

REVIEW ARTICLE

Open Access



Cognitive perspectives on maintaining physicians' medical expertise: II. Acquiring, maintaining, and updating cognitive skills

Zachary A. Caddick^{1,2†}, Scott H. Fraundorf^{1,2*†} , Benjamin M. Rottman^{1,2†} and Timothy J. Nokes-Malach^{1,2}

Abstract

Over the course of training, physicians develop significant knowledge and expertise. We review dual-process theory, the dominant theory in explaining medical decision making: physicians use both heuristics from accumulated experience (System 1) and logical deduction (System 2). We then discuss how the accumulation of System 1 clinical experience can have both positive effects (e.g., quick and accurate pattern recognition) and negative ones (e.g., gaps and biases in knowledge from physicians' idiosyncratic clinical experience). These idiosyncrasies, biases, and knowledge gaps indicate a need for individuals to engage in appropriate training and study to keep these cognitive skills current lest they decline over time. Indeed, we review converging evidence that physicians further out from training tend to perform worse on tests of medical knowledge and provide poorer patient care. This may reflect a variety of factors, such as specialization of a physician's practice, but is likely to stem at least in part from cognitive factors. Acquired knowledge or skills gained may not always be readily accessible to physicians for a number of reasons, including an absence of study, cognitive changes with age, and the presence of other similar knowledge or skills that compete in what is brought to mind. Lastly, we discuss the cognitive challenges of keeping up with standards of care that continuously evolve over time.

Keywords Diagnosis, Expertise, Dual-process theory, Memory, Retrieval failure, Aging

Significance statement

Physicians' expertise and ability to keep up with changing evidence is central for positive patient health outcomes. Here, we begin our review by first evaluating evidence of how expertise is acquired in the medical domain, considering dominant theories about the important benefits and detriments experience has in evaluating new information. We introduce complimentary evidence related to

the role memory plays in learning and, since physicians often have careers that spans several decades, the impact that aging and time since initial certification may have on physicians' clinical performance. Importantly, medicine is an ever-evolving field which leads to changes to the standards of care over time. In light of this reality, we discuss how evidence from the cognitive science literature can inform how best to address the challenge of this complex information environment. Given the wide scope of this endeavor, we conclude with proposals to fill in important gaps in knowledge that may continue to propel the field of medicine and physician assessment forward to ensure the highest standards of care possible.

[†]Zachary A. Caddick, Scott H. Fraundorf and Benjamin M. Rottman have contributed equally to this work.

*Correspondence:

Scott H. Fraundorf
scottfraundorf@gmail.com

¹ Learning Research and Development Center, University of Pittsburgh, 3420 Forbes Ave., Pittsburgh, PA 15260, USA

² Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Introduction

A primary goal of continuing certification programs is to ensure that board-certified physicians maintain at least a certain minimum level of expertise. In this article, we examine what principles and findings from cognitive science imply about acquiring, maintaining, and updating medical expertise. We first discuss psychological theories of the acquisition of higher-order cognitive medical skills, diagnosis in particular. Next, we discuss how learned skills can be maintained in the face of forgetting and age-related changes in cognition. Lastly, we discuss the process of updating and acquiring new knowledge as medical practice evolves.

We focus on medical decision making and expertise from the traditional information-processing perspective of cognition that undergirds cognitive psychology. We acknowledge that medical decision making is much more complex in that it occurs situated in a dynamic environment with other physicians and healthcare professionals and in the larger context of medical systems (see the 2020 special issue of the journal *Diagnosis*, Volume 7, Issue 3, for many articles on this perspective). However, because continuing certification program assessments only test a physician's cognitive abilities independently, not their performance in a clinical environment, we focus on the individual physician's cognitive skills.

To situate the strength of the evidence and claims made, we attach evidence levels (EL) to in-text citations for empirical claims (see Table 1). Evidence levels range from 1 to 6, with 1 being the strongest evidence (meta-analyses) and 6 being the weakest (opinion papers).

This article is part of a collection of five articles in this special issue focused on how physicians maintain medical expertise across their careers. These reviews are narrative reviews, not systematic, because they cover a wide variety of topics, not a single narrow topic.

Acquiring medical expertise

Expertise is marked by the acquisition of large amounts of knowledge, which in turn affects how information is organized, represented, and processed. General aptitude

measures struggle to predict expert performance, suggesting that expertise is not just reserved for the highly intelligent (Moneta-Koehler et al., 2017, EL: 5). Rather, experiences play an important role in the development of expertise. For instance, the amount of *deliberate practice*—activities designed to improve targeted aspects of performance—that an individual has completed predicts their level of expertise (Ericsson et al., 1993, EL: 5). It is likely the *quality* of practice, rather than quantity, that is necessary to develop expertise; the mere number of deliberate practice hours on their own does not adequately explain expert performance (Macnamara et al., 2014, EL: 1). Ample and accurate feedback is also crucial (Kahneman & Klein, 2009, EL: 2). In the process of developing expertise, individuals learn to categorize information based on abstract principles, whereas novices categorize based on superficial details (Chi et al., 1981, EL: 3).

Dual-process theories in medical decision making

A common theme in the cognitive psychology literature is the existence of two distinct systems for information processing (for overviews, see Evans, 2008; Kahneman & Frederick, 2002; Sloman, 1996). (The term “systems” in this literature refers to cognitive strategies and habits, not necessarily to neural or anatomical distinctions.) Most dual-processing theories hold that System 1 is fast, unconscious, evolutionarily old, associative, and universal. In contrast, System 2 is slow, conscious, evolutionarily new, and rule based. An important difference between the systems is that System 2 is under the control and guidance of the individual whereas System 1 occurs automatically. Although it may seem intuitive that the conscious and controlled System 2 is superior to System 1, this is not always the case. System 1, at times, produces highly accurate decisions and does so quickly and from little information (Marewski & Gigerenzer, 2012).

Within medicine, the dual-process theory is widely accepted as the dominant paradigm for understanding clinical decision making generally, and especially for diagnosis (Croskerry, 2009a, EL: 2; 2009b, EL: 2; Croskerry et al., 2013a, 2013b, EL: 2; Norman & Eva, 2010, EL: 2; Pelaccia et al., 2011, EL: 2). For example, the dual-process theory provides the theoretical backbone of the Institute of Medicine's report on Improving Diagnosis in Healthcare (National Academies of Sciences, Engineering, & Medicine, 2015, Chapter 2, EL: 2).

System 2 is understood as the analytical, hypothetico-deductive reasoning or problem-solving process. For example, a physician might diagnose a patient by systematically running one test, ruling out one diagnosis, and then following it up with a different test relevant to a different potential diagnosis—a process that involves a series of carefully thought-out decisions with

Table 1 Evidence levels for in-text citations for empirical claims

Evidence level	Type of work
1	Quantitative meta-analysis
2	Narrative review
3	Multiple original experiments/randomized controlled trials (RCTs)
4	Single original experiment/RCT
5	Correlational or quasi-experimental study
6	Opinion paper

logical reasoning. This sort of logical reasoning was studied extensively in the earlier years of research on medical decision making (e.g., Elstein et al., 1978; Barrows et al., 1982; Coderre et al., 2003; Newfeld et al., 1981). However, one of the broad conclusions of this research is that changes in hypothetico-deductive problem solving do not appear to explain the transition from novices to experts (e.g., Boshuizen & Schmidt, 1992; Elstein & Schwarz, 2002; Groen & Patel, 1985; Neufeld et al., 1981; Norman, 2005). Instead, it seems that experts use a variety of different forms of knowledge and representations. This is not to say that hypothetico-deductive problem solving is not used by experts; of course, someone with insufficient medical training cannot effectively engage in this sort of reasoning, and of course experts do slow down, ask for second opinions, consult resources, and engage in other forms of careful analytical thinking. Rather, the point is that experts have additional skills and knowledge, one of which is extensive experience with individual patients.

In contrast to System 2, System 1 is the non-analytic decision making in medicine is the very fast pattern recognition process. For instance, pattern recognition allows a physician to quickly think of hyperthyroidism when seeing a patient who is skinny, tremulous, perspiring a lot, and has bulging eyes and a swollen neck. Pattern recognition involves classifying a current patient as similar to a prior patient (called an *exemplar*) or similar to an abstracted pattern of multiple prior patients with the same disease (called a *prototype*). Though most often discussed in terms of diagnosis, this pattern recognition process is also relevant to other decisions, such as deciding whether to order further diagnostic testing, choosing a treatment, or deciding whether to refer to a specialist. Pattern recognition is believed to rely on the same cognitive processes that people use every day, e.g., for categorizing animals as dogs versus cats or for identifying different species of trees (Cohen & Lefebvre, 2017).

Another aspect of non-analytic System 1 decision making in medicine is the use of *heuristics*—mental shortcuts that allow decisions to be reached quickly and efficiently. Pattern recognition through categorization can in fact be viewed as one such heuristic (e.g., Nilsson et al., 2008), though heuristics are broader than just pattern recognition (e.g., National Academies of Sciences, Engineering, & Medicine, 2015, Chapter 2; Whelehan et al., 2020). For example, the *representativeness* heuristic leads physicians to judge the probability that a patient has a given disease based on the *sensitivity* of the diagnostic information (probability of a positive test given that the patient has the disease) rather than its *positive predictive value* (probability that a patient has a disease given a positive test; Casscells et al., 1978; Eddy, 1982; Rottman, 2017). This is appropriate when the two diseases are roughly

equally prevalent, but leads to *base rate neglect* when one is more common.

Heuristics are neither inherently bad nor inherently good. They can help physicians make fast decisions, which can be critical in situations with time pressure. And, simple rule-based heuristics sometimes outperform formal statistical regression analysis (Marewski & Gigerenzer, 2012, EL: 2). On the other hand, sometimes heuristics are applied in the wrong context or are overly simple, such as in the base rate neglect example, which can lead to suboptimal decisions.

Despite the fact that the dual-process theory is widely accepted as the dominant model of clinical decision making, there are important debates, open questions, and ambiguities with this model. Some researchers question whether there is a clear distinction between the two systems and whether there are only two systems (e.g., De Neys, 2021, 2022; Evans, 2008). Other research has challenged the assumption that they are always opposed to each other and instead may work together (Cushman & Morris, 2015; see Kool et al., 2018 for a review). Though the two systems are often presented as always coming to different decisions, it is likely that they would often come to the same decision in a given situation (De Neys, 2022), which raises a question of how to distinguish which of the two systems is responsible for a given decision.

Perhaps the most pressing question is how people coordinate between the two systems. For example, does some sort of signal of low confidence in System 1 lead to the engagement of System 2, or are both systems engaged simultaneously with a discrepancy-monitoring system noticing when they are coming to different decisions (De Neys, 2022)? These are not just important questions for psychology but are directly related to medicine. Moulton et al. (2007) proposed that being an expert in medical decision making involves “slowing down when you should.” This emphasis on “when you should” raises the questions of when an individual should slow down and how this slowing down works. Moulton et al. (see also Croskerry, 2009b) argue that these questions are exactly what the field needs to address to truly understand medical expertise and how to support and train expertise. Stated another way, instead of focusing on fast and slow thinking processes individually, it is more important to understand how experts know that they need to slow down and think a bit harder in certain situations; unfortunately this question is hard to address and there is little existing research.

The role of experience in medical decision making

Here we take a slightly different approach from the dual-process model. This review focuses on the role that experience (interactions with many patients over time) plays

in expertise. Experience with prior patients is at the core of non-analytic System 1: recognizing the pattern in the current case as similar to prior cases. In contrast, relying on rules, evidence, guidelines, and knowledge of pathophysiological and pharmacology rather than one's own idiosyncratic past experience fits with System 2. We do not claim that prior experiences with individual patients cannot be part of the analytical model of decision making and slow deliberative thought. Indeed, of course it is possible that when making a decision about a current patient that a physician may carefully and analytically make comparisons to individual prior patients; in naturalistic decision making it is not possible to know whether a physician is engaged in 'slow' or 'fast' thinking at a given moment (or that only one process occurred). However, to the extent that extensive experience can produce very fast decisions, it is more likely to be involved in System 1 thinking.

The strength of utilizing experience is that it allows for fast pattern recognition, which frequently leads to accurate diagnoses. For example, one study (Norman et al., 1989, EL: 5) examined the accuracy of dermatologists reading slides. They found that not only did the dermatologists demonstrate high levels of accuracy, but they also answered significantly faster on slides they got correct, showing how diagnosis can often be both extremely fast and accurate. A number of studies with primary care and emergency medicine physicians have found similar results: clinicians think of a few potential diagnoses within seconds to minutes and are usually right (Barrows et al., 1982, EL: 5; Elstein et al., 1978, EL: 5; Gruppen et al., 1988, EL: 5; Pelaccia et al., 2014, EL: 5). This has led to the provocative question by Norman et al. (2007): "How can it be that experts with minimal information are able to advance tentative hypotheses about the diagnoses, seemingly effortlessly, and apparently without conscious awareness of the retrieval process? ... Where do the hypotheses come from?" The answer, according to this line of research, is that with enough experience clinicians can quickly pattern-match a target case from a large set of prior cases.

However, there are also downsides to relying heavily on experience. Even though pattern recognition and reliance on the diagnosis and treatment decisions made for past patients can often work out well, it can also lead to biases. For example, in one experiment, family medicine residents were given a set of cases to practice interpreting ECGs. In the initial set of cases, a brief clinical scenario accompanied each case, along with the correct diagnoses. In the latter set of test cases, participants had to identify the correct diagnosis. For some of the test cases, the accompanying clinical scenario involved irrelevant features, such as the patient's job, that matched

features from the initial cases. When an irrelevant feature matched a prior case, the residents were more likely to give the same diagnosis as the prior case, which turned out to be wrong (Hatala et al., 1999, EL: 4; for other similar studies, see Brooks et al., 1991, EL: 4; Young et al., 2011, EL: 4).

A few studies that shown a similar role of prior experience in real-world medical decision making. One study investigated how often physicians prescribed warfarin for patients with atrial fibrillation in order to prevent a stroke, which despite the risks, is a standard of practice (Choudhry et al., 2006, EL: 5). When one of the physician's patients who was on warfarin experienced a severe bleeding event that was likely a side effect of the warfarin, the physicians were about 20% less likely to prescribe warfarin for patients with atrial fibrillation for at least a year afterward. Another study found that after delivering a baby and experiencing labor and delivery complications, a physician was a bit more likely to use a vaginal delivery instead of cesarean, or vice versa, depending on the prior delivery method (Singh, 2021). In sum, physicians' decisions can be impacted by recent experiences with other patients.

A particular challenge is that these experiences are idiosyncratic. If a physician works in a specialized clinic, they may see certain types of patients even though they still need to be able to diagnose and treat a broader set of patients that they see less frequently. This means that physicians are systematically missing out on experience with certain types of patients. For example, in one study, residents' beliefs about the prevalence of a disease were correlated with their probability of providing it as a potential diagnosis (Rottman et al., 2016, EL: 5). In general, this tendency makes sense from the rational Bayesian perspective of diagnosis, in which general "prior" beliefs about the likelihood of diseases in the population are updated with knowledge of the signs and symptoms and diagnostic tests of the specific patient to form a "posterior" probability of each disease on the differential (Ledley & Lusted, 1959; Pauker & Kassirer, 1980). However, to the extent that prevalence beliefs are distorted by one's experience, some diagnoses could be overlooked. Additionally, the appearance of patients with rare diseases or rare side effects from treatments (e.g., the bleeding events discussed above) is governed by chance, so physicians may be influenced by the vicissitudes of daily practice. Thus, it is important to receive corrective feedback and not to overly rely on one's own experiences.

Another problem with relying on one's experience is that experience provides an imperfect feedback system. Feedback is vital for developing expertise (e.g., Ericsson, 2015, EL: 2; Hattie & Timperley, 2007, EL: 2; Kahneman & Klein, 2009, EL: 2), and lack of feedback is believed to

contribute to overconfidence (Kahneman & Klein, 2009, EL: 2). However, the medical system is poor at providing feedback (National Academies of Sciences, Engineering, & Medicine, 2015, EL: 6; Schiff, 2008, EL: 6). An error in diagnosis or treatment may never be discovered, in which case feedback is never received. Additionally, because of the complex nature of modern medicine and the fact that an individual patient often has contact with many physicians, an individual physician often never knows the outcomes of patients that they encountered, resulting in a lack of both negative and positive feedback. For these reasons, two Institute of Medicine reports (McGinnis et al., 2013; National Academies of Sciences, Engineering, & Medicine, 2015; see also Rosner et al., 2023) have called for healthcare organizations to create better feedback systems.

In summary, accumulating experience with many patients is believed to be a critical aspect of how doctors become experts. However, due to working in specialized practices, chance encounters with certain patients but not others, and imperfect feedback systems, doctors will not always experience (and re-experience) certain types of patients. For this reason, one value of a longitudinal continuing certification program is to provide physicians with a well-rounded set of vignettes with feedback to supplement their real-world experiences. We recommend prioritizing gaps that are likely to occur in a physician's practice as well as those that have a meaningful impact on the quality of care delivered.

Maintaining expertise

Cognitive skills decline over time

Once learned, cognitive skills must be maintained: in the absence of study, learned information and procedures are forgotten over time. Decades of research suggest that, across content domains and types of tasks, forgetting tends to follow a negatively accelerated *power law* function such that a great deal of material is forgotten initially, but the remaining material is forgotten more slowly (Ebbinghaus, 1885, EL: 4; Rubin & Wenzel, 1996, EL: 2; Wickelgren, 1974, EL: 4; Wickens, 1998, EL: 6; Wixted, 2004, EL: 3; Wixted & Carpenter, 2007, EL: 3). That is, people forget much of what they see and hear in the initial minutes and hours afterward, somewhat more in the following days and weeks, and comparatively little of what remains in the months and years ahead (see Fig. 1). This power-law function may reflect the rate at which people are likely to stop re-encountering those topics (Anderson & Schooler, 1991, EL: 5).

In terms of medical expertise, the power law suggests that some knowledge will be retained relatively well over long periods of time, but it is almost inevitable that other material will be quickly forgotten after being encountered

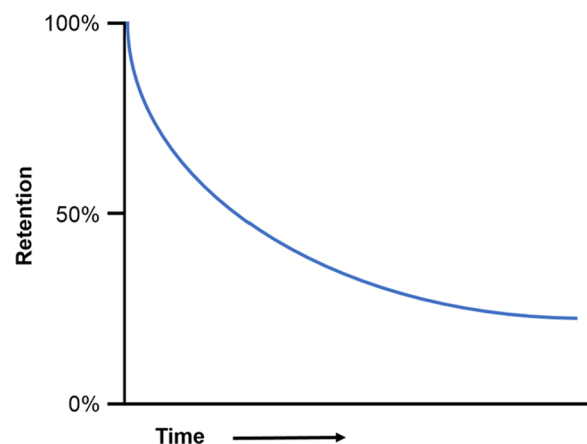


Fig. 1 Prototypical forgetting curve

in training if it is not deliberately practiced. How often important information needs to be practiced is likely to vary across individuals and be influenced by other variables. Nevertheless, the rapid decline in retention after any given study episode suggests that it is beneficial to distribute study over time to alleviate these losses.

Reasons for forgetting

Why do people forget? One intuitive hypothesis might be that we simply run out of mental “storage space” and that prior knowledge is forced out to make room for the new. And, it is indeed clear that there are sharp restrictions on how much can be held in *working memory*, or what we are currently thinking about (although the specification and cause of those limits remain debated; e.g., Cowan, 2010, EL: 2; Miller, 1956, EL: 2). However, it is not clear that there are practical limits on the total capacity of long-term learning and knowledge (Drachman, 2005; Landauer, 1986). Laboratory studies have demonstrated that people can readily learn and remember hundreds or even thousands of pictures or sentences even after only seconds of exposure to each (e.g., Shepard, 1967, EL: 4; Standing, 1973, EL: 4; Standing et al., 1970, EL: 4). Indeed, the cortex of the human brain contains approximately 150 trillion (1.5×10^{14}) synapses (Drachman, 2005, EL: 3). That is orders of magnitude more than what the average individual knows: 40,000 (4×10^4) words (Brybaert et al., 2016, EL: 3), 750 (7.5×10^2) people (Zheng et al., 2006, EL: 4), or, more generally, approximately 1 billion (10^9) bits of information (Landauer, 1986, EL: 2).

Another intuitive hypothesis as to why we forget is that skills and knowledge are lost to *decay*; that is, memories fade and are simply lost over time. This hypothesis receives more support; given the regularities in how memories decline over time (discussed above), it is likely that the passage of time contributes to forgetting

(Wixted, 2004: EL 2; Sadeh et al., 2014: EL 2). However, decay is not likely to be the *whole* story, since not all memories are fated to be lost over time: People can remember the names and locations of buildings on their college campus (Bahrick, 1983, EL: 5) or the names and faces of their high school classmates (Bahrick et al., 1975, EL: 5) even after decades of disuse.

Thus, cognitive psychologists often emphasize *interference* from other, similar information as an additional cause of forgetting. Many things commonly forgotten in daily life are those that compete with many other similar memories. For example, it is often difficult to remember where I left my phone this morning because I have many other competing memories of other places where I left my phone at different times. One experimental demonstration of interference is the *fan effect* (Anderson, 1974, EL: 3; Anderson & Reder, 1999, EL: 4): Learning multiple overlapping associations makes any individual association harder to retrieve (e.g., learning *the lawyer is in the cave* and *the lawyer is on the beach* is harder than learning *the lawyer is in the cave* and *the fireman is on the beach*). Thus, categorizing individual exemplars is more difficult (slower) for broader categories (e.g., “cancer” or “plants”) than for more narrow ones (e.g., “red-green colorblindness” or “flowers”; Landauer & Freedman, 1968, EL: 3; Landauer & Meyer, 1972, EL: 2; c.f., Collins & Quillian, 1970, EL: 3). Interference can happen both *proactively*, when old knowledge makes it harder to learn competing new knowledge (Watkins & Watkins, 1975, EL: 2), and *retroactively*, when new knowledge, once acquired, interferes with retrieving old knowledge (Postman & Underwood, 1973, EL: 2).

Indeed, interference-based failures to retrieve *some* information may be an inevitable consequence of remembering *other*, competing information (*retrieval-induced forgetting*; Anderson et al., 1994, EL: 2; Roediger, 1978, EL: 2). Imagine the process of diagnosing a patient with chest pain. Retrieving *myocardial infarction* in response to the cue *chest pain* reinforces thinking of *myocardial infarction* for future cases, and it also correspondingly weakens the likelihood of considering *aortic dissection* as a diagnosis. Thus, less common concepts and information are particularly vulnerable to interference (Anderson et al., 1994, EL: 3). This suggests it should be particularly important for physicians to practice similar and easily confusable concepts, especially those similar to more common concepts (e.g., common diagnoses) and thereby vulnerable to retrieval-induced forgetting.

It is not always adaptive to retrieve

The phenomenon of retrieval-induced forgetting relates to another key property of human memory: At least in some cases, it is beneficial or adaptive *not* to bring all of

one’s knowledge to mind (Bjork, 1989, EL: 2; Kuhl et al., 2007, EL: 4; MacLeod, 1998: EL 2; Nørby, 2005, EL: 2; Popov et al., 2019, EL: 3; Wimber et al., 2015, EL: 4). In the case of retrieval-induced forgetting, for instance, it is likely beneficial on the whole to prioritize frequently used facts and concepts over those less frequently used, so as to reduce interference and cognitive demands (Bäuml & Sameniéh, 2010, EL 4; Kuhl et al., 2007, EL: 4; Nørby, 2005, EL: 2; Popov et al., 2019, EL: 3; Wimber et al., 2015; EL: 4). For instance, standards of care change (as we discuss below), and it could be beneficial for physicians not to bring to mind outdated standards. Indeed, when explicitly told that some information is obsolete or otherwise should now be forgotten, people can prioritize study and retention of other, to-be-remembered information (for more discussion of the mechanisms of such *directed forgetting*, see MacLeod, 1998; EL: 2; Sahakyan et al., 2013, EL: 2). Forgetting the details of individual episodes or exemplars (e.g., individual patients) can also facilitate learning broader patterns or prototypes (e.g., diagnoses), supporting System 1 pattern recognition system, as discussed above (Nørby, 2005, EL: 2; Posner & Keele, 1968; EL: 3).

Thus, it is unlikely that it would be possible or even desirable to eliminate forgetting completely. Another implication is that is not necessarily advisable for physicians to try to remember every detail of every case, and they are perhaps better served by abstracting more general principles. Lastly, it may be valuable to explicitly highlight when standards of care or other information is out of date so that physicians can leverage directed forgetting to prioritize current, relevant knowledge and skills.

Inaccessible knowledge can often be recovered or relearned

Although people may sometimes be unable to bring to mind the desired knowledge or skills, that does not necessarily mean the learning is lost forever. Knowledge that is forgotten at one point in time can sometimes spontaneously be retrieved later (a phenomenon known as *hypermnesia*; Erdelyi & Becker, 1974, EL: 3). A classic example is the *tip-of-the-tongue* phenomenon (e.g., Burke et al., 1991, EL: 5), when one has a sense of knowing a particular word or name but being unable to retrieve it, only to spontaneously recover it later. Thus, failure to retrieve an idea at any point in time is not necessarily, or even likely, diagnostic of permanent loss. The fact that inaccessible knowledge and skills are not fully lost also leads to *savings* in that previously encountered knowledge can be relearned more quickly than it was initially acquired (Ebbinghaus, 1885, EL: 4; Nelson, 1978, EL: 4).

Although inaccessible knowledge can sometimes be retrieved spontaneously, it is more apt to be retrieved with appropriate *retrieval cues* (e.g., Tullis & Benjamin, 2015, EL: 2; Tullis & Fraundorf, 2017, EL: 3), characteristics of the environment that helps to “jog” one’s memory (although certain cues can be unhelpful if they disrupt a planned retrieval strategy; Basden & Basden, 1995, EL: 3; Roediger, 1978, EL: 3). In general, human memory is partially context-dependent, such that memories more readily come to mind when the environment relates to them or matches how they were initially learned or acquired (Bjork & Richardson-Klavehn, 1989, EL: 2). More frequent longitudinal assessment, rather than point-in-time assessment, could thus serve as a cue to keep this knowledge accessible and/or facilitate relearning.

Effects of age on memory and learning

Beyond the time that has elapsed since medical training, another source of skill decline may be aging. We first cover the basic science of aging and then address studies of aging specifically as it relates to physicians.

Age affects some cognitive skills more than others

A clear conclusion from the basic science of memory aging is that age differentially affects different types of knowledge. Beginning in early adulthood (e.g., age 20), performance steadily declines with age on tasks that require *fluid intelligence*; that is, those that involve novel learning or reasoning (Horn & Cattell, 1966, EL: 5; Horn & Cattell, 1967, EL: 5). This decline may be driven at least in part by declines in more fundamental aspects of cognition: The speed of even very basic cognitive processing (e.g., as measured by the speed of identifying whether two strings of letters are the same or different) declines with age, as does the ability to temporarily hold information in *working memory* (Park et al. 2002, EL: 5; Salthouse, 1991, EL: 5; Salthouse, 1996, EL: 2; Salthouse, 2004, EL: 2; Salthouse, 2005, EL: 5; Salthouse & Babcock, 1991, EL: 5; Stine-Morrow et al., 2008, EL: 5). For instance, declines in basic speed may drive age differences in more complex tasks insofar as cognitive skills may break down if people cannot retrieve or compute relevant information sufficiently quickly to be useful for the task at hand (Hertzog et al., 2003, EL: 5; Salthouse, 1991, EL: 5; Salthouse, 1996, EL: 2; Salthouse & Babcock, 1991, EL: 5; Salthouse, 2005, EL: 5; Stine-Morrow et al., 2008, EL: 5). Declines in processing speed are relevant to many areas of medicine because physicians often see high volumes of patients in a day, need to address multiple problems per visit, and in some settings need to switch quickly between patients. It is not enough simply to have acquired the relevant cognitive skills; physicians need to be able to bring to mind—or

know where to look up—the relevant knowledge in time to be practically useful.

By contrast, fixed knowledge, often referred to as *crystallized intelligence*, is preserved or even increases with age (Horn & Cattell, 1966, EL: 5; Horn & Cattell, 1967, EL: 5; Park et al., 2002, EL: 5; Salthouse, 2004, EL: 2; Zacks & Hasher, 2006, EL: 2). Even fixed knowledge may decline at especially advanced ages (e.g., age 80 or above; Park et al., 2002, EL: 5; Salthouse, 2004, EL: 2), but physicians would likely be retired at this age. In general, then, older adults rely less on novel (fluid) episodic learning and more on existing (crystallized) knowledge about the world (Castel, 2005, EL: 4; Castel, 2007, EL: 4; Castel et al., 2013, EL: 3; Koutstaal & Schacter, 1997, EL: 3; McGillivray & Castel, 2017, EL: 3; Stine-Morrow et al., 2008, EL: 4; Zacks & Hasher, 2006, EL: 3). This has mixed implications for the retention and use of medical expertise: On the one hand, physicians’ general medical knowledge might be expected to be relatively spared with age. On the other hand, older physicians might be less proficient at learning new techniques or remembering newly encountered cases and patients.

Further, although older adults underperform younger adults even in very basic memory tasks, age differences are larger in some types of learning and retrieval than others (Fraundorf et al., 2019, EL: 1). For instance, it has been argued that older adults are especially challenged by cognitive skills that require self-initiated or controlled processing, such as deliberately committing novel information to memory (e.g., learning new standards of care) or systematically reviewing one’s memory (e.g., deliberately considering each of a series of potential diagnoses). By comparison, age is less deleterious for relatively automatic or habitual uses of memory, such as applying a familiar set of actions (e.g., ordering a frequent diagnostic test) or recognizing a stimulus (e.g., recognizing a familiar set of symptoms as a particular disease) (e.g., Craik, 1986, EL: 2; Hoyer & Verhaeghen, 2005, EL: 2; Luo & Craik, 2008, EL: 2; c.f., Fraundorf et al., 2019, EL: 1). This pattern is consistent with age-related declines in the controlled, analytical System 2 but preserved or enhanced functioning of the automatic, experience-based System 1 (Eva, 2002, EL: 2; Eva, 2003, EL: 2). It suggests that older physicians may rely heavily on habitual, rather than new, cognitive skills and that they will better remember patients and treatments consistent with their experience.

Several other generalizations regarding memory aging highlight other situations where older physicians’ cognitive skills might be preserved. First, older adults perform comparatively well at remembering new information that is naturalistic (as opposed to arbitrary laboratory stimuli; Castel, 2007, EL: 4) or that allows the use of existing everyday memory strategies, such as establishing routines

or leaving reminders for oneself (Bailey et al., 2010, EL: 4; Moscovitch, 1982, EL: 5; Rendell & Craik, 2000, EL: 3; Rendell & Thomson, 1999, EL: 3). Second, older adults are *as* or even *more* effective than younger adults at working with familiar partners to remember information as a team. These *collaborative cognition* strategies can include dividing responsibilities for remembering different kinds of information and suggesting cues to support each other's memory (Dixon & Gould, 1996, EL: 3; Dixon, 1999, EL: 2). Third, older adults are sensitive to indicators of the *value* of to-be-retained information and perform comparatively well in remembering material that is important or that otherwise aligns with their motivational priorities. For instance, in laboratory experiments, older adults are adept at prioritizing material that prioritizing material that is worth more "points" toward a goal (Castel, 2007, EL: 3; Castel et al., 2002, EL: 3; Castel et al., 2007, EL: 3), that a speaker emphasizes intentionally (Fraundorf et al., 2012, EL: 4), that aligns with a motivational bias for positivity (Charles et al., 2003, EL: 3; Mather & Carstensen, 2005, EL: 3; May et al., 2005, EL: 3), or that comes from a more trustworthy source (Rahhal et al., 2002, EL: 3). Indeed, even non-physician older adults better remember fictive medications with severe side effects than those with less severe side effects (Hargis & Castel, 2018, EL: 3). All three of these age-related changes would be expected to favor retention of medical skill and learning even with increasing age insofar as physicians use their medical expertise in everyday life, often work with well-established teams, and (presumably) value their medical knowledge and skills.

However, a final generalization is that there is clear meta-analytic evidence that older adults are especially impaired in remembering the *source* or *context* of information (Fraundorf et al., 2019, EL: 1; Old & Naveh-Benjamin, 2008, EL: 1; Spencer & Raz, 1995, EL: 1). This could have deleterious consequences in medicine if older physicians confuse or misattribute the symptoms or treatments prescribed to several patients they have recently seen.

In sum, there is reason to be optimistic that physicians can retain much of their general medical knowledge with increasing age. However, older physicians may be vulnerable to reduced memory for specific cases or patients, and they may access their knowledge more slowly.

Aging as it relates to physicians

The role of aging in physicians' cognitive skills has been addressed in a number of narrative reviews (e.g., Ajmi & Aase, 2021, EL: 2; Eva, 2002, 2003; Durning et al., 2010; Williams, 2006; EL: 2, Council on Medical Education, 2015, EL: 2). Assessing the role of age in a physician's ability to provide high-quality care is quite complicated

because a number of factors are so highly correlated that it is usually impossible to distinguish them.

First, it is possible that a physician's abilities decline with age due to memory or processing decline. Second, it is possible that knowledge and skills could decline due to the passage of time out of medical school and residency; this variable is highly confounded with age among physicians insofar as most physicians enter medical school at roughly similar ages. Third, it is possible that as the number of years since residency increases, a physician's knowledge becomes out of date due to shifting standards that they did not learn in medical school or residency. Fourth, over time a physician accumulates more direct patient experience, which as explained above can have both positive and potentially negative impacts on performance. Fifth, some physicians specialize over time, which could lead them to lose broader skills that have become less relevant to their practice. In the following paragraphs, we unpack evidence relevant to aging physicians. When we mention correlations with age, we acknowledge that many other factors, explained above, are highly correlated with age. Thus, we are using "age" as a proxy variable, and this is not meant to implicate cognitive aging as the reason for these associations.

Reliable evidence indicates that the quality of health-care provided decreases with physician age. A systematic review of 62 studies found that 45 studies (73%) reported a decrease in performance for some or all outcomes (Choudhry et al., 2005, EL: 2). Another 13 (21%) found no association. The remaining four (6%) found a non-linear (inverted U) trend or an increase in some or all outcomes. This pattern held across a wide variety of measures, including knowledge measures, health outcomes, and adherence to standards of care for diagnosis, screening, prevention, and therapy. However, one potential limitation of this review is that it covered a period of time during which evidence-based medicine and quality-assurance techniques, such as performance evaluation, were becoming adopted. So, it is possible that the apparent age-related declines may instead be driven by the fact that the older physicians were trained prior to this shift and that the newer generation of physicians, who were trained to value evidence-based medicine, may not exhibit declines in quality of care as they age if they stay up to date with the evidence.

Since this systematic review, several notable studies reinforce this pattern of decreases in quality of care provided by older physicians. A population-based study of adherence to guidelines for antibiotic prescribing in treating urinary tract infections in children in Taiwan found that adherence dropped gradually from 87% in physicians younger than 35 to 45% in physicians older than 55 (Chen et al., 2011, EL: 5). Holmboe et al., (2008,

EL: 5) found that physicians more than 20 years out of medical school performed considerably worse on the maintenance of certification exam compared to physicians fewer than 20 years out. Physicians who scored lower on the assessment also exhibited worse performance on measures of treating patients: whether they had diabetes patients obtain eye exams, lipid tests, and HbA1c tests, whether they had female patients receive a mammogram in the past year, and whether they had patients with coronary artery disease obtain a lipid test in the past year. St-Onge et al. (2015, EL: 5) similarly found worse diagnostic performance for clinical vignettes among older physicians. In sum, there is evidence from multiple sources of a general decrease in both conceptual knowledge and quality of care with increased age.

However, the specific reasons for the decreasing quality of care with age is less certain. One possibility already discussed (see also Eva, 2002, EL: 2), is that performance decline may be caused by negative changes in cognitive processing. Another possibility is that older physicians fail to learn and/or retain changing standards of care. In fact, another study of scores on the ABIM test found that age predicted poorer performance on questions that tested knowledge for standards of care that had changed over the preceding 30 years, but age did not predict poorer performance on questions about standards of care that had not changed (Day et al., 1988, EL: 5; see also Holmboe et al., 2008). However, this study is quite dated, and it is not certain that this finding would still hold among the current cohort of physicians, who participate in different forms of continuing education than physicians 30 years ago. More studies of this nature could help elucidate why knowledge and performance appear to decline with age.

Although performance generally declines with age, there are some cases where it may not. System 1 (non-analytical processing or pattern recognition) may remain stable or even improve with age and experience; in the previous section, we discussed how more automatic forms of memory or habitual responses, as well as fixed knowledge, remain intact until advanced ages of 80 or above (see also Eva, 2002, EL: 2). Being able to continue to rely on automatic forms of memory accords with the important role of non-analytical medical decision making. Some studies have indeed found that older physicians tend to both identify correct diagnoses very quickly and settle on a diagnosis quickly. This is a double-edged sword. On the one hand, it can lead to quick and accurate diagnoses. For example, two studies (Eva et al., 2010, EL: 5; Hobus & Schmidt 1993, EL: 5; see discussion in Eva, 2002) found a positive relation between age/years of experience and diagnostic accuracy. However, quickly settling on a diagnosis can also lead to premature closure

(i.e., failing to consider alternatives after reaching a decision), and some research has found that older physicians focus more heavily on information presented earlier in a case (Eva & Cunningham, 2006, EL: 5).

In sum, the majority of the evidence suggests that older physicians perform worse in a variety of ways, even though gaining experience over one's career may mitigate this decline to some extent. However, doctors are tasked not only with maintaining current standards of care but also keeping up with changing standards, which we discuss in the next section.

Keeping up with changing standards of care

One of the fundamental challenges in medicine is keeping up with the ever-changing standards of care. An Institute of Medicine report (McGinnis et al., 2013) concluded that diagnostic and treatment options are changing at an accelerating rate, making it ever more important to keep up with changing standards. Two major reviews have systematized the barriers to using current standards of care (Cabana et al., 1999, EL: 2; Cochrane et al., 2007, EL: 2). Though the reviews differ in many ways, there is substantial agreement in terms of the cognitive and attitudinal barriers identified. Imagine a physician learning about a new treatment standard of care. First, the physician must become *familiar* with and *aware* of this new standard of care. Second, the physician must develop knowledge or skill, for example, knowledge about indications and dosages of a therapy. Third, the physician must form a high *outcome expectancy* (believe the treatment or standard would be beneficial) and *agree* with the new standard of care. Fourth, the physician should feel confident they can implement it, termed *self-efficacy*. By contrast, a physician may feel (for example) uncomfortable providing treatment for a condition at the boundary of their scope of practice. Fifth, the physician must overcome *habits* or *inertia*, doing things the same way as they have always been done.

How might a longitudinal assessment program affect these barriers? In Fraundorf et al. (2022), we discuss the overwhelming evidence that repeated testing benefits learning and retention and protects against interference from previously learned practices. Thus, longitudinal assessment would likely improve familiarity, awareness, and knowledge. It is less clear whether longitudinal assessment could impact the attitudinal barriers, such as outcome expectancy and self-efficacy, though it is possible that providing feedback (correct answers, explanations, and citations) might alleviate attitudinal barriers. Aside from longitudinal assessment, continuing medical education (CME) is the primary system currently in place designed to help physicians maintain cognitive skills and gain new skills. In the first article in this

collection (Rottman et al., 2023), we discuss the strength and limitations of CME and compare CME to longitudinal assessment.

Proposed studies and future directions

Providing feedback about strengths and weaknesses and tracking age gaps over time

Two benefits of a longitudinal assessment program are that it can potentially help physicians learn about standards of care that have changed since their training and that it can provide useful feedback to physicians about their relative strengths and weaknesses. We propose that Boards prospectively classify items as testing new standards of care versus testing old—but still relevant—standards of care. This classification can then be used in three ways.

First, physicians can be provided with feedback about performance on these two different types of questions. This will provide physicians with a sense of whether they are challenged more by staying current versus by maintaining older knowledge.

Second, we propose that boards use an approach pioneered in a clever study (Day et al., 1988). This study looked at older versus younger physicians' performance on the continuing certification program assessment, contrasting questions for which standards of care have changed over time versus questions for which they have not. The finding was that older physicians performed worse for questions testing knowledge about standards that had changed, but not for standards that had remained the same. Ideally, if the assessment program works to keep physicians up to date, this interaction for older vs. younger physicians on changed vs. unchanged standards should diminish over time. This analysis could be conducted both before and after implementing the longitudinal assessment program to evaluate the contributions of the educational component of the longitudinal program. And, it could be conducted on an ongoing basis to measure the success of the program over time, with the goal of continuously optimizing the test to minimize age differences.

Third, extending this analysis pioneered by Day et al. (1988) can help to uncover the reasons for poor performance. In particular, Day et al. found decreases in performance over time only on questions for which standards of care have changed over time, which seems to implicate challenges of staying current rather than aging or time since residency *per se*. However, since 1988, the landscape of CME has changed considerably, so it is not clear whether the same pattern would be found. Furthermore, one possibility is that physicians selectively keep up with standards that they think are especially relevant to their practice. This possibility could be assessed by having

physicians rate the relevance of each question, and testing whether relevance interacts with age and whether or not a standard has changed. Still other analyses would be possible if physicians also rate their confidence in their answers. For example, one possibility is that if a physician is wrong on a question that involves a new standard, but is highly confident, that might mean old knowledge is interfering with learning new knowledge or that they never learned the new standards. In contrast, if a physician is wrong but not very confident on a question involving a new standard, that might indicate that a new standard has not been learned.

Measuring response time during testing

Some prior research has examined the relationships between age, response time, accuracy, and case difficulty (Barrows et al., 1982, EL: 5; Elstein et al., 1978, EL:5; Gruppen et al., 1988, EL: 5; Norman et al., 1989, EL: 5; Pelaccia, et al., 2014, EL: 5). However, these findings are nuanced and not entirely consistent, and this research could benefit from broader case materials and from larger, more representative samples of physicians across many specialties. In fact, given that many items on continuing certification program assessments already incorporate questions about diagnosis and treatment, and response time is easy to record in a computer system, data could be easily obtained to test these relationships.

Identifying out-of-date information

As we discussed above, laboratory evidence indicates that learners have the capability to engage in directed forgetting of material that has been explicitly cued as to-be-forgotten (e.g., because it is out of date or incorrect). There may be opportunities to leverage this capacity in longitudinal assessment; for example, by presenting outdated standards of care and explicitly indicating they are no longer current and should not be retained. We hypothesize that this should lead to better retention of current standards of care than a procedure in which out-of-date material is not explicitly addressed.

Debiasing and clinical reasoning interventions

A more open-ended suggestion is to consider using longitudinal assessment programs as a platform to test debiasing and clinical reasoning interventions. In particular, the dual-process theory presumes that there are two systems of reasoning, one that is faster and one that is slower, and that these systems need to be coordinated (Croskerry, 2009a, EL: 2; 2009b, EL: 2; Norman & Eva, 2010, EL: 2; Pelaccia et al., 2011, EL: 2). For example, one proposal is that physicians need to learn how to switch from faster automatic judgment for routine problems to slower, more effortful reasoning for more unusual or ill-defined

problems (Moulton et al., 2007; see also Graber, 2009). Croskerry and colleagues have reviewed potential ways to attempt to teach physicians to avoid common biases (e.g., Croskerry et al., 2013a; b).

Though several debiasing and clinical reasoning interventions have been proposed and tested, there is only mixed evidence whether they work (e.g., Schmidt & Mamede, 2015, EL: 2; see also Isler et al., 2020, for research on debiasing training outside medicine). Still, a number of consenses still express interest in such interventions (Olson et al., 2019; National Academies of Sciences, Engineering, & Medicine, 2015, pp. 4–32 to 4–34; Parodis et al., 2021), and new debiasing interventions are being tested with some promise (Kuhn et al., 2020; Mamede et al., 2020). Longitudinal assessment programs could serve as a testing ground for brief interventions that could be embedded right before or during a question.

In summary, there are a number of potential ways that data collected from longitudinal assessment programs, or from interventions embedded inside the programs, could be used to assess the efficacy of the programs themselves, provide guidance about how to improve feedback, and test basic questions about medical expertise and diagnosis that are hard to study in other settings and for which studies are rare and small. Many of these proposals can be accomplished with minimal changes to the duration of the program, and physicians may find them insightful if the results can demonstrate strong evidence regarding the roles of aging and keeping up with standards of care, speed versus accuracy, and debiasing techniques.

Examining basic mechanisms of non-analytical reasoning in diagnosis

Longitudinal assessment also provides opportunities to further uncovering the basic mechanisms of non-analytical reasoning in medical decision making. Experiments on the role of non-analytical reasoning in diagnosis can be directly embedded into a longitudinal assessment program and, by doing so, would help to firmly establish an empirical base of knowledge regarding non-analytical reasoning in medicine. The research on non-analytical reasoning in medicine has only been tested in a few studies with only modest sample sizes (Brooks et al., 1991, EL: 4; Hatala et al., 1999, EL: 4; Young et al., 2011, EL: 4). Though these findings are intuitive, and though they build upon an extensive literature on the basic science of categorization from cognitive science (Cohen & Lefebvre, 2017), there exist important gaps in knowledge. First, most of the basic science research has been conducted with abstract stimuli, and with undergraduates who are trained for only short periods of time, not complex real-world medical stimuli that require years for physicians to

master. Thus, conducting more and larger studies with physicians will help to establish these phenomena within medicine with more certainty.

A longitudinal assessment also provides a remarkable opportunity to design studies embedded into the training programs to test the role of irrelevant information from prior cases in biasing the diagnosis of future cases. This research could test how long the bias lasts, how strong the bias is, how the prevalence of a disease or a physician's knowledge of a disease affects the bias, and whether age of the physician interacts with non-analytical reasoning. Further, given that most of the literature on non-analytical reasoning in diagnosis has focused on diagnosis based on visual information (e.g., reading ECGs, pathology, dermatology, radiology), another question is whether the bias is different for diagnosis that requires integrating multiple signs, symptoms, and lab reports (emergency medicine, internal medicine, etc.; Norman et al., 2007).

In sum, although there is broad consensus that non-analytic reasoning plays some role in diagnosis, we know comparatively little about *when*, *where*, and *how much* it matters. Delineating how experience both bolsters and biases subsequent decisions could help make physicians more aware of how such non-analytic factors could impact decision making, which could increase motivation to follow evidence-based guidelines or be used in debiasing efforts.

The effect of feedback on outcome expectancy and self-efficacy

Another opportunity is to study the feedback provided that explains whether an answer is right or wrong. Among the multiple barriers to keeping up with changing standards of care, the main barriers that a longitudinal assessment program is intended to address are awareness, familiarity, and knowledge about the new standard of care. However, it is possible that feedback could also address other barriers, such as not believing that the new standard is better (outcome expectancy) and not being confident in how to implement it (self-efficacy). Experiments could be designed that manipulate the provided feedback, and questions could be embedded about a physician's feelings of outcome expectancy and self-efficacy in order to test ways to maximize the broader utility of the feedback for overcoming multiple barriers.

Summary and conclusion

We discussed four topics related to how physicians acquire, maintain, and update cognitive skills. First, we reviewed the dual-process theory of medical expertise, which proposes that medical decision making is a combination of (a) fast intuitive thinking (non-analytical processing) shaped by experience and (b) slow analytical

thinking guided by logic. According to this theory, the benefits of non-analytical thinking are that decisions can be reached very fast and are often correct. However, intuitive, non-analytical thinking also has downsides. Idiosyncratic experiences can shape a physician's decisions (e.g., about treatment decisions). And, idiosyncrasies in which patients a physician does and does not see affect the maintenance of expertise; for instance, they could distort the physician's beliefs about the prevalence of a diagnosis and the likelihood of that diagnosis coming to mind.

Second, we discussed the basic science of why people forget or fail to retrieve information. Forgetting can sometimes be an adaptive in that it allows people to discard information that is out of date or less important, and forgetting individual details can help people to learn broader patterns. Nevertheless, it is clear that people sometimes fail to bring relevant information to mind. But, this does not necessarily indicate it is permanently lost. Rather, it may become accessible again later, especially with the right cues. One important cause of retrieval failures is the inability to access knowledge quickly enough to be useful. Another is interference from competing concepts and skills. The literatures both on interference and on analytical thinking point toward an opportunity for a longitudinal continuing certification program to attempt to fill in potential gaps in experience by testing cases that are somewhat rare but of high importance to patient care.

Third, whereas crystallized intelligence (e.g., medical knowledge) remains intact until age 80, fluid intelligence (e.g., novel learning or using balancing multiple tasks in working memory) declines with age. The majority of the evidence suggests that the quality of healthcare declines with physician age. This could be driven in part by the decline in fluid intelligence; however, another contributor could be a failure to keep up with changing standards of care.

Fourth, we reviewed physician-level barriers to keeping up with changing standards of care, including not being aware of or knowledgeable about a new standard, not believing that the new standard is better, not being confident in how to implement the new standard, and habits. The goal of continuing certification programs has traditionally been to assess whether physicians are keeping up with changing standards. The switch to a longitudinal assessment program presents the opportunity of serving both as assessment and as an educational program to make physicians more knowledgeable about new standards and their application in patient care.

Acknowledgements

We thank Andrew Bazemore, Rebecca S. Lipner, David B. Swanson, and Thomas O'Neill for feedback on earlier drafts of this work.

Author contributions

ZC, SF, and BR wrote the first draft of the manuscript. TN-M provided feedback. All authors contributed to revising the manuscript.

Funding

This work was funded by a grant from the American Board of Internal Medicine (ABIM), American Board of Medical Specialties (ABMS), and American Board of Family Medicine (ABFM). Individuals from ABIM, ABMS, and ABFM provided feedback on the overall goals of the review and on earlier drafts of the manuscript, but approval of the final manuscript rested with the authors alone.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors were not involved with the peer review process of this work.

Received: 1 March 2022 Accepted: 20 June 2023

Published online: 25 July 2023

References

- Ajmi, S. C., & Aase, K. (2021). Physicians' clinical experience and its association with healthcare quality: A systematised review. *BMJ Open Quality*, *10*(4), e001545.
- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451–474.
- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186–197.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087.
- Bahrick, H. P. (1983). The cognitive map of a city: Fifty years of learning and memory. *The Psychology of Learning and Motivation*, *17*, 125–163.
- Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*(1), 54–75.
- Bailey, P. E., Henry, J. D., Rendell, P. G., Phillips, L. H., & Kliegel, M. (2010). Dismantling the "age-prospective memory paradox": The classic laboratory paradigm simulated in a naturalistic setting. *Quarterly Journal of Experimental Psychology*, *63*(4), 646–652.
- Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1982). The clinical reasoning of randomly selected physicians in general medical practice. Clinical and investigative medicine. *Clinical and Investigative Medicine*, *5*(1), 49–55.
- Basden, D. R., & Basden, B. H. (1995). Some tests of the strategy disruption interpretation of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1656.
- Bäuml, K.-H.T., & Sameni, A. (2010). The two faces of memory retrieval. *Psychological Science*, *21*(6), 793–795.
- Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309–330). Erlbaum.
- Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Flowerree symposium on cognition* (pp. 313–344). Lawrence Erlbaum Associates Inc.

- Boshuizen, H. P., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16(2), 153–184.
- Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, 120(3), 278–287.
- Brybaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, 7, 1116.
- Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, 30(5), 542–579.
- Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P. A. C., & Rubin, H. R. (1999). Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*, 282(15), 1458–1465.
- Casscells, W., Schoenberger, A., & Graboyes, T. B. (1978). Interpretation by physicians of clinical laboratory tests. *The New England Journal of Medicine*, 299(18), 999–1001.
- Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, 20(4), 718–721.
- Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. *Psychology of Learning and Motivation*, 48, 225–270.
- Castel, A. D., Benjamin, A. S., Craik, F. I., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory and Cognition*, 30(7), 1078–1085.
- Castel, A. D., Farb, N. A., & Craik, F. I. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory and Cognition*, 35(4), 689–700.
- Castel, A. D., McGillivray, S., & Worden, K. M. (2013). Back to the future: Past and future era-based schematic support and associative memory for prices in younger and older adults. *Psychology and Aging*, 28(4), 996–1003.
- Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, 132(2), 310–324.
- Chen, C. C., Wu, L. C., Li, C. Y., Liu, C. K., Woung, L. C., & Ko, M. C. (2011). Non-adherence to antibiotic prescription guidelines in treating urinary tract infection of children: A population-based study in Taiwan. *Journal of Evaluation in Clinical Practice*, 17(6), 1030–1035.
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152.
- Choudhry, N. K., Anderson, G. M., Laupacis, A., Ross-Degnan, D., Normand, S. L. T., & Soumerai, S. B. (2006). Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: Matched pair analysis. *BMJ*, 332(7534), 141–145.
- Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic review: The relationship between clinical experience and quality of health care. *Annals of Internal Medicine*, 142(4), 260–273.
- Cochrane, L. J., Olson, C. A., Murray, S., Dupuis, M., Tooman, T., & Hayes, S. (2007). Gaps between knowing and doing: Understanding and assessing the barriers to optimal health care. *Journal of Continuing Education in the Health Professions*, 27(2), 94–102.
- Coderre, S., Mandin, H. H. P. H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education*, 37(8), 695–703.
- Cohen, H., & Lefebvre, C. (Eds.). (2017). *Handbook of categorization in cognitive science* (2nd ed.). Elsevier.
- Collins, A. M., & Quillian, M. R. (1970). Does category size affect categorization time? *Journal of Verbal Learning and Verbal Behavior*, 9(4), 432–438.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1), 51–57.
- Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Flix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 409–422). North-Holland.
- Croskerry, P. (2009a). Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. *Advances in Health Sciences Education*, 14(1), 27–35.
- Croskerry, P. (2009b). A universal model of diagnostic reasoning. *Academic Medicine*, 84(8), 1022–1028.
- Croskerry, P., Singhal, G., & Mamede, S. (2013a). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality and Safety*, 22(Suppl 2), ii58–ii64.
- Croskerry, P., Singhal, G., & Mamede, S. (2013b). Cognitive debiasing 2: Impediments to and strategies for change. *BMJ Quality and Safety*, 22(Suppl 2), ii65–ii72.
- Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences*, 112(45), 13817–13822.
- De Neys, W. (2021). On dual-and single-process models of thinking. *Perspectives on Psychological Science*, 16(6), 1412–1427.
- Dixon, R. A. (1999). Exploring cognition in interactive situations: The aging of N+1 minds. In T. M. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 267–290). Elsevier.
- Dixon, R. A., & Gould, O. N. (1996). Adults telling and retelling stories collaboratively. In P. B. Baltes & U. M. Staudinger (Eds.), *Interactive inds: Life-span perspectives on the social foundation of cognition* (pp. 221–241). Cambridge University Press.
- Drachman, D. A. (2005). Do we have brain to spare? *Neurology*, 64(12), 2004–2005.
- Durning, S. J., Artino, A. R., Holmboe, E., Beckman, T. J., van der Vleuten, C., & Schuwirth, L. (2010). Aging and cognitive performance: Challenges and implications for physicians practicing in the 21st century. *Journal of Continuing Education in the Health Professions*, 30(3), 153–160.
- Eddy, D. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge University Press.
- Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *BMJ*, 324(7339), 729–732.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press.
- Erdelyi, M. H., & Becker, J. (1974). Hypernesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, 6(1), 159–171.
- Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: A perspective from the expert-performance approach with deliberate practice. *Academic Medicine*, 90(11), 1471–1486.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406.
- Eva, K. W. (2002). The aging physician: Changes in cognitive processing and their impact on medical practice. *Academic Medicine*, 77(10), S1–S6.
- Eva, K. W. (2003). Stemming the tide: Cognitive aging theories and their implications for continuing education in the health professions. *Journal of Continuing Education in the Health Professions*, 23(3), 133–140.
- Eva, K. W., & Cunningham, J. P. (2006). The difficulty with experience: Does practice increase susceptibility to premature closure? *Journal of Continuing Education in the Health Professions*, 26(3), 192–198.
- Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278.
- Fraundorf, S. H., & Benjamin, A. S. (2016). Conflict and metacognitive control: The mismatch-monitoring hypothesis of how others' knowledge states affect recall. *Memory*, 24(8), 1108–1122.
- Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: A meta-analysis. *Psychological Bulletin*, 145(4), 339–371.
- Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging*, 27(1), 88–98.
- Graber, M. L. (2009). Educational strategies to reduce diagnostic error: Can you teach this stuff? *Advances in Health Sciences Education*, 14(1), 63–69.
- Groen, G. J., & Patel, V. L. (1985). Medical problem-solving: Some questionable assumptions. *Medical Education*, 19(2), 95–100.

- Hargis, M. B., & Castel, A. D. (2018). Younger and older adults' associative memory for medication interactions of varying severity. *Memory*, 26(8), 1151–1158.
- Hatala, R., Norman, G. R., & Brooks, L. R. (1999). Influence of a single example on subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine*, 11(2), 110–117.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hertzog, C., Dixon, R. A., Hulstsch, D. F., & MacDonald, S. W. S. (2003). Latent change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory? *Psychology and Aging*, 18, 755–769.
- Holmboe, E. S., Lipner, R., & Greiner, A. (2008). Assessing quality of care: Knowledge matters. *JAMA*, 299(3), 338–340.
- Horn, J. L., & Cattell, R. B. (1966). Age differences in primary mental ability factors. *Journal of Gerontology*, 21(2), 210–220.
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129.
- Hoyer, W. J., & Verhaeghen, P. (2005). Memory aging. In J. Birren & K. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 209–232). Elsevier.
- Isler, O., Yilmaz, O., & Dogruoy, B. (2020). Active reflective thinking with decision justification and debiasing training. *Judgment and Decision Making*, 15(6), 926–938.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment*, 49(49–81), 74.
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6), 515–526.
- Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, 37(4), 555–583.
- Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience*, 10(7), 908–914.
- Kuhn, J., van den Berg, P., Mamede, S., Zwaan, L., Diemers, A., Bindels, P., & van Gog, T. (2020). Can we teach reflective reasoning in general-practice training through example-based learning and learning by doing? *Health Professions Education*, 6(4), 506–515.
- Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science*, 10(4), 477–493.
- Landauer, T. K., & Freedman, J. L. (1968). Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior*, 7(2), 291–295.
- Landauer, T. K., & Meyer, D. E. (1972). Category size and semantic-memory retrieval. *Journal of Verbal Learning and Verbal Behavior*, 11(5), 539–549.
- Ledley, R. S., & Lusted, L. B. (1959). Reasoning foundations of medical diagnosis: Symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science*, 130, 9–21.
- Luo, L., & Craik, F. I. (2008). Aging and memory: A cognitive approach. *The Canadian Journal of Psychiatry*, 53, 346–353.
- MacLeod, C. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Lawrence Erlbaum Associates Publishers.
- Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science*, 25(8), 1608–1618.
- Mamede, S., de Carvalho-Filho, M. A., de Faria, R. M. D., Franci, D., Nunes, M. D. P. T., Ribeiro, L. M. C., Biegelmeyer, J., Zwaan, L., & Schmidt, H. G. (2020). 'Immunising' physicians against availability bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Quality and Safety*, 29(7), 550–559.
- Marewski, J. N., & Gigerenzer, G. (2012). Heuristic decision making in medicine. *Dialogues in Clinical Neuroscience*, 14(1), 77–89.
- Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*, 9(10), 496–502.
- May, C. P., Rahhal, T., Berry, E. M., & Leighton, E. A. (2005). Aging, source memory, and emotion. *Psychology and Aging*, 20(4), 571–578.
- McGillivray, S., & Castel, A. D. (2017). Older and younger adults' strategic control of metacognitive monitoring: The role of consequences, task experience, and prior knowledge. *Experimental Aging Research*, 43(3), 233–256.
- McGinnis, J. M., Stuckhardt, L., Saunders, R., & Smith, M. (Eds.). (2013). *Best care at lower cost: The path to continuously learning health care in America*. National Academies Press.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The limitations of the GRE in predicting success in biomedical graduate school. *PLoS ONE*, 12(1), e0166742.
- Moulton, C. E., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007). Slowing down when you should: A new model of expert judgment. *Academic Medicine*, 82(10), S109–S116.
- National Academies of Sciences, Engineering, and Medicine. (2015). *Improving diagnosis in health care*. National Academies Press.
- Nelson, T. O. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory*, 4(5), 453–468.
- Neufeld, V., Norman, G., Feightner, J., & Barrows, H. (1981). Clinical problem-solving by medical students: A cross-sectional and longitudinal analysis. *Medical Education*, 15(5), 315–322.
- Nilsson, H., Juslin, P., & Olsson, H. (2008). Exemplars in the mist: The cognitive substrate of the representativeness heuristic. *Scandinavian Journal of Psychology*, 49(3), 201–212.
- Nørby, S. (2005). Why forget? On the adaptive value of memory loss. *Perspectives on Psychological Science*, 10(5), 551–578.
- Norman, G. R. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education*, 39(4), 418–427.
- Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical Education*, 44(1), 94–100.
- Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology*, 125(8), 1063–1068.
- Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: The role of experience. *Medical Education*, 41(12), 1140–1145.
- Old, S. R., & Naveh-Benjamin, M. (2008). Memory for people and their actions: Further evidence for an age-related associative deficit. *Psychology and Aging*, 23(2), 467–472.
- Olson, A., Rencic, J., Cosby, K., Ruz, D., Papa, F., Croskerry, P., Zierler, B., Harkless, G., Giuliano, M. A., Schoenbaum, S., & Colford, C. (2019). Competencies for improving diagnosis: An interprofessional framework for education and training in health care. *Diagnosis*, 6(4), 335–341.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, 17(2), 299–320.
- Parodis, I., Andersson, L., Durning, S. J., Hege, I., Knez, J., Kononowicz, A. A., Lidskog, M., Petreski, T., Szopa, M., & Edelbring, S. (2021). Clinical reasoning needs to be explicitly addressed in health professions curricula: recommendations from a European consortium. *International Journal of Environmental Research and Public Health*, 18(21), 11202.
- Pauker, S., & Kassirer, J. (1980). The threshold approach to clinical decision making. *New England Journal of Medicine*, 302, 1109–1117.
- Pelaccia, T., Tardif, J., Triby, E., Ammirati, C., Bertrand, C., Dory, V., & Charlin, B. (2014). How and when do expert emergency physicians generate and evaluate diagnostic hypotheses? A qualitative study using head-mounted video cued-recall interviews. *Annals of Emergency Medicine*, 64(6), 575–585.
- Pelaccia, T., Tardif, J., Triby, E., & Charlin, B. (2011). An analysis of clinical reasoning through a recent and comprehensive approach: The dual-process theory. *Medical Education Online*, 16(1), 5890.
- Popov, V., Marevic, I., Rummel, J., & Reder, L. M. (2019). Forgetting is a feature, not a bug: Intentionally forgetting some things helps us remember others by freeing up working memory resources. *Psychological Science*, 30(9), 1303–1317.
- Posner, M., & Keele, S. K. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3), 353–363.
- Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, 1(1), 19–40.

- Rahhal, T. A., May, C. P., & Hasher, L. (2002). Truth and character: Sources that older adults can remember. *Psychological Science*, *13*(2), 101–105.
- Rendell, P. G., & Craik, F. I. (2000). Virtual week and actual week: Age-related differences in prospective memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *14*(7), S43–S62.
- Rendell, P. G., & Thomson, D. M. (1999). Aging and prospective memory: Differences between naturalistic and laboratory tasks. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *54*(4), P256–P269.
- Roediger, H. L. (1978). Recall as a self-limiting process. *Memory and Cognition*, *6*(1), 54–63.
- Rosner, B. I., Zwaan, L., & Olson, A. P. (2023). Imagining the future of diagnostic performance feedback. *Diagnosis*, *10*(1), 31–37.
- Rottman, B. M. (2017). Physician Bayesian updating from personal beliefs about the base rate and likelihood ratio. *Memory and Cognition*, *45*(2), 270–280.
- Rottman, B. M., Prochaska, M. T., & Deaño, R. C. (2016). Bayesian reasoning in residents' preliminary diagnoses. *Cognitive Research: Principles and Implications*, *1*(5), 1–5.
- Rottman, B. M., Caddick, Z. A., Nokes-Malach, T. J., & Fraundorf, S. H. (2023). Cognitive perspectives on maintaining physicians' medical expertise: I. Reimagining Maintenance of Certification to promote lifelong learning. *Cognitive Research: Principles and Implications* [advance online publication]. <https://doi.org/10.1186/s41235-023-00496-9>.
- Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734–760.
- Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Sciences*, *18*(1), 26–36.
- Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab B. (2013). *List-Method Directed Forgetting in Cognitive and Clinical Research*. Elsevier (pp. 131–189).
- Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science*, *2*, 179–183.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.
- Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*(4), 140–144.
- Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*, 532–545.
- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776.
- Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: A narrative review and a proposal. *Medical Education*, *49*(10), 961–973.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.
- Singh, M. (2021). Heuristics in the delivery room. *Science*, *374*(6565), 324–329.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3.
- Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, *10*(4), 527–539.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2,500 visual stimuli. *Psychonomic Science*, *2*, 43–53.
- Stine-Morrow, E. A., Soederberg Miller, L. M., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging*, *23*(1), 131–153.
- St-Onge, C., Landry, M., Xhignesse, M., Voyer, G., Tremblay-Lavoie, S., Mamede, S., Schmidt, H., & Rikers, R. (2016). Age-related decline and diagnostic performance of more and less prevalent clinical cases. *Advances in Health Sciences Education*, *21*(3), 561–570.
- Tullis, J. G., & Benjamin, A. S. (2015). Cueing others' memories. *Memory and Cognition*, *43*(4), 634–646.
- Tullis, J. G., & Fraundorf, S. H. (2017). Predicting others' memory performance: The accuracy and bases of social metacognition. *Journal of Memory and Language*, *95*, 124–137.
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442–452.
- Whelehan, D. F., Conlon, K. C., & Ridgway, P. F. (2020). Medicine and heuristics: Cognitive biases and medical decision-making. *Irish Journal of Medical Science*, *189*, 1477–1484.
- Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory and Cognition*, *2*(4), 775–780.
- Wickens, T. D. (1998). On the form of the retention function: Comment on Rubin and Wenzel (1996): A quantitative description of retention. *Psychological Review*, *105*(2), 379–386.
- Williams, B. W. (2006). The prevalence and special educational requirements of dyscompetent physicians. *Journal of Continuing Education in the Health Professions*, *26*(3), 173–191.
- Wimber, M., Alink, A., Charest, I., Kriegeskorte, N., & Anderson, M. C. (2015). Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nature Neuroscience*, *18*, 582–589.
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235–269.
- Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*, *18*(2), 133–134.
- Young, M. E., Brooks, L. R., & Norman, G. R. (2011). The influence of familiar non-diagnostic information on the diagnostic decisions of novices. *Medical Education*, *45*(4), 407–414.
- Zheng, T., Salganik, M. J., & Gelman, A. (2006). How many people do you know in prison? Using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association*, *101*(474), 409–423.
- Council on Medical Education. (2015). *Competency and the aging physician* [White paper]. American Medical Association. <https://www.cppph.org/wp-content/uploads/2016/02/AMA-Council-on-Medical-Education-Aging-Physician-Report-2015.pdf>
- Day, S. C., Norcini, J. J., Webster, G. D., Viner, E. D., & Chirico, A. M. (1988). The effect of changes in medical knowledge on examination performance at the time of recertification. In *Research in medical education: Proceedings of the annual conference on research in medical education* (Vol. 27, pp. 139–144).
- Ebbinghaus, H. (1885). Ueber das Gedächtnis.
- Fraundorf, S. H., Caddick, Z. A., Nokes-Malach, T. J., & Rottman, B. M. (2022). *Cognitive perspectives on maintaining physicians' medical expertise: IV. Best practices and open questions in using testing to enhance learning and retention*. Manuscript submitted for publication.
- Gruppen, L. D., Woolliscroft, J. O., & Wolf, F. M. (1988). The contribution of different components of the clinical encounter in generating and eliminating diagnostic hypotheses. In *Research in medical education: Proceedings of the annual conference on research in medical education* (Vol. 27, pp. 242–247).
- Hobus, P. P. M., & Schmidt, H. G. (1993). In Schmidt, H. G. (Ed.), *The encapsulation framework in the presentation of physicians' recall of clinical cases*. Presented at the annual meeting of the American Educational Research Association. Seattle, Washington, April 10–14, 2001.
- Kool, W., Cushman, F. A., & Gershman, S. J. (2018). Competition and cooperation between multiple reinforcement learning systems. *Goal-Directed Decision Making*, 153–178.
- Moscovitch, M. (1982). A neuropsychological approach to perception and memory in normal and pathological aging. In *Aging and cognitive processes* (pp. 55–78). Springer.
- De Neys, W. (2022). Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, 1–68.
- Zacks, R. T., & Hasher, L. (2006). Aging and long-term memory: Deficits are not inevitable. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 162–177).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.