

'This time is different': physician knowledge in the age of artificial intelligence

Gurpreet Dhaliwal 

Department of Medicine,
University of California San
Francisco, San Francisco,
California, USA

Correspondence to
Dr Gurpreet Dhaliwal;
gurpreet.dhaliwal@ucsf.edu

Accepted 24 April 2024
Published Online First
3 May 2024

Great diagnosticians are often portrayed as recognising rare diseases that evade the efforts of mere mortals. This makes for great TV and local legends, but does not reflect daily practice, where the most common diagnostic challenge is discriminating between common conditions like pneumonia and heart failure or appendicitis and gastroenteritis.

Questions about how to train the brain to make those distinctions are central to the efforts of many clinician educators. An unresolved issue is whether the structure of knowledge (about diseases and diagnostic pathways) in the physician's long-term memory or the clinician's mode of cognition (intuitive or analytical thinking) is more deterministic of diagnostic success. A study¹ in this issue of *BMJQS* sheds light on this issue, but also invites a broader question: is physician cognition still essential for this task at all?

A TEST OF LOOKALIKES

In a two-phase experiment, Mamede *et al*¹ asked 68 internal medicine residents to recall from memory the key clinical features of six conditions (vitamin B₁₂ deficiency, inflammatory bowel disease, hyperthyroidism, adrenal insufficiency, appendicitis, endocarditis). Physicians were categorised as high knowledge (HK) or low knowledge (LK) based on their recall of discriminating features, which are essential to differentiate one condition from common competing diagnoses.

One week later, the residents were given related clinical vignettes and asked to render a diagnosis. Half of the vignettes had a *salient distracting feature* (SDF), a clinical finding that may prompt the physician to suspect a condition other than the correct diagnosis. For example, a vignette of a confused patient included a family history of dementia, which was

irrelevant in the face of strong evidence for vitamin B₁₂ deficiency. The authors used the SDF as a model for activating the *anchoring heuristic*, which is a tendency to adhere to an early judgement triggered by a data point. Essentially, the authors created a trap and wanted to see who fell for it—and why.

The main outcome was the frequency of a specific diagnostic error—selecting the SDF-concordant diagnosis (eg, selecting dementia instead of vitamin B₁₂ deficiency). The authors found that the LK and HK residents had equal diagnostic accuracy for cases without an SDF. But when a case had an SDF, LK physicians were more likely to offer the diagnosis linked to the SDF than HK physicians. This raised the question of why HK physicians were less susceptible to the trap. Was it due to their greater knowledge or because they were more analytical? The authors used a proxy of analytical reasoning, time taken to make a diagnosis, to find out.

When faced with distracting information, both LK and HK residents thought longer about the case, but the LK physicians were still more likely to get it wrong. Extended deliberation did not explain the diagnostic differences between the two groups, but knowledge did.

ANALYSING THE ANALYTICAL MODE

Mamede *et al*'s study aligns with other investigations which find that steering the brain toward the analytical mode has limited prospects for improving diagnosis.²⁻⁵ Vignette studies lack ecological validity (eg, text cases cannot capture the context of a patient encounter),⁶ but their design allows for the interrogation of concepts that inform medical education based on sound logic but limited evidence. Such studies have helped dispel



► <http://dx.doi.org/10.1136/bmjqs-2023-016621>



© Author(s) (or their employer(s)) 2024. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Dhaliwal G.
BMJ Qual Saf
2024;**33**:549–551.

notions that there is a hierarchy between intuition and analysis,^{7,8} that heuristics and biases are reliably mitigated by analytical thought,⁹ or that heuristics trade accuracy for efficiency.¹⁰

Heuristics are human predispositions to think; they attract the name ‘bias’ when things go wrong. When we are well informed and well practised on a matter, these predispositions (eg, early perception of mottled skin in a patient with impending sepsis) are incredibly powerful and are a hallmark of expertise. When we are ill-informed or inexperienced, these predispositions (eg, latching on to a report of night sweats and immediately favouring tuberculosis) are a liability and get assigned a name like anchoring bias.¹¹ Heuristics turn into biases in the void of expertise. The study by Mamede *et al* illustrates that knowledge can immunise against bias¹² but that analytical thinking cannot reliably overcome it.

The equal time the LK and HK residents spent contemplating the SDF cases suggests that the analytical mode is heterogeneous across clinicians. When diagnosing, a physician can spend their time in analytical mode seeking contradictory evidence, rejoicing in confirmatory evidence or pondering irrelevant evidence. One thing is certain though: we all have the best chance of making use of that mode when we possess the relevant knowledge. If I am unaware that ataxia, macrocytosis and neutrophil hypersegmentation point to vitamin B₁₂ deficiency, no amount of analysis will get me to the right diagnosis. But maybe a computer will.

MACHINE LEARNING

This may be an odd moment to highlight the power of the brain in sorting between competing diagnoses. Machine learning systems have convincingly demonstrated the capacity to distinguish melanoma from a benign nevus¹³ or detect diabetic retinopathy,¹⁴ often matching or exceeding the accuracy of physicians.¹⁵ To some, these milestones herald the declining need for physicians to store information in long-term memory. For others, this is a familiar story.

Every advance in information technology, from the written word to the printing press to the internet, has sparked debate about the impact on learning and knowing. Mass-produced books in the 15th century triggered fears that the incentive to memorise and internalise concepts would be diminished. The internet brought a similar prediction—that humans (particularly students) would be liberated from memorisation because every fact would be available at their fingertips. Memorisation would be replaced with critical thinking skills. Recall would be replaced by search skills.

But it was not so. Infinite access did not produce infinite wisdom for any subject.¹⁶ It still takes a sustained intellectual investment to speak about a topic critically. Owning an English–Japanese dictionary does

not make someone fluent in Japanese; that still takes unrelenting memorising, practice and failure. And physicians rarely have time to research the questions that arise in their daily work¹⁷; instead, they continue to practise almost entirely using their knowledge and experience. But now, each of these tasks—erudite academic treatises produced with one well-crafted prompt, real-time language translation, and instantaneous answers to medical questions—have all been performed admirably by artificial intelligence systems. Such remarkable feats make me think that perhaps this technological revolution *is* truly different from its predecessors.

Yet, back in the clinic and hospital, patients arrive with unique scenarios and intersecting problems that neither the clinician nor the computer has ever seen before. For the time being, the upper hand in generating new solutions at the point of care goes to clinicians. Although the computer has the advantage in accessing and remixing all the old solutions that humans have created, it lacks understanding of context, emotion and common sense and operates without deep conceptual knowledge that physicians draw on when old solutions fail.¹⁸

MEMORY IS DEAD. LONG LIVE MEMORY

Memorisation in the sense of keeping basic facts on speed dial forever is waning in relevance. But the importance of human knowledge in solving complex health problems remains as relevant as ever.

Medicine is a multifaceted endeavour that asks more of the clinician than rendering a diagnosis at one point in time. It requires skills in teamwork, communication, health system navigation, and troubleshooting. For the focused portion of the job that requires making A versus B diagnostic comparisons, computers may soon learn enough to outperform clinicians. I suspect they would do a splendid job on the cases in Mamede *et al*'s study¹⁹ and that they will continue to amaze us by making insights that evade the human eye and brain like detecting the fingerprints of atrial fibrillation in a normal sinus rhythm ECG.²⁰ But if experience is any guide, whenever technology solves one problem, it creates another.

Genetic tests establish variants of unknown significance, head-to-toe imaging leaves incidentalomas in its wake and complex risk equations change adjectives into numbers (eg, ‘medium risk’ becomes 9%). Technology rarely resolves uncertainty—it just shifts the frontier.²¹ And when it does, it is still up to the physician to manage it. As the complexity and stakes of the uncertainty relocate—perhaps from ‘is this congestive heart failure or pneumonia?’ to ‘what are the stakes for this matriarch if we get that diagnosis wrong given her recent transplant, recurrent *Clostridium difficile* infections and family turmoil?’—the situation will call for a compassionate clinician who has deep knowledge, rich experience and an endless desire to grow both.

I love the optimism in the phrase ‘this time is different,’ and I sincerely hope it is true. But for now, the results of Mamede *et al*’s study bolster my standing advice to students and residents: do not worry about machine learning; worry more about how you can become a learning machine.

Contributors GD conceived of, drafted and revised the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests GD is a board member of the Society to Improve Diagnosis in Medicine.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iD

Gurpreet Dhaliwal <http://orcid.org/0000-0002-6134-5913>

REFERENCES

- Mamede S, Zandbergen A, de Carvalho-Filho MA, *et al*. Role of knowledge and reasoning processes as predictors of resident physicians’ susceptibility to anchoring bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Qual Saf* 2024;33:563–72.
- Sherbino J, Kulasegaram K, Howey E, *et al*. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. *CJEM* 2014;16:34–40.
- Sherbino J, Dore KL, Wood TJ, *et al*. The relationship between response time and diagnostic accuracy. *Acad Med* 2012;87:785–91.
- Tung A, Melchiorre M. n.d. Debiasing and educational interventions in medical diagnosis: A systematic review. *UTMJ*100.
- Staal J, Hooftman J, Gunput STG, *et al*. Effect on diagnostic accuracy of cognitive reasoning tools for the workplace setting: systematic review and meta-analysis. *BMJ Qual Saf* 2022;31:899–910.
- Croskerry P, Tait G. Clinical decision making: the need for meaningful research. *Acad Med* 2013;88:149–50.
- Croskerry P. A universal model of diagnostic reasoning. *Acad Med* 2009;84:1022–8.
- Custers EJFM. Medical education and cognitive continuum theory: an alternative perspective on medical problem solving and clinical reasoning. *Acad Med* 2013;88:1074–80.
- Hollingworth C, Barker L. New frontiers: re-establishing System1/system 2 truths. Research Live; 2019. Available: <https://www.research-live.com/article/opinion/new-frontiers-reestablishing-system1system-2-truths/id/5057422>
- McLaughlin K, Eva KW, Norman GR. Reexamining our bias against Heuristics. *Adv Health Sci Educ Theory Pract* 2014;19:457–64.
- Zwaan L, Monteiro S, Sherbino J, *et al*. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case Workups. *BMJ Qual Saf* 2017;26:104–10.
- Mamede S, de Carvalho-Filho MA, de Faria RMD, *et al*. Immunising’ physicians against availability bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Qual Saf* 2020;29:550–9.
- Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
- Lim JI, Regillo CD, Sadda SR, *et al*. Artificial intelligence detection of diabetic retinopathy: subgroup comparison of the Eyeart system with Ophthalmologists’ dilated examinations. *Ophthalmol Sci* 2023;3.
- Liu X, Faes L, Kale AU, *et al*. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health* 2019;1:e271–97.
- Goldberg C. In praise of Paper(S). *Am J Med* 2012;125:1047–8.
- Cook DA, Sorensen KJ, Wilkinson JM, *et al*. Barriers and decisions when answering clinical questions at the point of care: a grounded theory study. *JAMA Intern Med* 2013;173:1962–9.
- Mylopoulos M, Borschel DT, O’Brien T, *et al*. Exploring integration in action: Competencies as building blocks of expertise. *Acad Med* 2017;92:1794–9.
- Kanjee Z, Crowe B, Rodman A. Accuracy of a Generative artificial intelligence model in a complex diagnostic challenge. *JAMA* 2023;330:78–80.
- Yuan N, Duffy G, Dhruva SS, *et al*. Deep learning of electrocardiograms in sinus rhythm from US veterans to predict atrial fibrillation. *JAMA Cardiol* 2023;8:1131–9.
- Jha S. Uncertainty and the diagnostic Leviathan. *JAMA Intern Med* 2015;175:1085–6.