

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 completed 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. I would like to thank Wolfson College for providing financial support, as well as personal support for my time here at Oxford.

It really took a village to make this thesis a reality. There are a lot of people to thank, and some I may forget to mention, which in itself speaks to how lucky I have been to have had so much support around me. I would firstly like to thank my amazing supervisors, Nick and Helen, for their 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. I definitely feel like I have become a better researcher because of you both. Thank you for your patience and for being willing to help me on this journey, as well as just being good people. I could not have asked for better, more supportive supervisors.

I also want to thank everyone who was part of 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 those within the group who directly helped with my research, including Nathan, Kapil, Archie, Anil, Kaushila and Laura. Your inputs, feedback and help are so very appreciated! And I would like to thank Mo for his hard work. I have also been lucky to be a part of the ACCLab for several years, 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 an amazing support system 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 and 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 your friendships have helped me a lot over the course of this PhD.

Last, and by no means 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 was enough for me). For now, I will say that I love you all so very much and you share in this achievement with me, alongside my thathas and paatis, who I am at this stage because of. I miss them every day.

Sriraj Aiyer
Wolfson College, Oxford
14 February 2025

Abstract

Medical decisions are often made within high-pressure situations, with their outcomes having a major impact on patients. There has been limited work aiming to understand the cognitive mechanisms of medical diagnoses, and the potential drivers of both diagnostic error and overconfidence. This thesis explores the role of confidence during diagnoses, and its contribution to both the information that clinicians seek and the hypotheses that clinicians consider during the diagnostic decision process. The research presented in this thesis adopts several methodologies that increase in naturalism with each successive study, including interactive patient vignettes, a think-aloud protocol, virtual reality and ethnography based on observations within real medical settings. Our first line of research investigates the interplay between information seeking and both confidence and accuracy during medical diagnoses. To this end, we found that confidence was associated with the amount of information sought, but accuracy was associated with an efficiency in information seeking, which meant seeking the right information for each patient. Our second line of research involved detecting different reasoning strategies that medical students use during diagnoses. We found that medical students used a range of reasoning strategies that could not simply be explained by their individually preferred decision making style, the type of patient case being diagnosed or even the students' educational background. The third key avenue of research involved investigating whether medical students use an approach of broadening or narrowing their diagnostic hypotheses as they receive information. We find evidence of an initial broadening of diagnoses, which can then become a narrowing of diagnoses as medical students seek hypothesis-driven tests and observe a patient's reaction to treatment (also increasing their diagnostic confidence). Collectively, this thesis shows how important confidence is as a marker during medical diagnoses, as it signals to other clinicians both the information known about a patient and how much a clinician's thinking has evolved with respect to the diagnoses being considered for the patient. This thesis has implications for medical education on how diagnostic reasoning is taught and whether there can be several 'good' approaches to making diagnoses. This thesis also has implications for cognitive psychology, both methodologically (for the richness of data from think-aloud and simulation-based tasks) and theoretically (for how information seeking may reflect differences in reasoning strategies and the consideration of multiple hypotheses at a time).

Contents

List of Figures	ix
List of Tables	xi
List of Abbreviations	xvi
1 Introduction	1
1.1 Diagnosis and Error	1
1.2 Cognitive Biases in Diagnoses	5
1.3 Confidence and Miscalibration	8
1.4 Information Seeking and its link to Confidence	13
1.5 Evaluation of Diagnostic Hypotheses	18
1.6 Current Work	21
1.7 Thesis Structure	24
2 Systematic Scoping Review on Confidence and Certainty in Diagnoses	27
2.1 Introduction	27
2.2 Methods	29
2.3 Results	31
2.4 Discussion	43
3 Information Seeking and Confidence During Medical Diagnoses	50
3.1 Introduction	50
3.2 Methods	57
3.3 Results	67
3.4 Discussion	87

4	Characterising Diagnostic Reasoning Strategies via a Think-Aloud Paradigm	94
4.1	Introduction	94
4.2	Methods	101
4.3	Results	110
4.4	Discussion	142
5	Diagnostic Uncertainty and Information Seeking in Virtual Reality Paediatric Scenarios	153
5.1	Introduction	153
5.2	Methods	158
5.3	Results	171
5.4	Discussion	180
6	Ethnography Based on In-Situ Observations within an Intensive Care Unit and Emergency Department	189
6.1	Reflexivity	193
6.2	The Adult Intensive Care Unit	194
6.3	The Emergency Department	196
6.4	Commonalities in Decision Making Between AICU and ED	197
6.5	Uncertainty	200
6.6	Information Seeking	205
6.7	Differential Evaluation	208
6.8	Discussion	210
7	Overall Discussion	217
7.1	An Integrative Model of Diagnosis	220
7.2	Constraints on Generalisability	228
7.3	Implications for Cognitive Psychology Research	229
7.4	Implications for Medical Practice and Education	236
7.5	Prompting Uncertainty During the Diagnostic Process	240
7.6	Conclusion	243
Appendices		
A	Chapter 2 Appendices	245
A.1	Full Table of Included Review Papers	245

Contents

B Chapter 3 Appendices	266
B.1 Vignette Information Requests	266
B.2 Vignette Marking Scheme (Online and Think-Aloud Studies)	268
B.3 Analysis of Expert Participants	269
B.4 Calibration of Confidence to Alternative Accuracy Measures	278
C Chapter 4 Appendices	280
C.1 Debrief Questionnaire from Think-Aloud Study	280
D Chapter 5 Appendices	281
D.1 Diagnostic Appropriateness Marking Scheme for VR Study	281
E R Environment and Packages	283
References	298

List of Figures

1.1	Visual representation of confidence calibration	11
1.2	Graph representing the results from Oskamp (1965)	17
2.1	Scoping Review: PRISMA Diagram	32
2.2	Scoping Review: Paper Distribution by Publication Year	34
2.3	Scoping Review: Conceptual Model	43
3.1	Online Study: Paradigm	59
3.2	Online Study: Screenshot 1	60
3.3	Online Study: Screenshot 2	61
3.4	Online Study: Accuracy and Confidence Values by Stage	70
3.5	Online Study: Average Differentials by Stage (Violin Plot)	73
3.6	Online Study: Differentials Plotted Against Confidence/Accuracy (Scatter Plots)	76
3.7	Online Study: Information Seeking Plotted Against Confidence/Accuracy (Scatter Plots)	79
3.8	Online Study: Information Seeking Patterns Predicting Accuracy (ROC Curve)	81
3.9	Online Study: Accuracy Plotted Against Information Seeking Vari- ability (Scatter Plot)	83
3.10	Online Study: Information Seeking Variability Split by Accuracy and Case (Bar Graph)	84
3.11	Online Study: Information Sought on Each Case (Heatmap)	87
4.1	Think-Aloud Study: Differential Evaluations (Bar Graph)	114
4.2	Think-Aloud Study: Reasoning Strategies by Case (Bar Graph)	116
4.3	Think-Aloud Study: Initial Differentials and Reasoning Strategy against Accuracy (Regression)	142
5.1	VR Study: Screenshot 1	159
5.2	VR Study: Screenshot 2	159

List of Figures

5.3	VR Study: Scoring Criteria Example	165
5.4	VR Study: Confidence by Timepoint and Scenario (Violin Plots) . .	173
5.5	VR Study: Differentials by Timepoint and Scenario (Violin Plots) .	175
5.6	VR Study: Effect Sizes of Initial Confidence Predicting Subsequent Information Seeking	177
5.7	VR Study: History Value Against Diagnostic Appropriateness (Scat- ter Plot)	179
7.1	Overall Discussion: Conceptual Model for Prompting Uncertainty .	241
B.1	Online Study Appendix: Average Differentials by Stage for Expert Participants (Violin Plot)	270
B.2	Online Study Appendix: Expert Participants' Confidence and Ac- curacy Across Stages	272
B.3	Online Study Appendix: Expert Information Seeking by Case (Heatmap)	274
B.4	Online Study Appendix: Expert Information Seeking against Con- fidence/Accuracy (Scatter Plots)	275
B.5	Online Study Appendix: Expert Information Seeking Variability by Accuracy (Bar Graph)	277

List of Tables

2.1	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 Appendices (Table A.1).	34
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, 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).	69
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.	85

List of Tables

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. . .	86
4.1	Table of data from this think-aloud study showing, by case (AD = Aortic Dissection, GBS = Guillain Barré Syndrome, MTB = Miliary Tuberculosis, TA = Temporal Arteritis, TTP = Thrombotic Thrombocytopenic Purpura, UC = Ulcerative Colitis), from left to right, average values across participants for Accuracy (the proportion of participants who mentioned a correct differential during the case), Final Confidence (reported at the Testing stage), Difficulty (as rated by participants at the end of the case on a scale of 1-10) and Information Seeking (the proportion of available information sought).	112
4.2	Descriptive Statistics for subcodes within the Differential Evaluation main code as detailed above in the Data Analysis section. Shown above are mean values for the number of instances/utterances for each of the following subcodes (from left to right): a new differential being considered, a differential being removed from consideration, a differential being seen as more likely given a piece of information, a differential being seen as less likely given a piece of information, the average total of these subcodes.	112
4.3	Strategy coded for each case by participant (rows) and by patient condition (column) after resolving conflicts between both independent coders. The rightmost column shows the subjectively reported strategy that participants felt they tended to use. Anonymised participant IDs are used.	115
4.4	Examples of cases that were coded for a particular reasoning strategy and key quotes that suggest the coded reasoning strategy. We provide the quotes from each case, as well as how we interpret the quotes in line with the particular reasoning strategy.	117

List of Tables

4.5	Mean values for dependent variables broken down by the reasoning strategy coded after resolving conflicts between the two independent coders. From left to right, Accuracy refers to the proportion of cases where a correct differential was mentioned. Differential Evaluations refers to the number of coded utterances under one of the subcodes (Differential Added, Differential Removed, Increased Likelihood, Decreased Likelihood). Information Seeking refers to the proportion of available information sought across cases. Information Value refers to how useful the information sought by participants was by calculating the difference in accuracy between participants who did and those who did not seek each piece of information for a given case (see Data Analysis section for the online experiment). Confidence Change refers to difference between initial confidence and final confidence.	119
4.6	Table showing average accuracy values by cases where the participants used or did not use the dominant reasoning strategy for that case. Dominant strategies are decided based on which of the reasoning strategies was utilised by the majority of participants in the think-aloud study. Cases without a coded reasoning strategy are excluded from this table. The first column refers to the dominant strategy for that condition, whilst the second column refers to whether the cases' coded strategy matches the condition's dominant strategy.	121
4.7	Dependent variables by cases where the reasoning strategy used (as categorised by the independent coders) matches the subjective strategy coded for that participant (as per responses to the debrief interview, see Table 4.8 below).	122
4.8	Categorisation of participants under one of three possible reasoning strategies based on their responses during the debrief interview. We capture here the subjective reasoning strategy for each participant based on how they reflect on how they tend to make diagnostic decisions. In the second column are key highlighted quotes related to each of the participants' diagnostic decision making processes. In the fourth column, we provide our summary of the quote and then in the fifth column, our interpretation of the quote that explains the choice of reasoning strategy for that participant. Participants have been sorted/ordered by their coded strategy.	123

List of Tables

5.1	Average values for dependent variables by scenario. The n column denotes the number of participants (or 'observations') per scenario. We show mean values for the Diagnostic Appropriateness, Performance Score, Amount of Information Seeking (the number of actions taken belonging to History, Physical Examinations or Testing), Initial Confidence (as reported at the pause point in the scenario), Final Confidence (reported at the end of the scenario), Initial Diagnoses (the number of differentials reported at the pause point) and Final Diagnoses (the number of differentials reported at the end of the scenario).	172
7.1	Integrative model of diagnostic decisions, with respect to information seeking and hypothesis generation, based on findings from our studies. The Description column details the three stages of the diagnostic decision process. The main findings from our studies that support each of these stages is provided in the Associated Study Findings column, including the section/figure of the thesis from which these findings are taken. We note that these stages are not considered as constituting a linear process. Rather, it is envisioned that clinicians reflect on their decision making by 'going back' to previous steps (e.g. revisiting initial differential generation whilst determining a diagnosis, cycling between broadening and narrowing differentials).	222
A.1	Full list of papers that were included in the systematic scoping review of papers on confidence and certainty in medical diagnoses. Papers are arranged in alphabetical order. Studies marked with ** next to their title were included via citation tracking.	265
B.1	Full list of possible information requests that participants can make. This set of information is the same for all cases. The same vignettes and corresponding information are used for the online and think-aloud vignette studies.	267
B.2	Marking scheme used to denote differentials that are considered as correct for each of the six patient cases/vignettes. The same marking scheme is applied for online and think-aloud vignette studies. The presenting complaint is shown to participants at the start of the case, before they start seeking information.	268

List of Tables

D.1	Marking criteria for the VR Study. Differentials are shown for each scenario that were marked as either probable/possible and those categorised as improbable/unlikely. Any differentials not included in this table were marked as incorrect.	282
-----	--	-----

List of Abbreviations

ABCDE	...	Airway Breathing Circulation Disability Exposure (assessment tool)
AD	...	Aortic Dissection
AICU	...	Adult Intensive Care Unit
AUC	...	Area Under the Curve
CI	...	Confidence Interval
CRP	...	C-Reactive Protein blood Test
DKA	...	Diabetic Ketoacidosis
ECG	...	Electrocardiogram
ED	...	Emergency Department
EPR	...	Electronic Patient Record
FBC	...	Full Blood Count
GBS	...	Guillain-Barré Syndrome
GI	...	Gastrointestinal
GLM	...	Generalised Logistic Regression
HD	...	Hypothetico-Deductive Reasoning
LOOCV	...	Leave One Out Cross-Validation
M	...	Mean
MDiff	...	Mean Difference
MRC	...	Medical Research Council
MRI	...	Magnetic Resonance Imaging
MTB	...	Miliary Tuberculosis
NHS	...	National Health Service
OMS	...	Oxford Medical Simulation
OSF	...	Open Science Framework

List of Abbreviations

OUH	Oxford University Hospitals (NHS Trust)
PE	Pulmonary Embolism
PR	Pattern Recognition
PRISMA	. . .	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROC	Receiver Operating Characteristic
SBAR	Situation Background Assessment Recommendation (handover framework)
SD	Standard Deviation
SI	Scheme-Inductive Reasoning
TA	Temporal Arteritis
TTP	Thrombotic Thrombocytopenic Purpura
UC	Ulcerative Colitis
U&E	Urea & Electrolytes
VBG	Venous Blood Gas
VR	Virtual Reality