE = 0.23 t = 3.76, p < .001), whilst the fixed effects were not found to be significant. This indicates that higher changes in the number of differentials being considered (i.e. how much a participant was able to broaden or narrow their differentials) was associated with higher information seeking.
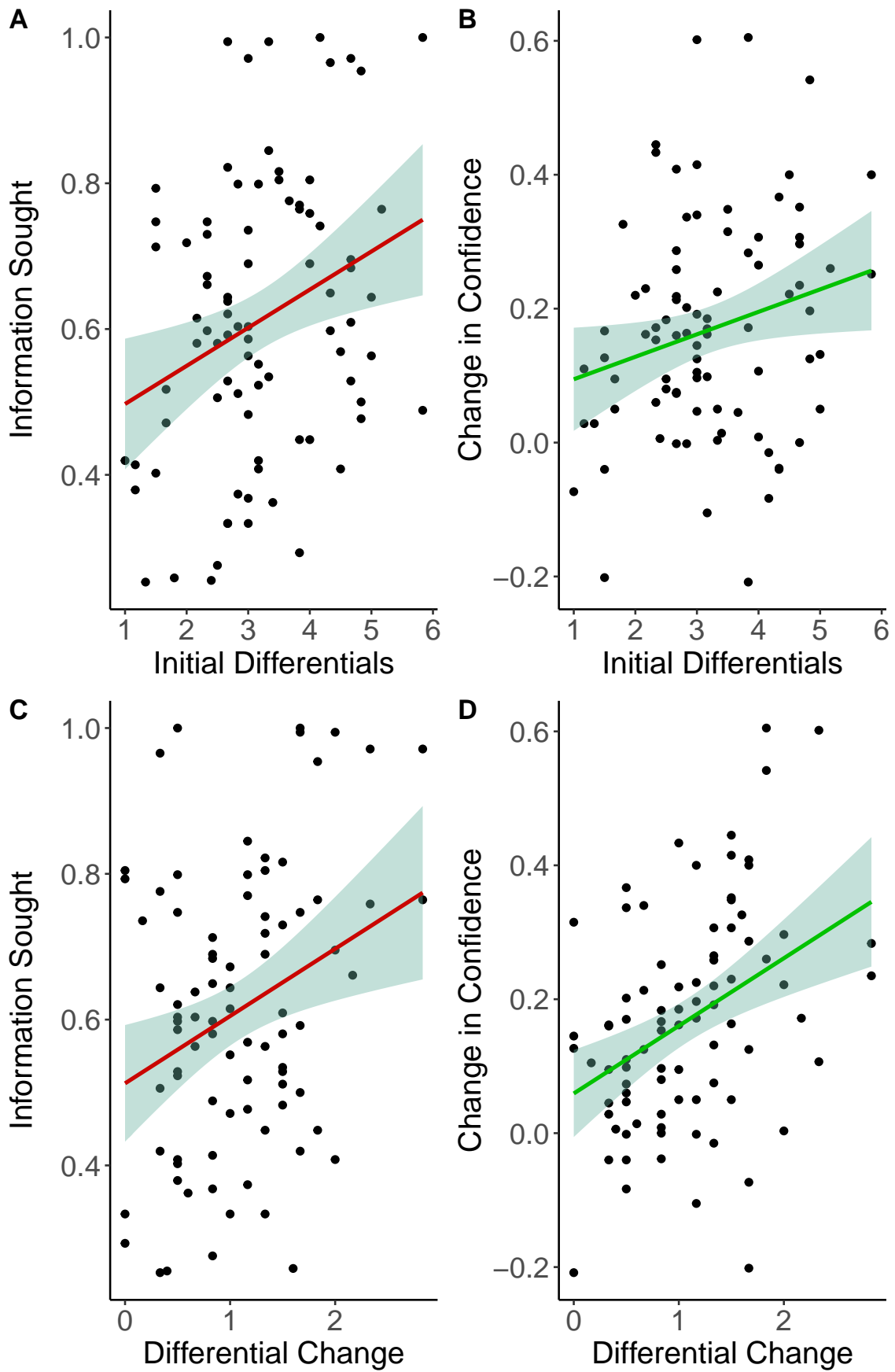
Figure 3.6: Scatter plot showing the relationship between the number of initial differentials reported at the Patient History stage (x-axis, figures 3.6A and 3.6B) and the change in differentials (x-axis, figures 3.6C and 3.6D) against both the proportion of available information sought (y-axis, figures 2.6A & 3.6C) and change in confidence (y-axis, figures 3.6B & 3.6D). Each point represents a single participant with all three variables averaged across the six cases that each participant performs. Initial Differentials refers to the average number of differentials that participants report in their list at the Patient History stage. Differential Change refers to the absolute difference in the number of Initial Differentials (at the Patient History Stage) and the number of Final Differentials (at the Testing Stage). Information Sought refers to the average proportion of available information sought, with each case containing 29 pieces of information across the Patient History, Physical Examination and Testing stages. Change in Confidence refers to the difference in reported confidence at the Patient History and Testing stages, such that a positive represents that the participant on average increased in their confidence over the course of the cases. The line of best fit is plotted using the geom_smooth function in R with a linear model. The shaded region shows the 95% confidence interval of the correlation.

Information Seeking

To investigate our research questions of how both confidence and accuracy interact with information seeking during the diagnostic process, we first look at broad characteristics of information seeking and then ask if they are predictive of differences in confidence and accuracy. When investigating whether participants became more selective in their information seeking over the course of cases, we find that the Proportion of Information Seeking decreased with each information stage (F(2, 252) = 57.26, η^2G = 0.31, p < .001). Participants sought more of the available information during the Patient History stage (M = 0.85, SD = 0.19 than during both during the Physical Examination (M = 0.59, SD = 0.24) and Testing stages (M = 0.5, SD = 0.22). All pairwise comparisons are significant (ps < .05). This selectivity in information seeking does not seem to reflect participants being less certain about their diagnoses, which the general pattern of broadening differentials may have indicated.

Given the design of our task, we ask if seeking all available information is in fact a helpful strategy for increasing diagnostic accuracy by testing for a correlation between the two. We do not find that participants who sought more information across cases were also more accurate in their diagnoses (r(83) = 0.16, 95% CI = [-0.05, 0.36], p = 0.13, Figure 3.7A). However, participants who sought more information tended to have increased their confidence more during cases (r(83) = 0.24, 95% CI = [0.02, 0.43], p = 0.03, Figure 3.7C). While seeking more information may imbue students with a greater level of confidence, we do not find evidence that this translates consistently into more accurate diagnoses. This finding links to the results presented in Figure 3.4, in which confidence and accuracy were related to one another but imperfectly (especially during the Testing stage, during which subjective confidence was higher than objective accuracy across participants). We also fit a linear mixed model to predict Information Seeking (after Patient History) on individual cases, using the number of initial differentials and Initial Confidence

as fixed effects, as well as both the patient case/condition and participant as random effects. We find that the number of initial differentials was positively predictive of the amount of information seeking ($\beta = 0.01$, SE = 0.01 t = 2.44, p = 0.02) and that Initial Confidence was negatively predictive ($\beta = 0$, SE = 0 t = -3.81, p = 0). This indicates that higher information seeking was predicted by lower confidence and higher initial diagnostic breadth.

In order to examine more specifically what differences in information seeking are driving differences in both accuracy and confidence, we look at their relationship with informational value. We assess the degree to which each participant's accuracy is predicted by the quality of the information they sought and find evidence for a positive relationship between accuracy and information value (r(83) = 0.22, 95% CI = [0, 0.41], p = 0.05, Figure 3.7B), as well as between confidence and information value (r(83) = 0.29, 95% CI = [0.08, 0.47], p = 0.01, Figure 3.7D). When comparing the correlations between both information amount and information value to accuracy via a Fisher's z-Test of dependent correlations, we find they are not significantly different from one another (z = 1.08, p = 0.28). This means that we cannot make a valid comparison between the correlations with information amount and information value with respect to accuracy.

We also fit a linear mixed model to predict Change in Confidence on individual cases, using the number of initial differentials, the amount of information seeking (after Patient History) and change in differentials as fixed effects, as well as both the patient case/condition and participant as random effects. We find no significant fixed effects (ps > .1).
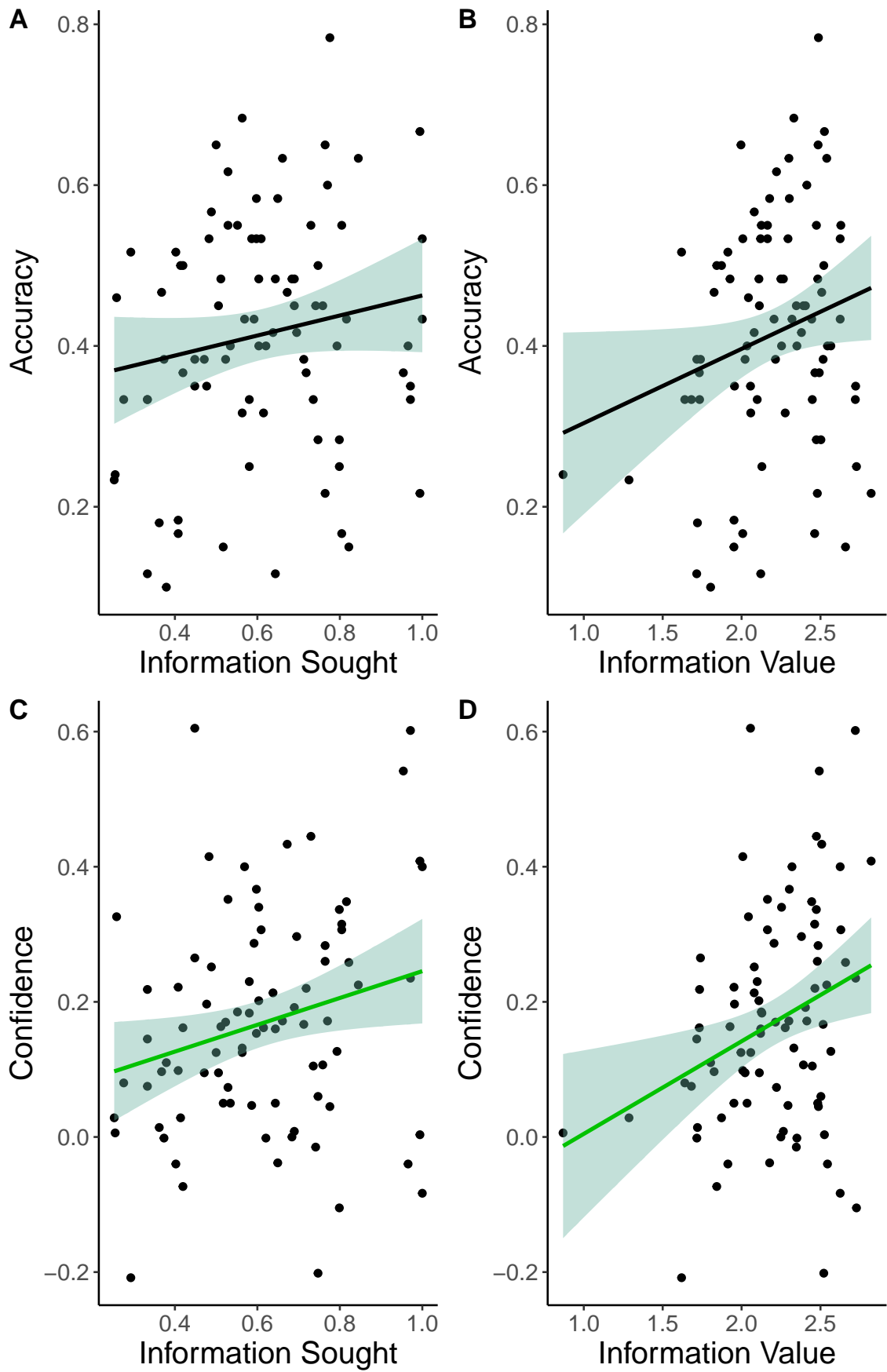
Figure 3.7: Scatter plots showing our information seeking variables (amount in figures 3.7A & 3.7C and value in 3.7B & 3.7D) against our key dependent variables of accuracy (the likelihood assigned to a correct differential if provided, figures 3.7A & 3.7B) and change in confidence (difference between final confidence and initial confidence, figures 3.7C & 3.7D). Information Sought refers to the proportion of available information sought across cases. Information Value refers to the sum of all mean information values across all 6 cases for a given participant. All data points are for a single participant where variables are averaged across all 6 cases they completed.

Whilst we do not find evidence that the amount of information sought is predictive of accuracy, it may be that there are identifiable 'fingerprints' reflected in information seeking patterns that differentiate between high and low accuracy diagnosticians. If this is the case, participants who are high and low accuracy participants could be predicted based on their information seeking patterns.

In order to test this, we investigate whether information seeking is predictive of participants who are higher or lower in their diagnostic accuracy using binary classification and receiver operating characteristic (ROC) analysis. ROC is a form of analysis that assesses how well a model performs at predicting a binary outcome (in this case, whether a case was performed by a high or low performing participant). We trained a binary classification algorithm using a generalised logistic regression (GLM) model with Leave One Out Cross-Validation (LOOCV) to identify if participants exhibited high or low accuracy based on the information they sought. LOOCV is where our classifier is trained on all data except one case to ask if, based on the learnt patterns from this data, the classifier is able to predict the participant's accuracy (high or low) on the remaining case. This process is then repeated with each case being left out of training and used as this 'test' case. We first split all cases into two groups by whether they were performed by a high and low Accuracy participant. This was done using a median split by

participants' average Accuracy across the six cases. By doing this, we can look at whether participants who perform better at diagnoses seek information in a markedly different way to participants who performed worse.

When plotting an ROC curve, the area under the curve (AUC) is indicative of how well a model performs at correctly categorising cases. An AUC of 0.5 would signify that our model is performing at chance and is not able to predict participant accuracy in any meaningful way. By plotting an ROC curve for our model, we find an AUC value of 0.72 (plotted in Figure 3.8). When conducting a DeLong test, to test the null hypothesis that the AUC is equal is 0.5 (i.e. that the classifier is unable to differentiate between high and low accuracy participants), we find $p < .001$, indicating that the AUC differs significantly from 0.5 and that the classifier is able to reliably predict high and low accuracy participants.
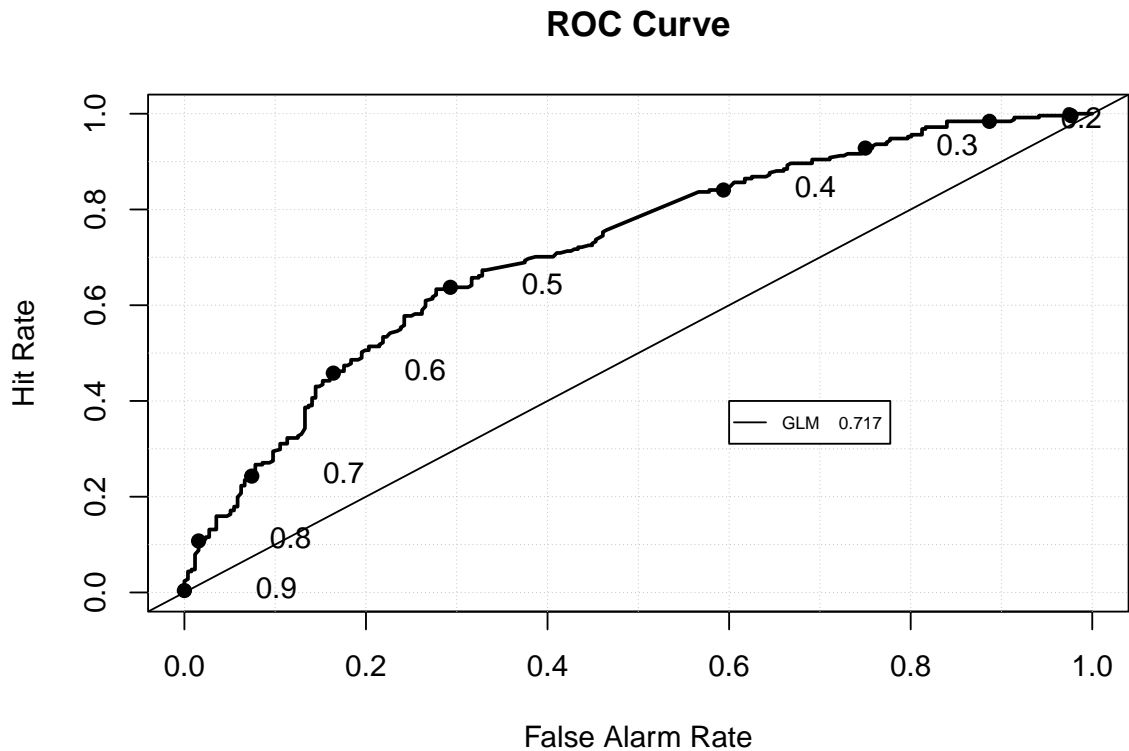
Figure 3.8: Receiver-Operator Characteristic (ROC) curve using a Generalised Linear Model to classify individual cases as being performed by either high or

low accuracy participants. The models are trained on the raw binary predictor variables for each of the 29 available pieces of information, with 0 indicating that the information was not sought for the case and 1 indicating that the information was sought. Participants were sorted as high or low accuracy based on a median split on their average Accuracy value across the six cases.

This result indicates overall that differences in information seeking are indeed predictive of a difference in participant ability at above chance, in terms of high and low accuracy participants seeking different sets of information. Essentially, information seeking patterns are different between high and low accuracy participants. This analysis alone, however, does not tell us what aspects of information seeking in particular are predictive of accuracy. We know from figure 3.7D that seeking more valuable information is associated with higher accuracy. We next seek to characterise the specific differences in information seeking that contribute to higher diagnostic performance.

By looking at the extent to which participants vary the information they seek by case, we can ask the following: is diagnostic accuracy characterised more carefully tailoring information seeking to each individual case, or is it characterised by adopting a more consistent information seeking approach regardless of the patient case? With our measure of how much participants vary in their information seeking across cases, we can see if the variability in information seeking is associated with higher diagnostic accuracy. If higher variability is associated with higher accuracy, this would indicate the former approach being more beneficial (tailored information seeking). If lower variability is associated with higher accuracy, this would indicate the latter approach being more beneficial (consistent information seeking).

We find marginal evidence for a negative association between Information Seeking Variability and Accuracy ($r(83) = -0.22$, 95% CI = [-0.42, -0.01], $p = 0.04$). This data is plotted below in Figure 3.9. We can also look at variability between groups

of participants for each case to ask: are higher performers (in terms of accuracy) more alike in their information seeking than lower performers? To do this, we median split participants into high and low overall accuracy across cases (similar to the ROC analysis in Figure 3.8). We then look at variability in information seeking between participants for each case. If variability is higher, this would indicate that for a given case, participants adopt information seeking approaches that are more different from one another. A plot of variability by case is shown in Figure 3.10. When performing a t-test across conditions, we find that higher performers are more alike in their information seeking (i.e. exhibit lower variability) (t(10) = 2.64, MDiff = 0.36, 0.31, p = 0.02). As can be seen in Figure 3.10, better performing participants show less variability in their information seeking patterns for 5 out of 6 cases, indicating that higher diagnostic accuracy is associated with a consistent 'optimal' information seeking strategy.
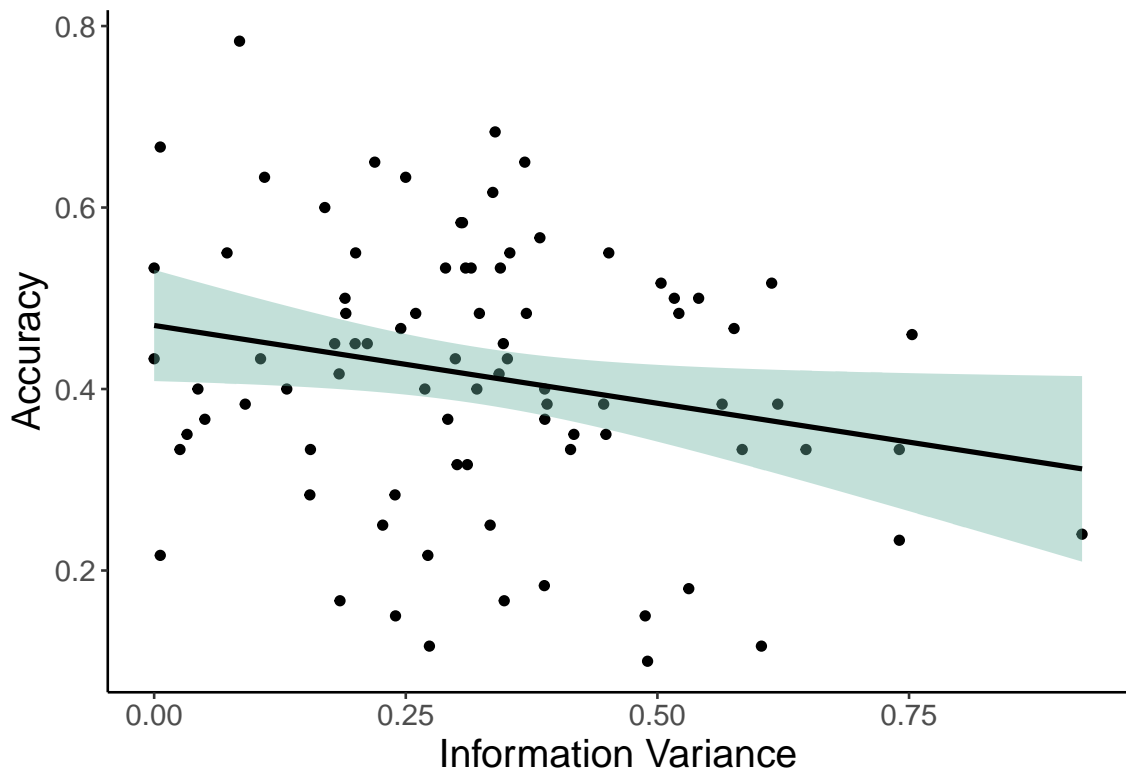
Figure 3.9: Scatter plot showing the relationship between Information Seeking Variability (x-axis, quantified as the average Dice Distance between all pairwise

Online Study

comparisons of cases for a given participant) and Accuracy (y-axis). Each data point represents a single participant.
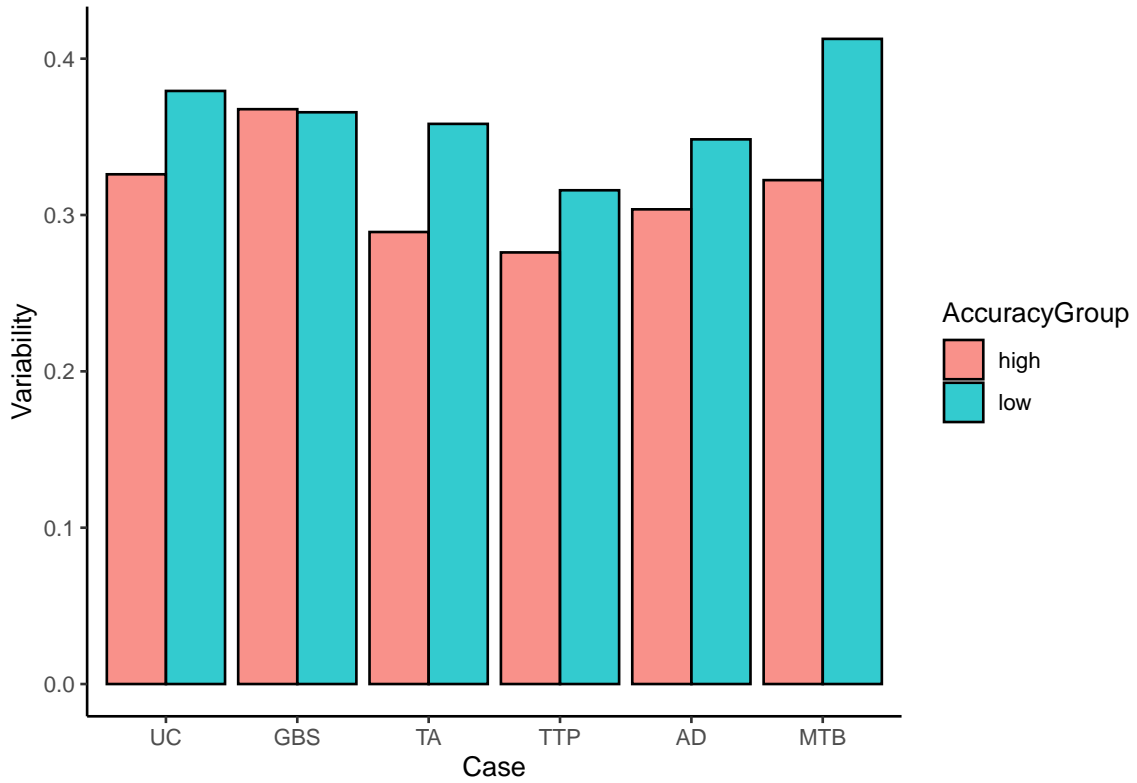
Figure 3.10: Information Seeking Variability (y-axis) for all cases of a given condition (x-axis), with cases median split by participant accuracy. Red bars indicate high performers and blue indicating lower performers. Cases are in descending order (UC = Ulcerative Colitis, GBS = Guillain Barré Syndrome, TA = Temporal Arteritis, TTP = Thrombotic Thrombocytopenic Purpura, AD = Aortic Dissection, MTB = Miliary Tuberculosis) by their average accuracy across participants. Higher variability values signify that participants were less alike one another in terms of the information they sought on a given case.

Given that information seeking variability has a weak negative association with accuracy on our task, we next ask if information seeking is also specific to patient conditions. If so, we would expect the information sought to be predictive of which case the participant is performing. To investigate this, we train a classifier using Penalized Multinomial Regression and Leave One Out Cross Validation (i.e. we train the classifier on all data except one case, and ask if, based on learnt patterns

in the remaining data, whether we are able to predict which case/condition it is based on information seeking patterns). Our input parameters are the available information requests as binary predictors (i.e. to denote whether they were sought on each case or not). The outcome variable of the classifier is the patient condition. We generate model predictions and then look at whether they correctly match the actual condition for that case. Across 510 cases (85 participants performing 6 cases each), the accuracy of the classifier was 57%, which is higher than the chance level of 16.6%. When breaking down accuracy of our classifier by condition, we find accuracy to be above chance across all conditions (see Table 3.2 below).

Condition	Prediction Accuracy
AD	0.65
GBS	0.46
MTB	0.48
TA	0.47
TTP	0.56
UC	0.78

Table 3.2: The accuracy of our multinomial classifier that predicts patient condition for each case based on the information sought/not sought as binary predictors. We then test the accuracy of the classifier by comparing the predicted condition from the model against the actual patient condition for each case. We then split cases by condition to look at accuracy on a case-by-case level. Given that participants perform 6 cases each, accuracy would be 1/6 (16.6%) when at chance.

Taking these findings together, keeping information seeking more constant (i.e. requesting similar high-value information) across cases was found to have an association with accuracy whilst there also being some information that is useful for clinicians to know for patients with specific conditions. To reconcile these, we derive which information requests were most weighted in our classifier models to find which were considered markers of accuracy (by being sought across cases) and

which were considered markers of identifying specific cases. We extract coefficients from the logistic classifier of accuracy (the ROC curve for which was shown in Figure 3.8) and the multinomial classifier (the accuracy of which was depicted above in Table 3.2). We identify the highest weighted information requests as input parameters for each model. The five highest weighted information requests for each model are shown below in Table 3.3. We also show how often each piece of information was sought for each of the cases in Figure 3.11 below. Viewing this figure shows individual tests that are useful for specific cases. For example, an ECG is sought by most participants for the AD (a heart condition) case.

	Accuracy			Condition		
Rank	Test Name	Coefficient	Odds Ratio	Test Name	Coefficient	Odds Ratio
1	FBC	0.45	1.57	Neurologic Exam Record	0.22	1.25
2	Venous Blood Gas	-0.44	0.65	Measure Blood Pressure	0.22	1.25
3	Urine Dipstick	-0.39	0.67	UREA and Electrolytes	0.2	1.22
4	Assess Extremities	0.38	1.47	Rectal Examination	0.17	1.19
5	Other Biochemistry Tests	-0.36	0.7	Urine Dipstick	0.16	1.17

Table 3.3: The five highest weighted parameters (by the absolute value of the coefficient values) for our logistic classifier of participant accuracy (under the "Accuracy" heading above) and our multinomial classifier of patient condition (under the "Condition" heading above). We also show coefficient values and odds ratio values for each parameter.
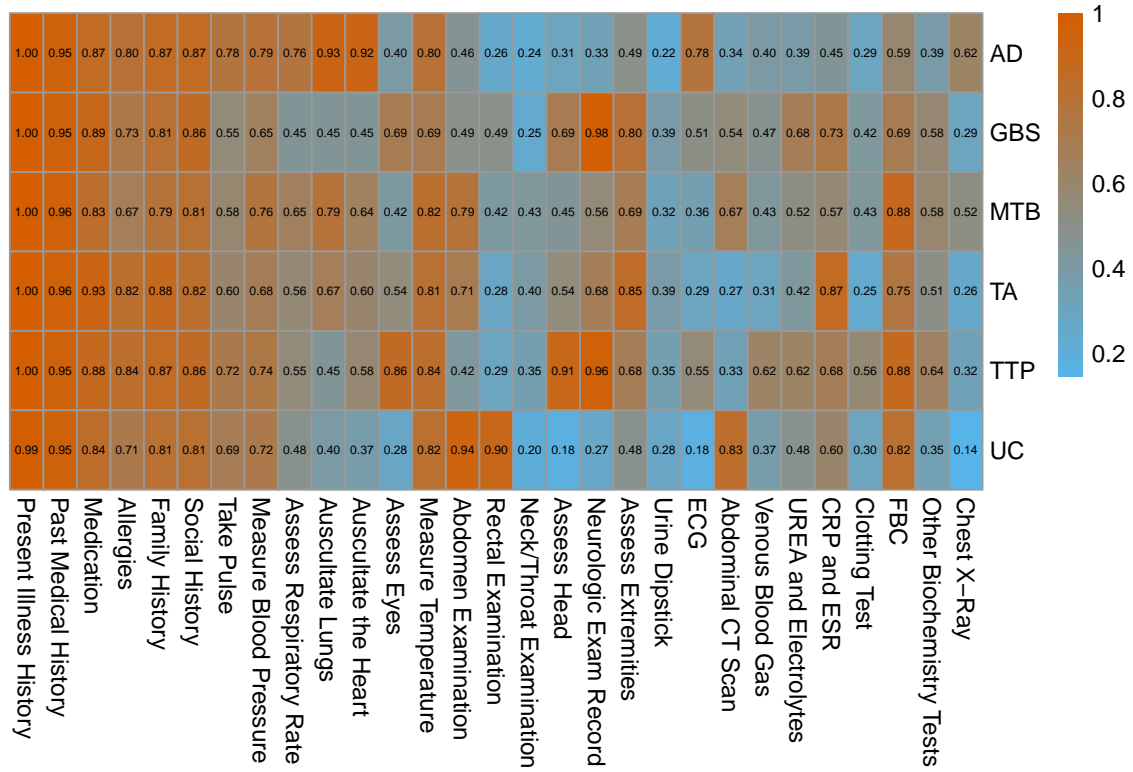
Figure 3.11: Visualisation of the proportion of participants who sought each available piece of information (columns, x-axis) broken down by case (rows, y-axis). Lighter blue colours indicate that fewer participants sought that information for a given case (i.e. towards 0%), whilst lighter orange colours indicate more participants sought that information for a given case (i.e. towards 100%)

Discussion

This study of medical students explored the interplay between confidence, accuracy and information seeking in a novel medical diagnosis task. Using an online interface, we explored how medical students work through diagnostic scenarios, freely seeking information to develop and test sets of possible differentials. Our aim was to look at how different aspects of information seeking impacts both diagnostic confidence and accuracy. The main strength of this study's paradigm is in allowing us to investigate the diagnostic process as it evolves over time and with more information, rather than as a single decision at a single point in time. By tracking how both

confidence and the diagnoses considered by participants changes over time, we gain a better understanding of how the manner in which information sought is key to the diagnostic process and to clinicians' subjective confidence.

Calibration of Confidence

On the question of whether medical students provided confidence judgements that were calibrated to their objective accuracy, we found that students become more accurate across successive stages of information seeking as well more confident. However, cases varied in difficulty as reflected in participant accuracy. In particular, the AD and MTB cases exhibited lower observed accuracy across participants. We observed overconfidence for these two cases, and underconfidence for the UC case (for which accuracy was highest). This indicates a classical hard-easy effect of confidence (Lichtenstein et al. 1977), whereby individuals have a greater tendency to be overconfident for more difficult decisions when compared to easier decision (Merkle 2009). Confidence also increased as participants received more information. However, students reported fairly low confidence overall to treat patients, with an average confidence of below 50% even after receiving all available information. This may indicate that part of ensuring appropriate confidence, or expressions of uncertainty could be related to properly evaluating all possible diagnostic differentials rather than forcing decisions to focus on a single diagnosis, which has been cited previously as a problematic tendency (Redelmeier and Shafir 2023). This may also be a function of undertaking the diagnostic process in isolation (i.e. without being able to discuss with colleagues, as would be the case in naturalistic medical environments). Such a reduction in confidence when making a decision alone rather than in group would be justified from a calibration perspective, as combining medical students' diagnoses has been found to improve accuracy (Kämmer, Hautz, et al. 2017).

Previous work (e.g. Meyer et al. (2013)) has revealed a gap between subjective confidence and objective accuracy. In particular, a general tendency has been

demonstrated for less experienced medical trainees to be underconfident and for more experienced medical professionals to be overconfident (Yang and Thompson 2010). Part of the discrepancy between our findings and past findings could stem from the way that diagnostic uncertainty is expressed by students in this study. Using our primary measure of accuracy, which is obtained by using the likelihood values assigned to correct differentials (if included), we find that accuracy tracks confidence quite closely at each information stage. We note however that our finding of calibrated confidence is highly contingent on the measure of accuracy used. When using a more lenient measure, the proportion of cases where a correct differential was reported (as used in previous papers, (Friedman, Gatti, Franz, et al. 2005; Meyer et al. 2013; Lambe et al. 2018; Küper et al. 2024)), participants were found to be underconfident. When using a stricter measure, the likelihood value assigned to the most likely differential if it is correct, participants are found to be overconfident. Calibration also varied across cases, with participants sometimes showing overconfidence and sometimes showing underconfidence. While we therefore temper our finding of calibration, this has implications for further research that looks at calibration during diagnoses, given that accuracy can be defined in multiple ways when participants record multiple differentials. In addition, our confidence measure is related to the participants' subjective readiness to treat the patient, rather than confidence in the set of differentials. Such a measure of confidence is novel to our study and has not been used in previous studies of diagnostic confidence. This limits the extent to which we can compare accuracy and confidence directly. However, rather than confidence being a subjective judgement, we connect to it to clinical action that would be taken by participants if the patient presented were real. This is similar to one paper in which confidence was measured as the subjective likelihood of seeking assistance to reach a diagnosis (Friedman, Gatti, Franz, et al. 2005), with the authors finding that medical students had a lower tendency toward confidence than both medical residents and faculty. When considering this result alongside our own finding of low confidence across medical students, it is possible that tempering overconfidence may stem from

tying judgements to specific clinical actions. Given that medical students lack the experience of more senior clinicians, they may generally be less confident as a result: the lower reported confidence is partly a reflection of their general aptitude/experience with the clinical action being prompted during the confidence judgement. Future work could then measure how confidence relates to specific aspects of the patient care pathway and differences in calibration.

Broadening of Differentials

On the question of whether participants tend to broaden or narrow their differentials with new information, participants exhibited a general pattern of broadening the range of differentials they were considering across successive information seeking stages. In addition, we observed that participants did not tend to remove differentials from consideration despite having the option to do so. This marks a novel finding when situated within past research, which has not studied how the differentials being considered evolves over time. We can interpret this as students being careful not to miss differentials from consideration, indicating a focus on being comprehensive in their generation of differentials rather than a focus on narrowing in on a single diagnosis. It is therefore worth considering whether students are explicitly taught not to disregard diagnoses completely, instead focusing on remaining open-minded to new possibilities for differentials. (Joseph and Patel 1990) found that clinicians with lower domain knowledge generated accurate hypotheses but were unable to differentiate eliminate hypotheses when receiving more information, unlike clinicians with higher domain knowledge who were able to confirm and eliminate hypotheses using the information received. This may help explain the broadening pattern of medical students, as their relative inexperience meant that they were not able to easily eliminate hypotheses.

We also found that the initial breadth of diagnoses considered from the patients' history was predictive of the amount of subsequent information seeking and changes in confidence. We also find that how much participants change the number of differentials they are considering is predictive of information seeking and changes in confidence. Relatedly, information seeking and confidence was associated, such that participants who sought more information tended to increase their confidence more over the course of the diagnoses. However, the amount of information sought was not predictive of diagnostic accuracy, with accuracy instead being associated with seeking more valuable/appropriate information for a given patient condition. When taken together, these findings give an interesting picture of the diagnostic process as we capture it within our task. Our account of how participants approach this task can be summarised as follows (note that this account requires follow-up study to elucidate further):

- Medical students generate an initial set of differentials from the patient history and use this to guide their information seeking.

- With more differentials to consider, students seek more information to 'test' each of these hypotheses.

- Seeking more information increases the likelihood that new differentials are brought to mind, resulting in more differentials being added as under consideration.

- When participants have more information and have considered a wider range of differentials, they are likely to increase their confidence due to being more comprehensive (i.e. considering more differentials) during their thought process.

- With more differentials being considered, participants are more likely to consider a 'correct' differential. However, considering a larger set of differentials makes it more difficult to focus on finding a differential that is most likely.

Aspects of Information Seeking

Given the flexibility afforded by our paradigm, we are able to monitor fine-grained aspects of how participants seek information. We find that the accuracy and confidence gained over the course of cases was related to the quality of the information sought. We also find that higher accuracy was associated with less variability in information seeking (i.e. seeking a similar set of information regardless of the patient case). Higher accuracy participants were found to be more alike in their information seeking compared to lower accuracy participants. Putting these findings together, we can surmise that each patient condition has associated valuable pieces of information that are worth seeking, but that there is a consistent set of information that accurate participants tend to seek across cases. When combined, each case can be seen to have an 'optimal' set of information that participants should seek. In addition, while seeking more information may increase confidence, having more information may be problematic for weighing up differentials against each other. This is because it can be harder to synthesise more information into a cohesive account of the patient. While past work has called for greater standardisation within healthcare (Wears 2015), what seems to constitute accurate diagnoses in our task is a degree of standardisation with certain selectivity of information given the patient condition. As depicted in Table 3.3, certain information is useful regardless of patient condition whilst others are useful for specific medical conditions. While we show certain tests/examinations as being most useful across patient cases or for specific cases, we recommend caution in interpreting these as representative of all diagnostic decisions outside of this task. These specific information requests were found to be useful for our task, but may not generalise to other patient conditions or diagnostic decisions.

Follow-up Think-Aloud Study

Given these results, we know what information is sought by medical students, but only have a limited insight into why they sought certain information and how

it directly affected the diagnosis they provided. For one, are all students using a similar decision making process when making diagnoses? As of now, we are inferring the participants' thought processes from data of their differentials and information seeking without context of how they are thinking about the task. One possibility is that there are differences in how medical students approach diagnoses that stem from differing reasoning strategies, which we cannot infer from this current dataset. In order to ascertain this, we would need to record the students' thought processes as they are doing the task. To this end, we conduct a follow-up study using a similar diagnostic paradigm conducted in-person where students think out loud as they make diagnoses.

Coderre et al. (2003) used a think-aloud paradigm to characterise distinct diagnostic reasoning strategies: a "hypothetico-deductive" strategy that is closest to the idealised process of elimination that is the typical characterisation of diagnosis, a "pattern matching" strategy where clinicians draw similarities between the current patient and either a past patient or prototypical case of a particular condition, and a "scheme-inductive" strategy in which clinicians follow a structured framework for diagnoses (e.g. a surgical sieve, that considers each pathophysiological system in turn). Of interest to our work was whether we would observe similar variation in reasoning strategies in our medical trainees and, if so, how these strategies related to patterns of information seeking and confidence We hypothesise not only that we can detect reasoning strategies based on the verbalisations of participants' thought process, but that different reasoning strategies for generating differentials are useful for some cases more than others. We also hypothesise that information seeking and changes in confidence vary as a function of the reasoning strategy employed.

Given the recording of qualitative data during this task, we can understand both how medical students are thinking about diagnoses as they are making them but also how they reflect on their thought process outside of the task. This detection of reasoning strategies, if successful, can then subsequently be used to detect the

same reasoning strategies in this online study dataset (where we do not have access to the participants' thought process) based on the information sought. Given the higher sample size afforded by the online study, we can more robustly look at differences between reasoning strategies and whether they can tell us about what makes more accurate and more confident diagnoses.

References

Abujudeh, Hani H. et al. (Feb. 2011). "Abdominopelvic CT Increases Diagnostic Certainty and Guides Management Decisions: A Prospective Investigation of 584 Patients in a Large Academic Medical Center". In: *American Journal of Roentgenology* 196.2. Number: 2 Publisher: American Roentgen Ray Society, pp. 238–243. DOI: 10.2214/AJR.10.4467. URL: https://ajronline.org/doi/full/10.2214/AJR.10.4467 (visited on 10/07/2024).

Adams, J. Stacy (1961). "Reduction of cognitive dissonance by seeking consonant information". In: *The Journal of Abnormal and Social Psychology* 62.1. Number: 1 Place: US Publisher: American Psychological Association, pp. 74–78. DOI: 10.1037/h0047029.

Ais, Joaquín et al. (Jan. 2016). "Individual consistency in the accuracy and distribution of confidence judgments". In: *Cognition* 146, pp. 377–386. DOI: 10.1016/j.cognition.2015.10.006. URL: https://www.sciencedirect.com/science/article/pii/S0010027715300846 (visited on 10/07/2024).

Aiyer, Sriraj et al. (July 2023). "Outcomes Affect Evaluations of Decision Quality: Replication and Extensions of Baron and Hershey's (1988) Outcome Bias Experiment 1". en. In: *International Review of Social Psychology* 36.1, p. 12. DOI: 10.5334/irsp.751. URL: https://account.rips-irsp.com/index.php/up-j-irsp/article/view/751 (visited on 11/29/2024).

Akaishi, Rei et al. (Jan. 2014). "Autonomous Mechanism of Internal Choice Estimate Underlies Decision Inertia". English. In: *Neuron* 81.1. Number: 1 Publisher: Elsevier, pp. 195–206. DOI: 10.1016/j.neuron.2013.10.018. URL: https://www.cell.com/neuron/abstract/S0896-6273(13)00921-5 (visited on 11/29/2024).

Albrechtsen, Simon S. et al. (Mar. 2022). "Impact of MRI on decision-making in ICU patients with disorders of consciousness". In: *Behavioural Brain Research* 421, p. 113729. DOI: 10.1016/j.bbr.2021.113729. URL: https://www.sciencedirect.com/science/article/pii/S0166432821006173 (visited on 10/07/2024).

Alby, Francesca, Cristina Zuccermaglio, and Mattia Baruzzo (Jan. 2015). "Diagnostic Decision Making in Oncology: Creating Shared Knowledge and Managing Complexity". In: *Mind, Culture, and Activity* 22.1. Number: 1 Publisher: Routledge _eprint: https://doi.org/10.1080/10749039.2014.981642, pp. 4–22. DOI: 10.1080/10749039.2014.981642. URL: https://doi.org/10.1080/10749039.2014.981642 (visited on 11/20/2024).

Alicke, Mark D. and Olesya Govorun (2005). "The Better-Than-Average Effect". In: *The Self in Social Judgment.* Num Pages: 22. Psychology Press.

References

Arocha, José F. and Vimla L. Patel (Oct. 1995). "Novice Diagnostic Reasoning in Medicine: Accounting for Evidence". EN. In: *The Journal of the Learning Sciences*. Publisher: Lawrence Erlbaum Associates, Inc. DOI: 10.1207/s15327809jls0404_1. URL: https://www.tandfonline.com/doi/abs/10.1207/s15327809jls0404_1 (visited on 11/18/2024).

Aron, Adrian et al. (2024). "Diagnostic momentum in physical therapy clinical reasoning". en. In: *Journal of Evaluation in Clinical Practice* 30.1. Number: 1 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jep.13884, pp. 73–81. DOI: 10.1111/jep.13884. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jep.13884 (visited on 11/26/2024).

Baker, G. Ross et al. (May 2004). "The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada". en. In: *CMAJ* 170.11. Number: 11 Publisher: CMAJ Section: Research article, pp. 1678–1686. DOI: 10.1503/cmaj.1040498. URL: https://www.cmaj.ca/content/170/11/1678 (visited on 10/07/2024).

Balsdon, Tarryn, Valentin Wyart, and Pascal Mamassian (Apr. 2020). "Confidence controls perceptual evidence accumulation". en. In: *Nature Communications* 11.1. Number: 1 Publisher: Nature Publishing Group, p. 1753. DOI: 10.1038/s41467-020-15561-w. URL: https://www.nature.com/articles/s41467-020-15561-w (visited on 11/26/2024).

Baron, Jonathan and John C. Hershey (1988). "Outcome bias in decision evaluation". In: *Journal of Personality and Social Psychology* 54.4. Number: 4 Place: US Publisher: American Psychological Association, pp. 569–579. DOI: 10.1037/0022-3514.54.4.569.

Ben-Assuli, Ofir et al. (June 2015). "Improving diagnostic accuracy using EHR in emergency departments: A simulation-based study". In: *Journal of Biomedical Informatics* 55, pp. 31–40. DOI: 10.1016/j.jbi.2015.03.004. URL: https://www.sciencedirect.com/science/article/pii/S1532046415000477 (visited on 10/07/2024).

Benvenuto-Andrade, Cristiane et al. (May 2006). "Level of Confidence in Diagnosis: Clinical Examination Versus Dermoscopy Examination". en-US. In: *Dermatologic Surgery* 32.5. Number: 5, p. 738. URL: https://journals.lww.com/dermatologicsurgery/abstract/2006/05000/level_of_confidence_in_diagnosis__clinical.34.aspx (visited on 10/07/2024).

Bergl, Paul A. et al. (Feb. 2024). "Factors influencing diagnostic accuracy among intensive care unit clinicians – an observational study". en. In: *Diagnosis* 11.1. Number: 1 Publisher: De Gruyter, pp. 31–39. DOI: 10.1515/dx-2023-0026. URL: https://www.degruyter.com/document/doi/10.1515/dx-2023-0026/html (visited on 10/07/2024).

Berner, Eta S. and Mark L. Graber (May 2008). "Overconfidence as a Cause of Diagnostic Error in Medicine". In: *The American Journal of Medicine*. Diagnostic Error: Is Overconfidence the Problem 121.5, Supplement. Number: 5, Supplement, S2–S23. DOI: 10.1016/j.amjmed.2008.01.001. URL: https://www.sciencedirect.com/science/article/pii/S0002934308000405 (visited on 10/07/2024).

References

Berner, Eta S. and Richard S. Maisiak (Sept. 1999). "Influence of Case and Physician Characteristics on Perceptions of Decision Support Systems". In: *Journal of the American Medical Informatics Association* 6.5. Number: 5, pp. 428–434. DOI: 10.1136/jamia.1999.0060428. URL: https://doi.org/10.1136/jamia.1999.0060428 (visited on 10/07/2024).

Berwick, Donald M. and Andrew D. Hackbarth (Apr. 2012). "Eliminating Waste in US Health Care". In: *JAMA* 307.14. Number: 14, pp. 1513–1516. DOI: 10.1001/jama.2012.362. URL: https://doi.org/10.1001/jama.2012.362 (visited on 10/07/2024).

Bhise, Viraj, Ashley N.D. Meyer, et al. (Aug. 2017). "Errors in Diagnosis of Spinal Epidural Abscesses in the Era of Electronic Health Records". en. In: *The American Journal of Medicine* 130.8. Number: 8, pp. 975–981. DOI: 10.1016/j.amjmed.2017.03.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S0002934317303236 (visited on 10/05/2024).

Bhise, Viraj, Suja S. Rajan, et al. (Jan. 2018). "Defining and Measuring Diagnostic Uncertainty in Medicine: A Systematic Review". en. In: *Journal of General Internal Medicine* 33.1. Number: 1, pp. 103–115. DOI: 10.1007/s11606-017-4164-1. URL: https://doi.org/10.1007/s11606-017-4164-1 (visited on 10/07/2024).

Boldt, Annika and Nick Yeung (Feb. 2015). "Shared Neural Markers of Decision Confidence and Error Detection". en. In: *Journal of Neuroscience* 35.8. Number: 8 Publisher: Society for Neuroscience Section: Brief Communications, pp. 3478–3484. DOI: 10.1523/JNEUROSCI.0797-14.2015. URL: https://www.jneurosci.org/content/35/8/3478 (visited on 10/07/2024).

Brannon, Laura A and Kimi L Carson (Nov. 2003). "Nursing expertise and information structure influence medical decision making". In: *Applied Nursing Research* 16.4. Number: 4, pp. 287–290. DOI: 10.1016/j.apnr.2003.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0897189703000788 (visited on 10/07/2024).

Brezis, Mayer et al. (Dec. 2019). "Does Physician's Training Induce Overconfidence That Hampers Disclosing Errors?" en-US. In: *Journal of Patient Safety* 15.4. Number: 4, p. 296. DOI: 10.1097/PTS.0000000000000255. URL: https://journals.lww.com/journalpatientsafety/abstract/2019/12000/does_physician_s_training_induce_overconfidence.10.aspx (visited on 10/07/2024).

Calman, Neil, R Hyman, and W Licht (Aug. 1992). "Variability in Consultation Rates and Practitioner Level of Diagnostic Certainty". In: *The Journal of family practice* 35, pp. 31–8.

Carpenter, Christopher R., Ali S. Raja, and Michael D. Brown (2015). "Overtesting and the Downstream Consequences of Overtreatment: Implications of "Preventing Overdiagnosis" for Emergency Medicine". en. In: *Academic Emergency Medicine* 22.12. Number: 12 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/acem.12820, pp. 1484–1492. DOI: 10.1111/acem.12820. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/acem.12820 (visited on 10/07/2024).

Chapman, Gretchen B., George R. Bergus, and Arthur S. Elstein (1996). "Order of information affects clinical judgment". In: *Journal of Behavioral Decision Making*

References

9.3, pp. 201–211. DOI:
`10.1002/(SICI)1099-0771(199609)9:3<201::AID-BDM229>3.0.CO;2-J`.

Charles, Lucie and Nick Yeung (2019). "Dynamic sources of evidence supporting
confidence judgments and error detection". In: *Journal of Experimental Psychology:
Human Perception and Performance* 45.1. Number: 1 Place: US Publisher:
American Psychological Association, pp. 39–52. DOI: `10.1037/xhp0000583`.

Chartan, Corey et al. (Oct. 2019). "Isolating red flags to enhance diagnosis (I-RED):
An experimental vignette study". In: *International Journal for Quality in Health
Care* 31.8. Number: 8, G97–G102. DOI: `10.1093/intqhc/mzz082`. URL:
`https://doi.org/10.1093/intqhc/mzz082` (visited on 10/07/2024).

Cheng, Joey T. et al. (2021). "The social transmission of overconfidence". In: *Journal of
Experimental Psychology: General* 150.1. Number: 1 Place: US Publisher: American
Psychological Association, pp. 157–186. DOI: `10.1037/xge0000787`.

Clayton, Dayna A. et al. (Feb. 2023). "Are Pathologists Self-Aware of Their Diagnostic
Accuracy? Metacognition and the Diagnostic Process in Pathology". en. In: *Medical
Decision Making* 43.2. Number: 2 Publisher: SAGE Publications Inc STM,
pp. 164–174. DOI: `10.1177/0272989X221126528`. URL:
`https://doi.org/10.1177/0272989X221126528` (visited on 10/07/2024).

Coderre, S et al. (2003). "Diagnostic reasoning strategies and diagnostic success". en. In:
Medical Education 37.8. Number: 8 _eprint:
https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2923.2003.01577.x,
pp. 695–703. DOI: `10.1046/j.1365-2923.2003.01577.x`. URL: `https:
//onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2923.2003.01577.x`
(visited on 10/07/2024).

Costa Filho, Galileu B. et al. (Aug. 2019). "Effects of deliberate reflection on diagnostic
accuracy, confidence and diagnostic calibration in dermatology". en. In: *Perspectives
on Medical Education* 8.4. Number: 4, pp. 230–236. DOI:
`10.1007/s40037-019-0522-5`. URL:
`https://doi.org/10.1007/s40037-019-0522-5` (visited on 10/07/2024).

Croskerry, Pat (Sept. 2009). "Clinical cognition and diagnostic error: applications of a
dual process model of reasoning". en. In: *Advances in Health Sciences Education*
14.1. Number: 1, pp. 27–35. DOI: `10.1007/s10459-009-9182-2`. URL:
`https://doi.org/10.1007/s10459-009-9182-2` (visited on 11/19/2024).

— (June 2013). "From mindless to mindful practice–cognitive bias and clinical decision
making". eng. In: *The New England Journal of Medicine* 368.26. Number: 26,
pp. 2445–2448. DOI: `10.1056/NEJMp1303712`.

Croskerry, Pat et al. (Feb. 2014). "Deciding About Fast and Slow Decisions". en-US. In:
Academic Medicine 89.2. Number: 2, p. 197. DOI: `10.1097/ACM.0000000000000121`.
URL: `https://journals.lww.com/academicmedicine/fulltext/2014/02000/
Deciding_About_Fast_and_Slow_Decisions.7.aspx` (visited on 11/19/2024).

Crowley, Rebecca S. et al. (Aug. 2013). "Automated detection of heuristics and biases
among pathologists in a computer-based system". en. In: *Advances in Health
Sciences Education* 18.3. Number: 3, pp. 343–363. DOI:
`10.1007/s10459-012-9374-z`. URL:
`https://doi.org/10.1007/s10459-012-9374-z` (visited on 10/07/2024).

Cutler, Brian L., Steven D. Penrod, and Hedy R. Dexter (1989). "The eyewitness, the
expert psychologist, and the jury." en. In: *Law and Human Behavior* 13.3. Number:

References

3, pp. 311–332. DOI: `10.1007/BF01067032`. URL: `https://doi.apa.org/doi/10.1007/BF01067032` (visited on 10/07/2024).

Desender, Kobe, Annika Boldt, and Nick Yeung (May 2018). "Subjective Confidence Predicts Information Seeking in Decision Making". en. In: *Psychological Science* 29.5. Number: 5 Publisher: SAGE Publications Inc, pp. 761–778. DOI: `10.1177/0956797617744771`. URL: `https://doi.org/10.1177/0956797617744771` (visited on 10/07/2024).

Dice, Lee R. (1945). "Measures of the Amount of Ecologic Association Between Species". In: *Ecology* 26.3. Number: 3 Publisher: [Wiley, Ecological Society of America], pp. 297–302. DOI: `10.2307/1932409`. URL: `https://www.jstor.org/stable/1932409` (visited on 11/06/2024).

Dreiseitl, Stephan and Michael Binder (Jan. 2005). "Do physicians value decision support? A look at the effect of decision support systems on physician opinion". In: *Artificial Intelligence in Medicine* 33.1. Number: 1, pp. 25–30. DOI: `10.1016/j.artmed.2004.07.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0933365704001071` (visited on 10/07/2024).

Dunlosky, John (2013). "Strengthening the Student Toolbox: Study Strategies to Boost Learning". en. In: *American Educator* 37.3. Number: 3 Publisher: American Federation of Teachers ERIC Number: EJ1021069, pp. 12–21. URL: `https://eric.ed.gov/?id=EJ1021069` (visited on 10/07/2024).

Eva, Kevin W (2001). "The influence of differentially processing evidence on diagnostic decision-making". PhD thesis. McMaster University. URL: `https://macsphere.mcmaster.ca/bitstream/11375/7187/1/fulltext.pdf` (visited on 12/02/2024).

Fawver, Bradley et al. (2020). "Seeing isn't necessarily believing: Misleading contextual information influences perceptual-cognitive bias in radiologists". In: *Journal of Experimental Psychology: Applied* 26.4. Number: 4 Place: US Publisher: American Psychological Association, pp. 579–592. DOI: `10.1037/xap0000274`.

Featherston, Rebecca et al. (Oct. 2020). "Decision making biases in the allied health professions: A systematic scoping review". en. In: *PLOS ONE* 15.10. Number: 10 Publisher: Public Library of Science, e0240716. DOI: `10.1371/journal.pone.0240716`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240716` (visited on 11/19/2024).

Fenwick, Samuel (1891). *Medical Diagnosis*. en. Google-Books-ID: GvRaAAAAQAAJ. J. and A. Churchill.

Fernández-Aguilar, Carmen et al. (2022). "Use of heuristics during the clinical decision process from family care physicians in real conditions". en. In: *Journal of Evaluation in Clinical Practice* 28.1. Number: 1 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jep.13608, pp. 135–141. DOI: `10.1111/jep.13608`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/jep.13608` (visited on 10/07/2024).

Feyzi-Behnagh, Reza et al. (Mar. 2014). "Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system". en. In: *Instructional Science* 42.2. Number: 2, pp. 159–181. DOI:

References

`10.1007/s11251-013-9275-4`. URL:
https://doi.org/10.1007/s11251-013-9275-4 (visited on 10/07/2024).

Fleming, Stephen M. and Nathaniel D. Daw (Jan. 2017). "Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation." en. In: *Psychological Review* 124.1. Number: 1, pp. 91–114. DOI: `10.1037/rev0000045`. URL: https://doi.apa.org/doi/10.1037/rev0000045 (visited on 10/07/2024).

Friedman, Charles, Guido Gatti, Arthur Elstein, et al. (2001). "Are Clinicians Correct When They Believe They are Correct? Implications for Medical Decision Support". In: *MEDINFO 2001*. IOS Press, pp. 454–458. DOI: `10.3233/978-1-60750-928-8-454`. URL: https://ebooks.iospress.nl/doi/10.3233/978-1-60750-928-8-454 (visited on 11/12/2024).

Friedman, Charles P., Guido G. Gatti, Timothy M. Franz, et al. (Apr. 2005). "Do physicians know when their diagnoses are correct?" en. In: *Journal of General Internal Medicine* 20.4. Number: 4, pp. 334–339. DOI: `10.1111/j.1525-1497.2005.30145.x`. URL: https://doi.org/10.1111/j.1525-1497.2005.30145.x (visited on 10/07/2024).

Garbayo, Luciana S. et al. (Mar. 2023). "A metacognitive confidence calibration (MCC) tool to help medical students scaffold diagnostic reasoning in decision-making during high-fidelity patient simulations". In: *Advances in Physiology Education* 47.1. Number: 1 Publisher: American Physiological Society, pp. 71–81. DOI: `10.1152/advan.00156.2021`. URL: https://journals.physiology.org/doi/full/10.1152/advan.00156.2021 (visited on 10/07/2024).

Gehlbach, Hunter, Carly D. Robinson, and Angus Fletcher (Oct. 2024). "The illusion of information adequacy". en. In: *PLOS ONE* 19.10. Number: 10 Publisher: Public Library of Science, e0310216. DOI: `10.1371/journal.pone.0310216`. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0310216 (visited on 12/10/2024).

Gherman, Sabina and Marios G. Philiastides (Feb. 2015). "Neural representations of confidence emerge from the process of decision formation during perceptual choices". In: *NeuroImage* 106, pp. 134–143. DOI: `10.1016/j.neuroimage.2014.11.036`. URL: https://www.sciencedirect.com/science/article/pii/S1053811914009537 (visited on 10/07/2024).

Gigerenzer, Gerd (Jan. 2008). "Why Heuristics Work". en. In: *Perspectives on Psychological Science* 3.1. Number: 1 Publisher: SAGE Publications Inc, pp. 20–29. DOI: `10.1111/j.1745-6916.2008.00058.x`. URL: https://doi.org/10.1111/j.1745-6916.2008.00058.x (visited on 11/19/2024).

Gilhooly, K. J. (1990). "Cognitive psychology and medical diagnosis". en. In: *Applied Cognitive Psychology* 4.4. Number: 4 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.2350040404, pp. 261–272. DOI: `10.1002/acp.2350040404`. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.2350040404 (visited on 10/28/2024).

Graber, Mark L. et al. (July 2012). "Cognitive interventions to reduce diagnostic error: a narrative review". en. In: *BMJ Quality & Safety* 21.7. Number: 7 Publisher: BMJ Publishing Group Ltd Section: Narrative review, pp. 535–557. DOI:

`10.1136/bmjqs-2011-000149`. URL:
`https://qualitysafety.bmj.com/content/21/7/535` (visited on 10/07/2024).

Greig, Paul R., Helen Higham, and Emma Vaux (Sept. 2015). "Lack of standardisation between specialties for human factors content in postgraduate training: an analysis of specialty curricula in the UK". en. In: *BMJ Quality & Safety* 24.9. Number: 9 Publisher: BMJ Publishing Group Ltd Section: Short report, pp. 558–560. DOI: `10.1136/bmjqs-2014-003684`. URL:
`https://qualitysafety.bmj.com/content/24/9/558` (visited on 10/07/2024).

Groopman, Jerome (2010). *How Doctors Think*. en. Scribe Publications.

Gruppen, Larry D., Fredric M. Wolf, and John E. Billi (Dec. 1991). "Information Gathering and Integration as Sources of Error in Diagnostic Decision Making". en. In: *Medical Decision Making* 11.4. Number: 4 Publisher: SAGE Publications Inc STM, pp. 233–239. DOI: `10.1177/0272989X9101100401`. URL:
`https://doi.org/10.1177/0272989X9101100401` (visited on 10/07/2024).

Gupta, Ashwin B. et al. (Oct. 2023). "Associations Between Hospitalist Shift Busyness, Diagnostic Confidence, and Resource Utilization: A Pilot Study". en-US. In: *Journal of Patient Safety* 19.7. Number: 7, p. 447. DOI: `10.1097/PTS.0000000000001157`. URL: `https://journals.lww.com/journalpatientsafety/abstract/2023/10000/associations_between_hospitalist_shift_busyness,.5.aspx?context=latestarticles` (visited on 10/07/2024).

Hageman, Michiel G. J. S. et al. (Dec. 2013). "Surgeon Confidence in an Outpatient Setting". en. In: *HAND* 8.4. Number: 4 Publisher: SAGE Publications Inc, pp. 430–433. DOI: `10.1007/s11552-013-9533-6`. URL:
`https://doi.org/10.1007/s11552-013-9533-6` (visited on 10/07/2024).

Hall, Katherine H (2002). "Reviewing intuitive decision-making and uncertainty: the implications for medical education". en. In: *Medical Education* 36.3. Number: 3 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2923.2002.01140.x, pp. 216–224. DOI: `10.1046/j.1365-2923.2002.01140.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2923.2002.01140.x` (visited on 10/07/2024).

Hampton, J R et al. (May 1975). "Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients." en. In: *BMJ* 2.5969. Number: 5969, pp. 486–489. DOI: `10.1136/bmj.2.5969.486`. URL: `https://www.bmj.com/lookup/doi/10.1136/bmj.2.5969.486` (visited on 11/18/2024).

Han, Paul K. J., William M. P. Klein, and Neeraj K. Arora (Nov. 2011). "Varieties of Uncertainty in Health Care: A Conceptual Taxonomy". en. In: *Medical Decision Making* 31.6. Number: 6 Publisher: SAGE Publications Inc STM, pp. 828–838. DOI: `10.1177/0272989X10393976`. URL: `https://doi.org/10.1177/0272989X10393976` (visited on 11/19/2024).

Hancock, Jason and Karen Mattick (Feb. 2020). "Tolerance of ambiguity and psychological well-being in medical training: A systematic review". In: *Medical Education* 54.2. Number: 2 Publisher: John Wiley and Sons, Ltd, pp. 125–137. DOI: `10.1111/medu.14031`. URL: `https://asmepublications.onlinelibrary.wiley.com/doi/full/10.1111/medu.14031` (visited on 11/20/2024).

References

Harvey, Christopher J. et al. (Apr. 1999). "Evacuation Proctography: A Prospective Study of Diagnostic and Therapeutic Effects". In: *Radiology* 211.1. Number: 1 Publisher: Radiological Society of North America, pp. 223–227. DOI: 10.1148/radiology.211.1.r99mr16223. URL: https://pubs.rsna.org/doi/abs/10.1148/radiology.211.1.r99mr16223 (visited on 10/07/2024).

Hausmann, Daniel et al. (Apr. 2019). "Sensitivity for multimorbidity: The role of diagnostic uncertainty of physicians when evaluating multimorbid video case-based vignettes". en. In: *PLOS ONE* 14.4. Number: 4 Publisher: Public Library of Science, e0215049. DOI: 10.1371/journal.pone.0215049. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0215049 (visited on 10/07/2024).

Hautz, Wolf E, Sebastian Schubert, et al. (2019). "Accuracy of self-monitoring: does experience, ability or case difficulty matter?" en. In: *Medical Education* 53.7. Number: 7 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/medu.13801, pp. 735–744. DOI: 10.1111/medu.13801. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/medu.13801 (visited on 10/07/2024).

Hautz, Wolf E., Juliane E. Kämmer, et al. (May 2019). "Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room". en. In: *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 27.1. Number: 1, p. 54. DOI: 10.1186/s13049-019-0629-z. URL: https://doi.org/10.1186/s13049-019-0629-z (visited on 10/07/2024).

Heereman, J. and P. Walla (Dec. 2011). "Stress, Uncertainty and Decision Confidence". en. In: *Applied Psychophysiology and Biofeedback* 36.4. Number: 4, pp. 273–279. DOI: 10.1007/s10484-011-9167-9. URL: https://doi.org/10.1007/s10484-011-9167-9 (visited on 10/07/2024).

Heller, Rachael F., Herbert D. Saltzstein, and William B. Caspe (Feb. 1992). "Heuristics in Medical and Non-Medical Decision-Making". en. In: *The Quarterly Journal of Experimental Psychology Section A* 44.2. Number: 2 Publisher: SAGE Publications, pp. 211–235. DOI: 10.1080/02724989243000019. URL: https://doi.org/10.1080/02724989243000019 (visited on 10/07/2024).

Hémon, Brivael et al. (Aug. 2020). "Speaking Up About Errors in Routine Clinical Practice: A Simulation-Based Intervention With Nursing Students". In: *Clinical Simulation in Nursing*. Patient Safety 45, pp. 32–41. DOI: 10.1016/j.ecns.2020.03.003. URL: https://www.sciencedirect.com/science/article/pii/S1876139920300207 (visited on 10/07/2024).

Henmon, V. A. C. (1911). "The relation of the time of a judgment to its accuracy". In: *Psychological Review* 18.3. Number: 3 Place: US Publisher: The Review Publishing Company, pp. 186–201. DOI: 10.1037/h0074579.

Higgs, Joy et al., eds. (2019). *Clinical reasoning in the health professions*. en. Fourth edition. Edinburgh London New York: Elsevier.

Higham, Helen et al. (Aug. 2019). "Observer-based tools for non-technical skills assessment in simulated and real clinical environments in healthcare: a systematic review". en. In: *BMJ Quality & Safety* 28.8. Number: 8 Publisher: BMJ Publishing Group Ltd Section: Systematic review, pp. 672–686. DOI:

References

10.1136/bmjqs-2018-008565. URL:
https://qualitysafety.bmj.com/content/28/8/672 (visited on 10/07/2024).

Hillson, Steven D., Donald P. Connelly, and Yuanli Liu (June 1995). "The Effects of Computer-assisted Electrocardiographic Interpretation on Physicians' Diagnostic Decisions". en. In: *Medical Decision Making* 15.2. Number: 2 Publisher: SAGE Publications Inc STM, pp. 107–112. DOI: 10.1177/0272989X9501500202. URL: https://doi.org/10.1177/0272989X9501500202 (visited on 10/07/2024).

Hunt, Laurence T. et al. (Nov. 2016). "Approach-Induced Biases in Human Information Sampling". en. In: *PLOS Biology* 14.11. Number: 11 Publisher: Public Library of Science, e2000638. DOI: 10.1371/journal.pbio.2000638. URL: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2000638 (visited on 10/07/2024).

Investigation report (June 2021). *Investigation report: Delayed Recognition of Acute Aortic Dissection*. en-gb. URL:
https://www.hssib.org.uk/patient-safety-investigations/delayed-recognition-of-acute-aortic-dissection/investigation-report/ (visited on 11/20/2024).

Investigation report (Mar. 2022). *Investigation report: Missed detection of lung cancer on chest X-rays of patients being seen in primary care*. en-gb. URL:
https://www.hssib.org.uk/patient-safety-investigations/missed-detection-of-lung-cancer-on-chest-x-rays-of-patients-being-seen-in-primary-care/investigation-report/ (visited on 11/20/2024).

Jans, Carley et al. (June 2023). "Examining the impact of virtual reality on clinical decision making – An integrative review". en. In: *Nurse Education Today* 125, p. 105767. DOI: 10.1016/j.nedt.2023.105767. URL:
https://linkinghub.elsevier.com/retrieve/pii/S0260691723000618 (visited on 10/11/2024).

Jaspan, Oren et al. (Mar. 2022). "Improving the Relationship Between Confidence and Competence: Implications for Diagnostic Radiology Training From the Psychology and Medical Literature". In: *Academic Radiology* 29.3. Number: 3, pp. 428–438. DOI: 10.1016/j.acra.2020.12.006. URL:
https://www.sciencedirect.com/science/article/pii/S1076633220306991 (visited on 10/07/2024).

Jonas, Eva et al. (2001). "Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information". In: *Journal of Personality and Social Psychology* 80.4. Number: 4 Place: US Publisher: American Psychological Association, pp. 557–571. DOI: 10.1037/0022-3514.80.4.557.

Jonassen, David H. (Mar. 1997). "Instructional design models for well-structured and Ill-structured problem-solving learning outcomes". en. In: *Educational Technology Research and Development* 45.1. Number: 1, pp. 65–94. DOI: 10.1007/BF02299613. URL: http://link.springer.com/10.1007/BF02299613 (visited on 10/05/2024).

Joseph, Guy-Marie and Vimla L. Patel (Feb. 1990). "Domain Knowledge and Hypothesis Genenation in Diagnostic Reasoning". en. In: *Medical Decision Making* 10.1. Number: 1 Publisher: SAGE Publications Inc STM, pp. 31–44. DOI: 10.1177/0272989X9001000107. URL:
https://doi.org/10.1177/0272989X9001000107 (visited on 11/14/2024).

References

Kaanders, Paula et al. (Apr. 2022). "Humans actively sample evidence to support prior beliefs". In: *eLife* 11. Ed. by Valentin Wyart, Michael J Frank, and Konstantinos Tsetsos. Publisher: eLife Sciences Publications, Ltd, e71768. DOI: 10.7554/eLife.71768. URL: https://doi.org/10.7554/eLife.71768 (visited on 10/07/2024).

Kahneman, Daniel (Oct. 2011). *Thinking, Fast and Slow*. en. Google-Books-ID: ZuKTvERuPG8C. Farrar, Straus and Giroux.

Kämmer, Juliane E., Wolf E. Hautz, et al. (Aug. 2017). "The Potential of Collective Intelligence in Emergency Medicine: Pooling Medical Students' Independent Decisions Improves Diagnostic Performance". en. In: *Medical Decision Making* 37.6. Number: 6 Publisher: SAGE Publications Inc STM, pp. 715–724. DOI: 10.1177/0272989X17696998. URL: https://doi.org/10.1177/0272989X17696998 (visited on 11/12/2024).

Kämmer, Juliane E., Stefan K. Schauber, et al. (2021). "Differential diagnosis checklists reduce diagnostic error differentially: A randomised experiment". en. In: *Medical Education* 55.10. Number: 10 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/medu.14596, pp. 1172–1182. DOI: 10.1111/medu.14596. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/medu.14596 (visited on 10/07/2024).

Katz, Jay (1984). "Why Doctors Don't Disclose Uncertainty". In: *The Hastings Center Report* 14.1. Number: 1 Publisher: [Hastings Center, Wiley], pp. 35–44. DOI: 10.2307/3560848. URL: https://www.jstor.org/stable/3560848 (visited on 10/07/2024).

Kiani, Roozbeh, Leah Corthell, and Michael N. Shadlen (Dec. 2014). "Choice Certainty Is Informed by Both Evidence and Decision Time". English. In: *Neuron* 84.6. Number: 6 Publisher: Elsevier, pp. 1329–1342. DOI: 10.1016/j.neuron.2014.12.015. URL: https://www.cell.com/neuron/abstract/S0896-6273(14)01096-4 (visited on 10/07/2024).

Ko, Yiu Hong et al. (Aug. 2022). "Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements". In: *Cognition* 225, p. 105125. DOI: 10.1016/j.cognition.2022.105125. URL: https://www.sciencedirect.com/science/article/pii/S0010027722001135 (visited on 10/07/2024).

Kohn, Linda T., Janet M. Corrigan, and Molla S. Donaldson (2000). "Errors in Health Care: A Leading Cause of Death and Injury". en. In: *To Err is Human: Building a Safer Health System*. National Academies Press (US). URL: https://www.ncbi.nlm.nih.gov/books/NBK225187/ (visited on 10/07/2024).

Koriat, Asher, Sarah Lichtenstein, and Baruch Fischhoff (1980). "Reasons for confidence". In: *Journal of Experimental Psychology: Human Learning and Memory* 6.2. Number: 2 Place: US Publisher: American Psychological Association, pp. 107–118. DOI: 10.1037/0278-7393.6.2.107.

Kostopoulou, Olga, Brendan C Delaney, and Craig W Munro (Dec. 2008). "Diagnostic difficulty and error in primary care—a systematic review". In: *Family Practice* 25.6. Number: 6, pp. 400–413. DOI: 10.1093/fampra/cmn071. URL: https://doi.org/10.1093/fampra/cmn071 (visited on 10/07/2024).

References

Kostopoulou, Olga, J. Edward Russo, et al. (Nov. 2012). "Information Distortion in Physicians' Diagnostic Judgments". en. In: *Medical Decision Making* 32.6. Number: 6 Publisher: SAGE Publications Inc STM, pp. 831–839. DOI: 10.1177/0272989X12447241. URL: https://doi.org/10.1177/0272989X12447241 (visited on 10/07/2024).

Kourtidis, Ploutarchos et al. (Dec. 2022). "Influences of early diagnostic suggestions on clinical reasoning". en. In: *Cognitive Research: Principles and Implications* 7.1. Number: 1, p. 103. DOI: 10.1186/s41235-022-00453-y. URL: https://doi.org/10.1186/s41235-022-00453-y (visited on 10/07/2024).

Kovacs, Roxanne J., Mylene Lagarde, and John Cairns (Jan. 2020). "Overconfident health workers provide lower quality healthcare". In: *Journal of Economic Psychology* 76, p. 102213. DOI: 10.1016/j.joep.2019.102213. URL: https://www.sciencedirect.com/science/article/pii/S0167487019300455 (visited on 10/07/2024).

Kruger, Justin and David Dunning (1999). "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments". In: *Journal of Personality and Social Psychology* 77.6. Number: 6 Place: US Publisher: American Psychological Association, pp. 1121–1134. DOI: 10.1037/0022-3514.77.6.1121.

Krupat, Edward et al. (2017). "Avoiding premature closure and reaching diagnostic accuracy: some key predictive factors". en. In: *Medical Education* 51.11. Number: 11 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/medu.13382, pp. 1127–1137. DOI: 10.1111/medu.13382. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/medu.13382 (visited on 10/07/2024).

Kuhn, Josepha, Silvia Mamede, et al. (Mar. 2023). "Learning deliberate reflection in medical diagnosis: does learning-by-teaching help?" en. In: *Advances in Health Sciences Education* 28.1. Number: 1, pp. 13–26. DOI: 10.1007/s10459-022-10138-2. URL: https://doi.org/10.1007/s10459-022-10138-2 (visited on 10/07/2024).

Kuhn, Josepha, Pieter van den Berg, et al. (Mar. 2022). "Improving medical residents' self-assessment of their diagnostic accuracy: does feedback help?" en. In: *Advances in Health Sciences Education* 27.1. Number: 1, pp. 189–200. DOI: 10.1007/s10459-021-10080-9. URL: https://doi.org/10.1007/s10459-021-10080-9 (visited on 10/07/2024).

Kuipers, Benjamin and Jerome P. Kassirer (Oct. 1984). "Causal reasoning in medicine: Analysis of a protocol". In: *Cognitive Science* 8.4. Number: 4, pp. 363–385. DOI: 10.1016/S0364-0213(84)80007-5. URL: https://www.sciencedirect.com/science/article/pii/S0364021384800075 (visited on 10/07/2024).

Küper, Alisa et al. (July 2024). "Mitigating cognitive bias with clinical decision support systems: an experimental study". In: *Journal of Decision Systems* 33.3. Number: 3 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/12460125.2023.2245215, pp. 439–458. DOI: 10.1080/12460125.2023.2245215. URL: https://doi.org/10.1080/12460125.2023.2245215 (visited on 11/29/2024).

Lambe, Kathryn Ann, David Hevey, and Brendan D. Kelly (Nov. 2018). "Guided Reflection Interventions Show No Effect on Diagnostic Accuracy in Medical

References

Students". English. In: *Frontiers in Psychology* 9. Publisher: Frontiers. DOI: 10.3389/fpsyg.2018.02297. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.02297/full (visited on 10/07/2024).

Lawton, Rebecca et al. (May 2019). "Are more experienced clinicians better able to tolerate uncertainty and manage risks? A vignette study of doctors in three NHS emergency departments in England". en. In: *BMJ Quality & Safety* 28.5. Number: 5 Publisher: BMJ Publishing Group Ltd Section: Original research, pp. 382–388. DOI: 10.1136/bmjqs-2018-008390. URL: https://qualitysafety.bmj.com/content/28/5/382 (visited on 10/07/2024).

Leblanc, Vicki R., Lee R. Brooks, and Geoffrey R. Norman (Oct. 2002). "Believing Is Seeing: The Influence of a Diagnostic Hypothesis on the Interpretation of Clinical Features". en-US. In: *Academic Medicine* 77.10. Number: 10, S67. URL: https://journals.lww.com/academicmedicine/fulltext/2002/10001/believing_is_seeing__the_influence_of_a_diagnostic.22.aspx (visited on 10/07/2024).

Lee, Ciara et al. (2021). "Towards a new understanding of uncertainty in medical education". en. In: *Journal of Evaluation in Clinical Practice* 27.5. Number: 5 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jep.13503, pp. 1194–1204. DOI: 10.1111/jep.13503. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jep.13503 (visited on 11/19/2024).

Levin, Phillip D. et al. (2012). "Antimicrobial use in the ICU: Indications and accuracy—an observational trial". en. In: *Journal of Hospital Medicine* 7.9. Number: 9 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jhm.1964, pp. 672–678. DOI: 10.1002/jhm.1964. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jhm.1964 (visited on 10/07/2024).

Li, Shan, Juan Zheng, and Susanne P. Lajoie (Apr. 2023). "The relationship between cognitive engagement and students' performance in a simulation-based training environment: an information-processing perspective". In: *Interactive Learning Environments* 31.3. Number: 3 Publisher: Routledge _eprint: https://doi.org/10.1080/10494820.2020.1848879, pp. 1532–1545. DOI: 10.1080/10494820.2020.1848879. URL: https://doi.org/10.1080/10494820.2020.1848879 (visited on 10/07/2024).

Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence D. Phillips (1977). "Calibration of Probabilities: The State of the Art". en. In: *Decision Making and Change in Human Affairs: Proceedings of the Fifth Research Conference on Subjective Probability, Utility, and Decision Making, Darmstadt, 1–4 September, 1975*. Ed. by Helmut Jungermann and Gerard De Zeeuw. Dordrecht: Springer Netherlands, pp. 275–324. DOI: 10.1007/978-94-010-1276-8_19. URL: https://doi.org/10.1007/978-94-010-1276-8_19 (visited on 11/12/2024).

Mackenzie, R. et al. (Apr. 1996). "Magnetic resonance imaging of the knee: Assessment of effectiveness". In: *Clinical Radiology* 51.4. Number: 4, pp. 245–250. DOI: 10.1016/S0009-9260(96)80340-0. URL: https://www.sciencedirect.com/science/article/pii/S0009926096803400 (visited on 10/07/2024).

Mahmoodi, Ali et al. (Mar. 2015). "Equality bias impairs collective decision-making across cultures". In: *Proceedings of the National Academy of Sciences* 112.12.

References

Number: 12 Publisher: Proceedings of the National Academy of Sciences, pp. 3835–3840. DOI: `10.1073/pnas.1421692112`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.1421692112` (visited on 10/07/2024).

Mamede, Sílvia et al. (Sept. 2024). "Role of knowledge and reasoning processes as predictors of resident physicians' susceptibility to anchoring bias in diagnostic reasoning: a randomised controlled experiment". en. In: *BMJ Quality & Safety* 33.9. Number: 9 Publisher: BMJ Publishing Group Ltd Section: Original research, pp. 563–572. DOI: `10.1136/bmjqs-2023-016621`. URL: `https://qualitysafety.bmj.com/content/33/9/563` (visited on 10/07/2024).

Mann, Doug (Apr. 1993). *The Relationship between Diagnostic Accuracy and Confidence in Medical Students*. en. Tech. rep. ERIC Number: ED358110. URL: `https://eric.ed.gov/?id=ED358110` (visited on 10/07/2024).

Mata, André, Mário B. Ferreira, and Steven J. Sherman (2013). "The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence". In: *Journal of Personality and Social Psychology* 105.3. Number: 3 Place: US Publisher: American Psychological Association, pp. 353–373. DOI: `10.1037/a0033640`.

Merkle, Edgar C. (Feb. 2009). "The disutility of the hard-easy effect in choice confidence". en. In: *Psychonomic Bulletin & Review* 16.1. Number: 1, pp. 204–213. DOI: `10.3758/PBR.16.1.204`. URL: `https://doi.org/10.3758/PBR.16.1.204` (visited on 11/12/2024).

Meyer, Ashley N. D. et al. (Nov. 2013). "Physicians' Diagnostic Accuracy, Confidence, and Resource Requests: A Vignette Study". In: *JAMA Internal Medicine* 173.21. Number: 21, pp. 1952–1958. DOI: `10.1001/jamainternmed.2013.10081`. URL: `https://doi.org/10.1001/jamainternmed.2013.10081` (visited on 10/07/2024).

Mullen, Brian and Catherine A. Riordan (1988). "Self-Serving Attributions for Performance in Naturalistic Settings: A Meta-Analytic Review". en. In: *Journal of Applied Social Psychology* 18.1. Number: 1 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.1988.tb00001.x, pp. 3–22. DOI: `10.1111/j.1559-1816.1988.tb00001.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1559-1816.1988.tb00001.x` (visited on 11/29/2024).

Navajas, Joaquin et al. (Nov. 2017). "The idiosyncratic nature of confidence". en. In: *Nature Human Behaviour* 1.11. Number: 11 Publisher: Nature Publishing Group, pp. 810–818. DOI: `10.1038/s41562-017-0215-1`. URL: `https://www.nature.com/articles/s41562-017-0215-1` (visited on 10/07/2024).

Nendaz, Mathieu and Arnaud Perrier (2012). "Diagnostic errors and flaws in clinical reasoning: mechanisms and prevention in practice". eng. In: *Swiss Medical Weekly* 142, w13706. DOI: `10.4414/smw.2012.13706`.

Neugebauer, M., M. Ebert, and R. Vogelmann (Mar. 2020). "A clinical decision support system improves antibiotic therapy for upper urinary tract infection in a randomized single-blinded study". en. In: *BMC Health Services Research* 20.1. Number: 1, p. 185. DOI: `10.1186/s12913-020-5045-6`. URL: `https://doi.org/10.1186/s12913-020-5045-6` (visited on 10/07/2024).

Nickerson, Raymond S. (June 1998). "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises". en. In: *Review of General Psychology* 2.2. Number: 2 Publisher: SAGE Publications Inc, pp. 175–220. DOI: `10.1037/1089-2680.2.2.175`. URL: `https://doi.org/10.1037/1089-2680.2.2.175` (visited on 10/07/2024).

References

Nickerson, Raymond S. (1999). "How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others". In: *Psychological Bulletin* 125.6. Number: 6 Place: US Publisher: American Psychological Association, pp. 737–759. DOI: `10.1037/0033-2909.125.6.737`.

Norman, Geoffrey et al. (Feb. 2014). "The Etiology of Diagnostic Errors: A Controlled Trial of System 1 Versus System 2 Reasoning". en-US. In: *Academic Medicine* 89.2. Number: 2, p. 277. DOI: `10.1097/ACM.0000000000000105`. URL: `https://journals.lww.com/academicmedicine/abstract/2014/02000/the_etiology_of_diagnostic_errors__a_controlled.26.aspx` (visited on 10/07/2024).

Oeberst, Aileen and Roland Imhoff (Nov. 2023). "Toward Parsimony in Bias Research: A Proposed Common Framework of Belief-Consistent Information Processing for a Set of Biases". en. In: *Perspectives on Psychological Science* 18.6. Number: 6 Publisher: SAGE Publications Inc, pp. 1464–1487. DOI: `10.1177/17456916221148147`. URL: `https://doi.org/10.1177/17456916221148147` (visited on 11/05/2024).

Oskamp, Stuart (1965). "Overconfidence in case-study judgments". In: *Journal of Consulting Psychology* 29.3. Number: 3 Place: US Publisher: American Psychological Association, pp. 261–265. DOI: `10.1037/h0022125`.

Pescetelli, Niccolò, Anna-Katharina Hauperich, and Nick Yeung (Oct. 2021). "Confidence, advice seeking and changes of mind in decision making". In: *Cognition* 215, p. 104810. DOI: `10.1016/j.cognition.2021.104810`. URL: `https://www.sciencedirect.com/science/article/pii/S0010027721002298` (visited on 10/07/2024).

Peters, Megan A. K., Thomas Thesen, et al. (July 2017). "Perceptual confidence neglects decision-incongruent evidence in the brain". en. In: *Nature Human Behaviour* 1.7. Number: 7 Publisher: Nature Publishing Group, pp. 1–8. DOI: `10.1038/s41562-017-0139`. URL: `https://www.nature.com/articles/s41562-017-0139` (visited on 11/05/2024).

Peters, Micah Dj, Christina Godfrey, et al. (2024). "Scoping reviews". In: *JBI Manual for Evidence Synthesis*. Ed. by Edoardo Aromataris et al. JBI. DOI: `10.46658/JBIMES-24-09`. URL: `https://jbi-global-wiki.refined.site/space/MANUAL/355862497/10.+Scoping+reviews` (visited on 12/02/2024).

Peterson, Michael C. et al. (Oct. 1992). "Contributions of the History, Physical Examination, and Laboratory Investigation in Making Medical Diagnoses:" en. In: *Obstetrical & Gynecological Survey* 47.10. Number: 10, pp. 711–712. DOI: `10.1097/00006254-199210000-00013`. URL: `http://journals.lww.com/00006254-199210000-00013` (visited on 11/18/2024).

Pouget, Alexandre, Jan Drugowitsch, and Adam Kepecs (Mar. 2016). "Confidence and certainty: distinct probabilistic quantities for different goals". en. In: *Nature Neuroscience* 19.3. Number: 3 Publisher: Nature Publishing Group, pp. 366–374. DOI: `10.1038/nn.4240`. URL: `https://www.nature.com/articles/nn.4240` (visited on 10/07/2024).

Price, Paul C. and Eric R. Stone (2004). "Intuitive evaluation of likelihood judgment producers: evidence for a confidence heuristic". fr. In: *Journal of Behavioral Decision Making* 17.1. Number: 1 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bdm.460, pp. 39–57. DOI:

References

10.1002/bdm.460. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.460 (visited on 10/07/2024).

Price, Tom, Matthew Tenan, et al. (Mar. 2016). "Acute stress causes over confidence in Situation Awareness". In: *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. ISSN: 2379-1675, pp. 1–6. DOI: 10.1109/COGSIMA.2016.7497778. URL: https://ieeexplore.ieee.org/abstract/document/7497778 (visited on 10/07/2024).

Putnam, Adam L., Victor W. Sungkhasettee, and Henry L. Roediger (Sept. 2016). "Optimizing Learning in College: Tips From Cognitive Psychology". en. In: *Perspectives on Psychological Science* 11.5. Number: 5 Publisher: SAGE Publications Inc, pp. 652–660. DOI: 10.1177/1745691616645770. URL: https://doi.org/10.1177/1745691616645770 (visited on 10/07/2024).

Rabbitt, Patrick M. A. (May 1968). "Three Kinds of Error-Signalling Responses in a Serial Choice Task". en. In: *Quarterly Journal of Experimental Psychology* 20.2. Number: 2 Publisher: SAGE Publications, pp. 179–188. DOI: 10.1080/14640746808400146. URL: https://doi.org/10.1080/14640746808400146 (visited on 12/18/2024).

Rausch, Manuel et al. (Sept. 2020). "Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring". In: *NeuroImage* 218, p. 116963. DOI: 10.1016/j.neuroimage.2020.116963. URL: https://www.sciencedirect.com/science/article/pii/S1053811920304493 (visited on 10/07/2024).

Redelmeier, Donald A. and Eldar Shafir (Feb. 2023). "The Fallacy of a Single Diagnosis". In: *Medical Decision Making* 43.2. Number: 2 Publisher: SAGE Publications Inc STM, pp. 183–190. DOI: 10.1177/0272989X221121343. URL: https://journals.sagepub.com/doi/full/10.1177/0272989X221121343 (visited on 10/07/2024).

Restrepo, Daniel, Katrina A. Armstrong, and Joshua P. Metlay (June 2020). "Annals Clinical Decision Making: Avoiding Cognitive Errors in Clinical Decision Making". In: *Annals of Internal Medicine* 172.11. Number: 11 Publisher: American College of Physicians, pp. 747–751. DOI: 10.7326/M19-3692. URL: https://www.acpjournals.org/doi/abs/10.7326/M19-3692 (visited on 10/07/2024).

Robbins, Jordan M. and Joachim I. Krueger (Feb. 2005). "Social Projection to Ingroups and Outgroups: A Review and Meta-Analysis". en. In: *Personality and Social Psychology Review* 9.1. Number: 1 Publisher: SAGE Publications Inc, pp. 32–47. DOI: 10.1207/s15327957pspr0901_3. URL: https://doi.org/10.1207/s15327957pspr0901_3 (visited on 11/29/2024).

Robinson, Lori B. and Reid Hastie (1985). "Revision of beliefs when a hypothesis is eliminated from consideration". In: *Journal of Experimental Psychology: Human Perception and Performance* 11.4. Number: 4 Place: US Publisher: American Psychological Association, pp. 443–456. DOI: 10.1037/0096-1523.11.4.443.

Roediger Iii, Henry L., John H. Wixted, and K. Andrew Desoto (July 2012). "The Curious Complexity between Confidence and Accuracy in Reports from Memory". en. In: *Memory and Law*. Ed. by Lynn Nadel and Walter P. Sinnott-Armstrong. Oxford University Press, pp. 84–117. DOI:

References

    `10.1093/acprof:oso/9780199920754.003.0004`. URL:
    `https://academic.oup.com/book/25835/chapter/193484433` (visited on
    10/07/2024).

Rollwage, Max et al. (May 2020). "Confidence drives a neural confirmation bias". en. In:
    Nature Communications 11.1. Number: 1 Publisher: Nature Publishing Group,
    p. 2634. DOI: `10.1038/s41467-020-16278-6`. URL:
    `https://www.nature.com/articles/s41467-020-16278-6` (visited on
    10/07/2024).

Rouault, Marion, Peter Dayan, and Stephen M. Fleming (Mar. 2019). "Forming global
    estimates of self-performance from local confidence". en. In: *Nature Communications*
    10.1. Number: 1 Publisher: Nature Publishing Group, p. 1141. DOI:
    `10.1038/s41467-019-09075-3`. URL:
    `https://www.nature.com/articles/s41467-019-09075-3` (visited on
    11/20/2024).

Rouault, Marion, Tricia Seow, et al. (Sept. 2018). "Psychiatric Symptom Dimensions
    Are Associated With Dissociable Shifts in Metacognition but Not Task
    Performance". In: *Biological Psychiatry.* Translating Biology to Treatment in
    Schizophrenia 84.6. Number: 6, pp. 443–451. DOI:
    `10.1016/j.biopsych.2017.12.017`. URL:
    `https://www.sciencedirect.com/science/article/pii/S0006322318300295`
    (visited on 10/07/2024).

Royce, Celeste S., Margaret M. Hayes, and Richard M. Schwartzstein (Feb. 2019).
    "Teaching Critical Thinking: A Case for Instruction in Cognitive Biases to Reduce
    Diagnostic Errors and Improve Patient Safety". en-US. In: *Academic Medicine* 94.2.
    Number: 2, p. 187. DOI: `10.1097/ACM.0000000000002518`. URL:
    `https://journals.lww.com/academicmedicine/fulltext/2019/02000/`
    `teaching_critical_thinking__a_case_for_instruction.20.aspx` (visited on
    11/19/2024).

Sætrevik, Bjørn et al. (Jan. 2020). *Anchoring, confirmation and confidence bias among
    medical decision-makers.* en-us. DOI: `10.31234/osf.io/f9382`. URL:
    `https://osf.io/f9382` (visited on 10/07/2024).

Salem-Schatz, Susanne R., Jerry Avorn, and Stephen B. Soumerai (July 1990).
    "Influence of Clinical Knowledge, Organizational Context, and Practice Style on
    Transfusion Decision Making: Implications for Practice Change Strategies". In:
    JAMA 264.4. Number: 4, pp. 476–483. DOI: `10.1001/jama.1990.03450040072034`.
    URL: `https://doi.org/10.1001/jama.1990.03450040072034` (visited on
    10/07/2024).

Sanger, Patrick C. et al. (Jan. 2017). "Diagnosing Surgical Site Infection Using Wound
    Photography: A Scenario-Based Study". In: *Journal of the American College of
    Surgeons* 224.1. Number: 1, 8–15.e1. DOI: `10.1016/j.jamcollsurg.2016.10.027`.
    URL:
    `https://www.sciencedirect.com/science/article/pii/S1072751516315460`
    (visited on 10/07/2024).

Saposnik, Gustavo et al. (Nov. 2016). "Cognitive biases associated with medical
    decisions: a systematic review". en. In: *BMC Medical Informatics and Decision
    Making* 16.1. Number: 1, p. 138. DOI: `10.1186/s12911-016-0377-1`. URL:
    `https://doi.org/10.1186/s12911-016-0377-1` (visited on 10/07/2024).

References

Schaefer, Peter S. et al. (Oct. 2004). "Overconfidence and the Big Five". In: *Journal of Research in Personality* 38.5. Number: 5, pp. 473–480. DOI: 10.1016/j.jrp.2003.09.010. URL: https://www.sciencedirect.com/science/article/pii/S0092656603001089 (visited on 10/07/2024).

Scherer, Laura D. et al. (2015). "Trust in deliberation: The consequences of deliberative decision strategies for medical decisions". In: *Health Psychology* 34.11. Number: 11 Place: US Publisher: American Psychological Association, pp. 1090–1099. DOI: 10.1037/hea0000203.

Schiff, Gordon D. et al. (Nov. 2009). "Diagnostic Error in Medicine: Analysis of 583 Physician-Reported Errors". In: *Archives of Internal Medicine* 169.20. Number: 20, pp. 1881–1887. DOI: 10.1001/archinternmed.2009.333. URL: https://doi.org/10.1001/archinternmed.2009.333 (visited on 10/07/2024).

Schmidt, Eric et al. (Mar. 2013). "Simulation Exercises as a Patient Safety Strategy". In: *Annals of Internal Medicine* 158.5_Part_2. Number: 5_Part_2 Publisher: American College of Physicians, pp. 426–432. DOI: 10.7326/0003-4819-158-5-201303051-00010. URL: https://www.acpjournals.org/doi/full/10.7326/0003-4819-158-5-201303051-00010 (visited on 10/07/2024).

Schoenherr, Jordan Richard, Jason Waechter, and Scott J. Millington (Oct. 2018). "Subjective awareness of ultrasound expertise development: individual experience as a determinant of overconfidence". en. In: *Advances in Health Sciences Education* 23.4. Number: 4, pp. 749–765. DOI: 10.1007/s10459-018-9826-1. URL: https://doi.org/10.1007/s10459-018-9826-1 (visited on 10/07/2024).

Schulz, Lion et al. (Dec. 2020). "Dogmatism manifests in lowered information search under uncertainty". In: *Proceedings of the National Academy of Sciences* 117.49. Publisher: Proceedings of the National Academy of Sciences, pp. 31527–31534. DOI: 10.1073/pnas.2009641117. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2009641117 (visited on 01/10/2025).

Schulz-Hardt, Stefan et al. (2000). "Biased information search in group decision making". In: *Journal of Personality and Social Psychology* 78.4. Number: 4 Place: US Publisher: American Psychological Association, pp. 655–669. DOI: 10.1037/0022-3514.78.4.655.

See, Kelly E. et al. (Nov. 2011). "The detrimental effects of power on confidence, advice taking, and accuracy". In: *Organizational Behavior and Human Decision Processes* 116.2. Number: 2, pp. 272–285. DOI: 10.1016/j.obhdp.2011.07.006. URL: https://www.sciencedirect.com/science/article/pii/S0749597811000975 (visited on 10/07/2024).

Shea, Nicholas et al. (Apr. 2014). "Supra-personal cognitive control and metacognition". English. In: *Trends in Cognitive Sciences* 18.4. Number: 4 Publisher: Elsevier, pp. 186–193. DOI: 10.1016/j.tics.2014.01.006. URL: https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(14)00023-0 (visited on 11/05/2024).

Silver, Ike, Barbara A. Mellers, and Philip E. Tetlock (Sept. 2021). "Wise teamwork: Collective confidence calibration predicts the effectiveness of group discussion". In: *Journal of Experimental Social Psychology* 96, p. 104157. DOI: 10.1016/j.jesp.2021.104157. URL:

References

https://www.sciencedirect.com/science/article/pii/S0022103121000603
(visited on 10/07/2024).

Soares, William E. et al. (Jan. 2019). "Accuracy Screening for ST Elevation Myocardial Infarction in a Task-switching Simulation". In: *Western Journal of Emergency Medicine* 20.1. Number: 1, pp. 177–184. DOI: 10.5811/westjem.2018.10.39962. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6324702/ (visited on 10/07/2024).

Staal, J., K. Katarya, et al. (Mar. 2024). "Impact of performance and information feedback on medical interns' confidence–accuracy calibration". en. In: *Advances in Health Sciences Education* 29.1. Number: 1, pp. 129–145. DOI: 10.1007/s10459-023-10252-9. URL: https://doi.org/10.1007/s10459-023-10252-9 (visited on 10/07/2024).

Staal, J., M. Speelman, et al. (Apr. 2022). "Does a suggested diagnosis in a general practitioners' referral question impact diagnostic reasoning: an experimental study". en. In: *BMC Medical Education* 22.1. Number: 1, p. 256. DOI: 10.1186/s12909-022-03325-7. URL: https://doi.org/10.1186/s12909-022-03325-7 (visited on 10/07/2024).

Stasser, Garold and William Titus (1985). "Pooling of unshared information in group decision making: Biased information sampling during discussion". In: *Journal of Personality and Social Psychology* 48.6. Number: 6 Place: US Publisher: American Psychological Association, pp. 1467–1478. DOI: 10.1037/0022-3514.48.6.1467.

Stephens, Georgina C., Charlotte E. Rees, and Michelle D. Lazarus (Mar. 2021). "Exploring the impact of education on preclinical medical students' tolerance of uncertainty: a qualitative longitudinal study". en. In: *Advances in Health Sciences Education* 26.1. Number: 1, pp. 53–77. DOI: 10.1007/s10459-020-09971-0. URL: https://doi.org/10.1007/s10459-020-09971-0 (visited on 11/19/2024).

Syzmanowicz, Agata and Adrian Furnham (Oct. 2011). "Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses". In: *Learning and Individual Differences* 21.5. Number: 5, pp. 493–504. DOI: 10.1016/j.lindif.2011.07.001. URL: https://www.sciencedirect.com/science/article/pii/S1041608011000781 (visited on 10/07/2024).

Tabak, Nili, Yoram Bar-Tal, and Jiska Cohen-Mansfield (Oct. 1996). "Clinical Decision Making of Experienced and Novice Nurses". en. In: *Western Journal of Nursing Research.* Publisher: SAGE PUBLICATIONS, INC.2455 Teller Road, Thousand Oaks, CA 91320. DOI: 10.1177/019394599601800505. URL: https://journals.sagepub.com/doi/abs/10.1177/019394599601800505 (visited on 10/07/2024).

Tarrant, Mark, Raff Calitri, and Dale Weston (Sept. 2012). "Social Identification Structures the Effects of Perspective Taking". en. In: *Psychological Science* 23.9. Number: 9 Publisher: SAGE Publications Inc, pp. 973–978. DOI: 10.1177/0956797612441221. URL: https://doi.org/10.1177/0956797612441221 (visited on 11/05/2024).

Thammasitboon, Satid and William B. Cutrer (Oct. 2013). "Diagnostic Decision-Making and Strategies to Improve Diagnosis". In: *Current Problems in Pediatric and Adolescent Health Care* 43.9. Number: 9, pp. 232–241. DOI: 10.1016/j.cppeds.2013.07.003. URL:

References

https://www.sciencedirect.com/science/article/pii/S1538544213000928
(visited on 11/20/2024).

The state of medical education and practice in the UK. Workplace experiences 2024
(2024). en. Tech. rep. GMC.

Thompson, Carl et al. (July 2004). "Nurses, information use, and clinical decision
making—the real world potential for evidence-based decisions in nursing". en. In:
Evidence-Based Nursing 7.3. Number: 3 Publisher: Royal College of Nursing Section:
EBN notebook, pp. 68–72. DOI: 10.1136/ebn.7.3.68. URL:
https://ebn.bmj.com/content/7/3/68 (visited on 10/07/2024).

Thorlacius-Ussing, Gorm et al. (Jan. 2021). "Comparing a Single Clinician Versus a
Multidisciplinary Consensus Conference Approach for Dementia Diagnostics". en.
In: *Journal of Alzheimer's Disease* 83.2. Number: 2 Publisher: IOS Press,
pp. 741–751. DOI: 10.3233/JAD-210278. URL:
https://content.iospress.com/articles/journal-of-alzheimers-
disease/jad210278 (visited on 10/07/2024).

Tio, René A. et al. (Aug. 2022). "The Effect of Information Presentation Order on
Residents' Diagnostic Accuracy of Online Simulated Patients With Chest Pain". In:
Journal of Graduate Medical Education 14.4. Number: 4, pp. 475–481. DOI:
10.4300/JGME-D-21-01053.1. URL:
https://doi.org/10.4300/JGME-D-21-01053.1 (visited on 10/07/2024).

Tranfield, David, David Denyer, and Palminder Smart (2003). "Towards a Methodology
for Developing Evidence-Informed Management Knowledge by Means of Systematic
Review". en. In: *British Journal of Management* 14.3. Number: 3 _eprint:
https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8551.00375, pp. 207–222. DOI:
10.1111/1467-8551.00375. URL:
https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8551.00375 (visited
on 10/07/2024).

Vallee-Tourangeau, Frederic, David M. Beynon, and Staunton A. James (Mar. 2000).
"The role of alternative hypotheses in the integration of evidence that disconfirms
an acquired belief". In: *European Journal of Cognitive Psychology* 12.1. Number: 1
Publisher: Routledge _eprint: https://doi.org/10.1080/095414400382226,
pp. 107–129. DOI: 10.1080/095414400382226. URL:
https://doi.org/10.1080/095414400382226 (visited on 11/19/2024).

Van den Berge, Kees and Sílvia Mamede (Sept. 2013). "Cognitive diagnostic error in
internal medicine". In: *European Journal of Internal Medicine* 24.6. Number: 6,
pp. 525–529. DOI: 10.1016/j.ejim.2013.03.006. URL:
https://www.sciencedirect.com/science/article/pii/S0953620513000903
(visited on 10/07/2024).

Van Wallendael, Lori Robinson and Reid Hastie (May 1990). "Tracing the footsteps of
Sherlock Holmes: Cognitive representations of hypothesis testing". en. In: *Memory
& Cognition* 18.3. Number: 3, pp. 240–250. DOI: 10.3758/BF03213878. URL:
https://doi.org/10.3758/BF03213878 (visited on 10/07/2024).

Vickers, Douglas and Jeanette Packer (Mar. 1982). "Effects of alternating set for speed
or accuracy on response time, accuracy and confidence in a unidimensional
discrimination task". In: *Acta Psychologica* 50.2. Number: 2, pp. 179–197. DOI:
10.1016/0001-6918(82)90006-3. URL:
https://www.sciencedirect.com/science/article/pii/0001691882900063
(visited on 10/07/2024).

References

Wason, P. C. (July 1960). "On the Failure to Eliminate Hypotheses in a Conceptual Task". en. In: *Quarterly Journal of Experimental Psychology* 12.3. Number: 3 Publisher: SAGE Publications, pp. 129–140. DOI: 10.1080/17470216008416717. URL: https://doi.org/10.1080/17470216008416717 (visited on 10/07/2024).

Wears, Robert L. (Dec. 2014). "Diagnosing Diagnosis". English. In: *Annals of Emergency Medicine* 64.6. Number: 6 Publisher: Elsevier, pp. 586–587. DOI: 10.1016/j.annemergmed.2014.08.009. URL: https://www.annemergmed.com/article/S0196-0644(14)01151-2/abstract (visited on 10/07/2024).

— (Feb. 2015). "Standardisation and its discontents". en. In: *Cognition, Technology & Work* 17.1. Number: 1, pp. 89–94. DOI: 10.1007/s10111-014-0299-6. URL: https://doi.org/10.1007/s10111-014-0299-6 (visited on 10/07/2024).

Webster, Jane and Richard T. Watson (2002). "Analyzing the Past to Prepare for the Future: Writing a Literature Review". In: *MIS Quarterly* 26.2. Number: 2 Publisher: Management Information Systems Research Center, University of Minnesota, pp. xiii–xxiii. URL: https://www.jstor.org/stable/4132319 (visited on 10/07/2024).

Wei, L. (2021). "Confidence and information seeking in decision making under uncertainty". English. http://purl.org/dc/dcmitype/Text. University of Oxford. URL: https://ora.ox.ac.uk/objects/uuid:39d7574e-5e7c-4346-890a-6ebb6cc9a7a2 (visited on 10/07/2024).

Wilson, Ross McL et al. (1999). "An analysis of the causes of adverse events from the Quality in Australian Health Care Study". en. In: *Medical Journal of Australia* 170.9. Number: 9 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.5694/j.1326-5377.1999.tb127814.x, pp. 411–415. DOI: 10.5694/j.1326-5377.1999.tb127814.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.5694/j.1326-5377.1999.tb127814.x (visited on 10/07/2024).

Xu, Yunjie (Calvin), Bernard C.Y. Tan, and Li Yang (2006). "Who will you ask? An empirical study of interpersonal task information seeking". en. In: *Journal of the American Society for Information Science and Technology* 57.12. Number: 12 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20339, pp. 1666–1677. DOI: 10.1002/asi.20339. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20339 (visited on 12/10/2024).

Xue, Kai, Medha Shekhar, and Dobromir Rahnev (Sept. 2023). *Challenging the Bayesian confidence hypothesis*. en-us. DOI: 10.31234/osf.io/mf5zp. URL: https://osf.io/mf5zp (visited on 10/07/2024).

Yang, Huiqin and Carl Thompson (2010). "Nurses' risk assessment judgements: a confidence calibration study". en. In: *Journal of Advanced Nursing* 66.12. Number: 12 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2648.2010.05437.x, pp. 2751–2760. DOI: 10.1111/j.1365-2648.2010.05437.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2648.2010.05437.x (visited on 10/07/2024).

Yang, Huiqin, Carl Thompson, and Martin Bland (Oct. 2012). "The effect of clinical experience, judgment task difficulty and time pressure on nurses' confidence calibration in a high fidelity clinical simulation". en. In: *BMC Medical Informatics*

References

      and Decision Making 12.1. Number: 1, p. 113. DOI: 10.1186/1472-6947-12-113.
      URL: https://doi.org/10.1186/1472-6947-12-113 (visited on 10/07/2024).

Yates, Scott W. (Feb. 2020). "Physician Stress and Burnout". In: *The American
      Journal of Medicine* 133.2. Number: 2, pp. 160–164. DOI:
      10.1016/j.amjmed.2019.08.034. URL:
      https://www.sciencedirect.com/science/article/pii/S0002934319307570
      (visited on 10/07/2024).

Zarnoth, Paul and Janet A. Sniezek (July 1997). "The Social Influence of Confidence in
      Group Decision Making". In: *Journal of Experimental Social Psychology* 33.4.
      Number: 4, pp. 345–366. DOI: 10.1006/jesp.1997.1326. URL:
      https://www.sciencedirect.com/science/article/pii/S0022103197913263
      (visited on 10/07/2024).