Sandra Monteiro*, Jonathan Sherbino, Jonathan S. Ilgen, Emily M. Hayden, Elizabeth Howey and Geoff Norman

The effect of prior experience on diagnostic reasoning: exploration of availability bias

Abstract

Objectives: Diagnostic reasoning has been shown to be influenced by a prior similar patient case. However, it is unclear whether this process influences diagnostic error rates or whether clinicians at all experience levels are equally susceptible. The present study measured the influence of specific prior exposure and experience level on diagnostic accuracy.
Methods: To create the experience of prior exposure, participants (pre-clerkship medical students, emergency medicine residents, and faculty) first verified diagnoses of clinical vignettes. The influence of prior exposures was measured using equiprobable clinical vignettes; indicating two diagnoses. Participants diagnosed equiprobable cases that were: 1) matched to exposure cases (in one of three conditions: a) similar patient features, similar clinical features; b) dissimilar patient features, similar clinical features; c) similar patient features, dissimilar clinical features), or 2) not matched to any prior case (d) no exposure). A diagnosis consistent with a matched exposure case was scored correct. Cases with no prior exposure had no matched cases, hence validated the equiprobable design.
Results: Diagnosis A represented 47% of responses in *condition d*, but there was no influence of specific similarity of patient characteristics for Diagnosis A, $F(3,712)=7.28$, $p=0.28$ or Diagnosis B, $F(3,712)=4.87$, $p=0.19$. When re-scored based on matching both equiprobable diagnoses, accuracy was high, but favored faculty (n=40) 98%, and residents (n=39) 98% over medical students (n=32) 85%, $F(2,712)=35.6$, $p<0.0001$. Accuracy for medical students was 84, 87, 94, and 73% for conditions a–d, respectively, interaction $F(2,712)=3.55$, $p<0.002$.
Conclusions: The differential diagnosis of pre-clerkship medical students improved with prior exposure, but this was unrelated to specific case or patient features. The accuracy of medical residents and staff was not influenced by prior exposure.

Keywords: availability bias; clinical expertise; diagnostic error; diagnostic reasoning.

*Corresponding author: Sandra Monteiro, PhD, 1280 Main Street West, 5th Floor, 5002 A/E, David Braley Health Sciences Centre, Hamilton, ON, L8S 4L8, Canada; Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada; and McMaster Faculty of Health Sciences Education Research, Innovation and Theory (MERIT) Program, McMaster University, Hamilton, ON, Canada, Phone: 905 525 9140, Fax: 905 572 7099, E-mail: monteisd@mcmaster.ca. https://orcid.org/0000-0001-8723-5942
Jonathan Sherbino: McMaster Faculty of Health Sciences Education Research, Innovation and Theory (MERIT) Program, McMaster University, Hamilton, ON, Canada; Division of Emergency Medicine, Department of Medicine, McMaster University, Hamilton, ON, Canada
Jonathan S. Ilgen: Department of Emergency Medicine, University of Washington, Seattle, WA, USA; Center for Leadership & Innovation in Medical Education, University of Washington, Seattle, WA, USA
Emily M. Hayden: Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, USA
Elizabeth Howey: McMaster Faculty of Health Sciences Education Research, Innovation and Theory (MERIT) Program, McMaster University, Hamilton, ON, Canada
Geoff Norman: Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada; McMaster Faculty of Health Sciences Education Research, Innovation and Theory (MERIT) Program, McMaster University, Hamilton, ON, Canada

Introduction

Classic work on diagnostic reasoning has established that clinicians propose an initial diagnostic hypothesis quite rapidly and very early in a patient encounter [1–4]. These early diagnostic hypotheses have been shown to determine final diagnostic accuracy; if the correct diagnostic hypothesis was advanced in the first few minutes of a clinical encounter, clinicians were likely to be correct (95%), whereas if the hypothesis was not advanced in the first few minutes, the likelihood of an eventual accurate diagnosis was much lower (20%) [4]. In another study in a primary care setting, the accuracy of clinicians' "diagnostic hypotheses" was 78% when based only on a review of the chief complaint alone [3]. Subsequent review of the patient history contributed a smaller amount to diagnostic performance (16%), while the physical exam and

diagnostic tests did not influence clinicians' diagnostic accuracy.

This early hypothesis generation process is presumed to be based on prior examples [1]. In the course of becoming an expert, a clinician will encounter hundreds of clinical cases. These encounters may be seen as representative of core distinguishing features; relying on memory for these prior cases may be both an accurate and efficient way to sort between the features of one potential diagnosis relative to another [5, 6]. Even for novices, prior examples can provide scaffolding to understand how current case features relate to prior experiences; given the relatively small subset of cases that novice clinicians have to draw upon; this approach may be less effective in terms of accuracy [7]. Indeed, several studies have shown that prior examples can have both positive and negative effects on diagnostic accuracy [7–12]. While none of these specifically evaluated the influence of expertise, the common sense conclusion from this classic research is that more experienced physicians generate more accurate hypotheses [1]. Yet recent literature [12–19] has unilaterally emphasized the potential for error due to reliance on prior examples. Kahneman described this as the process by which:

> "…people assess the … probability of an event by the ease with which instances or occurrences can be brought to mind. If the availability heuristic is applied, then such factors will affect the … subjective probability of events. Consequently, the use of the availability heuristic leads to systematic biases." [18]

Typical descriptions of this process, often called "availability bias," [12, 13, 15–19] do not specify the contextual factors (e.g., clinical signs or symptoms, patient demographics) that may influence ease of recall. However, results from previous studies of availability bias suggest two mechanisms that may act independently of expertise: the current case may facilitate recall of the clinical reasoning pathway and relevant clinical knowledge of differential diagnoses [19–21], or resemblance between a prior case and the current case—even related to non-clinical features such as hair color—may be sufficient to facilitate recall [8]. Of particular concern, research on the availability bias in domains outside of medicine rarely include expertise as a factor. With few exceptions, studies of cognitive bias outside medicine use materials that are "knowledge or experience sparse" and dependent more on statistical or logical analysis than specialist knowledge [22]. Consequently, they provide no insight into the effect of experience or expertise. Within medicine, a few studies have evaluated how expertise is associated with the influence of cognitive bias [21, 23]. In one study, exposure to prior similar cases was measured through self-report and

the nature of similarity was not controlled [21]. Additionally, participants reviewed very few cases. In the second study, clinical outcomes for 12 cases were presented in a positive versus a negative frame, but the effects of this biasing technique were minimal [23]. Indeed, medical students did not demonstrate bias. More experienced physicians demonstrated a small bias for only 2 cases and responded in idiosyncratic ways, suggesting a stronger influence of unique personal experience [23]. The empirical question remaining is whether the influence of a prior example varies when cases share similar features or dissimilar features, and also if these variations are influenced further by the clinician's level of experience.

This study aimed to measure the influence of a single similar prior case on the diagnostic hypothesis generation process. As it is difficult to control what prior cases physicians have seen or remember, study participants were exposed to identical sets of recent prior cases. A preliminary *exposure phase* created the experience of a prior case. In this exposure phase, participants confirmed a working diagnosis. A second *test phase* measured the influence of the prior case. The similarity of case features in the exposure and test phases was manipulated to evaluate the effect of case similarity on the diagnostic accuracy of clinicians with varying amounts of experience. Competing clinical reasoning theories predict two potential outcomes: 1) if experts rely more on analytical reasoning and discount prior examples, they demonstrate no influence of a prior similar case; 2) if experts rely *more* on similarity to prior exemplars, the influence of a prior similar case may be stronger for experts compared to novices [20, 24].

Research questions

(1) To what extent is diagnosis of a written clinical case influenced by the similar features of recent cases?
   a. How is the influence of recent case exposure related to patient contextual features that are unrelated to the clinical pathology itself?
   b. How does prior case exposure influence performance in clinicians at different levels of experience (student, resident, faculty)?

Materials and methods

Design summary

The study involved clinicians at three levels of expertise—pre-clerkship medical students, emergency medicine residents, and staff

emergency medicine physicians. In essence, we engineered experiences with specific prior cases. We then measured if there was an influence of these prior cases on a set of test cases. Each participant first went through an exposure phase involving a total of eight case vignettes. They then diagnosed 10 test cases. The exposure and test phases created four experimental conditions and one control condition. This study design is summarized in Figure 1.

Materials

Three board certified emergency physicians (EMH, JI, and JS) each with more than 10 years in clinical practice, developed all case materials. All case vignettes followed a standard format describing the chief complaint, past medical history, history of presenting illness, social history and physical examination findings relevant to a particular disease process. Additionally, seven specific and unique patient characteristics where coded for all test cases (See Appendix A, supplementary material, for a description of the case format).

In total 36 cases were created. Case developers first identified reasonable pairs of diagnoses that are often considered together when evaluating a common chief complaint (e.g., cholecystitis and pancreatitis both in the differential diagnosis of upper abdominal pain). The eight diagnostic pairs (referred to hereafter as Diagnosis A and B) are shown in Table 1. For each diagnostic pair, one ambiguous test case representative of both Diagnosis A and B was created (referred to hereafter as *50/50 cases*). Additionally, three prior exposure cases were created with carefully manipulated similarity to the 50/50 cases: similar patient context representative of Diagnosis A (referred to hereafter as *condition a*), similar patient context representative of Diagnosis B (referred to hereafter as *condition b*) and dissimilar patient context representative of Diagnosis A (referred to hereafter as *condition c*). Conditions a–c are explained in more detail below. Finally, four control filler cases were created: adrenal insufficiency, anaphylaxis, glyburide, and mastoiditis. The purpose of these cases is described below.

Prior exposure case design and validation: Prior exposure cases could be representative of Diagnosis A or B. For example, cases representative of Diagnosis A contained five clinical features consistent with Diagnosis A, and one clinical feature consistent with Diagnosis B. All
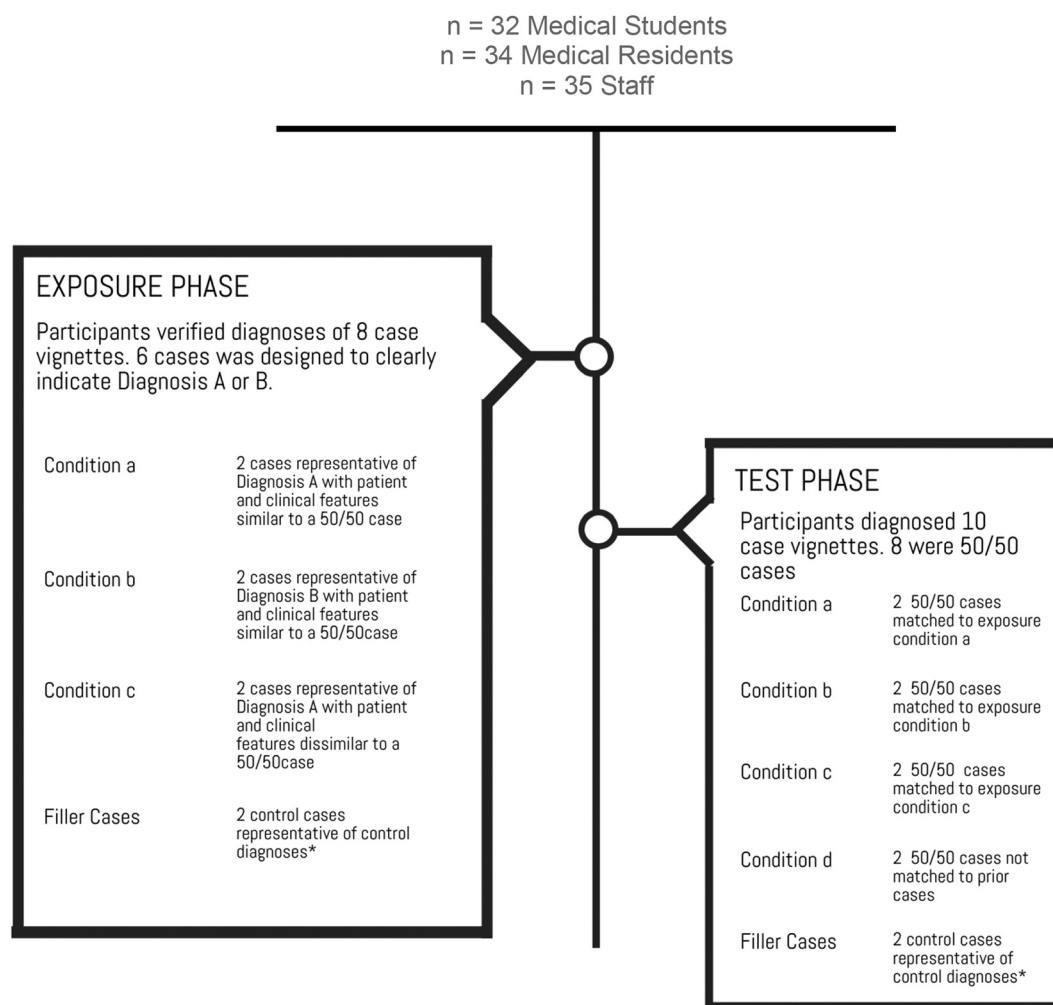

Figure 1: Study flow diagram – all participants completed an exposure phase and a test phase.
*Control case diagnoses: adrenal insufficiency, anaphylaxis, glyburide and mastoiditis.

24 exposure cases were blindly reviewed by 10 experienced emergency physicians not involved with the study. Using a 100-point scale, where zero represented Diagnosis A and 100 represented Diagnosis B, these experts rated the cases. Only exposure cases that had a mean score of less than 10 (related to diagnostic category B) or greater than 90 (related to diagnostic category A) were included.

50/50 case design and validation: Equiprobable, or 50/50 cases, were designed so that there were two diagnoses indicated; Diagnosis A and B (see Table 1). Each 50/50 test case contained five clinical features specific for Diagnosis A and five clinical features specific for Diagnosis B. This strategy ("50/50" cases) has been used in previous studies [20, 24, 25]. The premise is that without any other influence, participants should be equally likely to select Diagnosis A or B. But if there is a bias or undue influence, for example from a previous case of Diagnosis A, then participants should be more likely to select Diagnosis A. These 50/50 cases were blindly reviewed by 10 experienced emergency physicians not involved with the study. Reviewers rated the cases using a 100-point scale, where 0 represented Diagnosis A and 100 represented Diagnosis B. Only test cases that had a mean score of between 40 and 60 for both diagnostic categories were included.

Manipulation of similarity between exposure and 50/50 cases: Similarity to 50/50 cases was either increased or decreased, or not modulated. Case similarity of exposure cases was modulated through the seven patient characteristics and the proportion of diagnosis-specific clinical features. As stated previously, patient characteristics included in exposure cases were intentionally matched or mismatched to those presented in 50/50 test cases. Conditions a–c are described here:

a. **Similar to a case representative of Diagnosis A:** contains similar patient and clinical features. This manipulation was intended to increase similarity between a 50/50 test case and a prior exposure case – contains five Diagnosis A clinical features, one Diagnosis B clinical feature and seven similar patient characteristics

b. [the counterpart to *condition a*] **Similar to a case representative of Diagnosis B:** contains similar patient and clinical features. This manipulation is identical to *condition a*, but for Diagnosis B – contains five Diagnosis B clinical features, one Diagnosis A clinical feature and seven similar patient characteristics. This

condition was introduced to ensure that both diagnoses in each pair had equal influence within our study and there was no unintended systematic bias towards one diagnosis from each pair.

c. **Similar to case representative of Diagnosis A:** but contains *dissimilar* patient and clinical features. This manipulation was intended to decrease similarity between a 50/50 test case and a prior exposure case – contains five Diagnosis A clinical features, one Diagnosis B clinical feature but no similar patient features. This condition was introduced to evaluate the strength of mere exposure to a clinical diagnosis, independent of case feature similarity.

An additional *condition d* was created by including 50/50 cases at test that had no match at the exposure phase. To be clear, this condition did not involve the creation of new cases, but rather the absence of a matching case at exposure.

Participants

The study recruited participants at three levels of expertise: pre-clerkship medical students (n=32), residents (postgraduate year one or 2) in emergency medicine (n=34), and board-certified emergency physicians (n=35). Medical students were recruited from McMaster University. Resident and faculty participants were recruited from three sites: McMaster University, University of Washington, and Harvard University. Participants received a standard monetary honorarium for their participation.

Procedure

All cases were presented on a computer and participants were free to review cases at their own pace. All materials were presented using an open source web-based software (Livecode Ltd).

Prior exposure phase: Participants read a case vignette representative of a diagnosis (e.g., cholecystitis as a typical diagnosis of upper abdominal pain) and were asked to *verify the accuracy of the provided diagnosis*. Accuracy of their verification was not analyzed.

Participants were exposed to a random selection of eight cases (all taken from different pairs); two cases described by *condition a*, two described by *condition b* and two described by *condition c*. These cases were developed within one of the three conditions using block randomization so that all eight cases/diagnostic pairs were matched to *conditions a–d* equally across all participants. For emphasis, the experimental manipulation of similarity was applied to cases in the exposure phase, but the impact of these manipulations were measured later in the test phase. The participants were unaware of the manipulation of the case vignettes in the exposure phase to permit an experimental test of the hypothesized mechanism. An additional two control filler cases were included in this phase to help disguise the study design.

Test phase: Participants then diagnosed 10 test vignettes: six 50/50 cases from *condition a to c*; two 50/50 cases from *condition d*; two control filler cases. Written responses were recorded, scored and analyzed.

Table 1: These eight paired diagnoses were the basis for the test vignettes and the experimental manipulation introducing similarity to a prior case (priming case). Through counterbalancing, participants saw six different prototypical cases in the priming phase that strongly indicated one of the diagnoses.

Diagnosis A	Diagnosis B
Appendicitis	Tuboovarian abscess
Cellulitis	Deep vein thrombosis
Cholecystitis	Pancreatitis
Crystal arthropathy	Septic joint
Myocardial infarction	Aortic dissection
Pneumonia	Pulmonary embolism
Nephrolithiasis	Pyelonephritis
Subarachnoid hemorrhage	Meningitis

Scoring

Written responses for all cases were scored using a rubric created by the case developers. Any scoring ambiguity was resolved by consensus. Control cases had only one correct answer. Responses to test cases were scored as 'correct' if the diagnosis matched the similar prior exposure case. For example, if a participant saw an exposure case of Diagnosis A (e.g., Pneumonia) within *condition a*, then a response of Pneumonia to the similar test case was coded as correct (consistent with the prime). Conversely, a diagnosis of Pulmonary Emolism (PE) would be scored as incorrect (inconsistent with the prime). A high diagnostic accuracy in the presence of a prior exposure case of Pneumonia would then be interpreted as a positive influence of availability bias. Cases in *condition d* were not scored as correct or incorrect but confirmed the validity of the 50/50 case design, by providing a frequency count of indicating Diagnosis A in the absence of a bias.

Ethical approval

Pursuant to the Declaration of Helsinki regarding the ethical conduct of research with human subjects, research ethics approval was received from McMaster University REB #: 11-409;

University of Washington IRB #: 40249 EB; Partners Human Research Committee #: 2016P000605/MGH and consent from all participants was recorded.

Primary analysis

Analyses to explore specific similarity to a single prior case and level of experience: Previous research has demonstrated that exemplar-based reasoning may result in one to one matching with a prior case [7], and this may arise from case features that are similar. Using the scoring method described, *conditions a* and *b* should have equal accuracy. *Condition d* should have 50% accuracy. *Condition c* might be 50% or lower; *condition d* was designed to produce a 50% accuracy while *condition c* was designed to mislead. This hypothesis is tested by examining accuracy with a mixed model ANOVA with Expertise and Condition as between subject factors, and Case as within subject factor. We conducted separate analyses to evaluate the influence of condition on accuracy for Diagnosis A and B.

Results

For the eight 50/50 cases overall, Diagnosis A represented 41%, Diagnosis B 53% and other diagnoses 6% of all responses. The percent by condition for Diagnosis A: a) 36%, b) 36%, c) 44% and d) 47%. Accuracy for filler cases varied: 68% for medical students and 98% for residents and staff.

Specific similarity

As shown in Figure 2, the results examining accuracy assessed by Diagnosis A only did not reveal any influence of specific similarity of patient characteristics ($F_{(3,712)}=7.28$, $p=0.28$). (Similar results are found for Diagnosis B ($F_{(3,712)}=4.87$, $p=0.19$)). Contrary to the proposed hypothesis there was no evidence of a gradient of accuracy. Although *condition d* did produce roughly 50/50 accuracy, this was not unique.

Experience

There was no main effect of expertise when analyzed for Diagnosis A. There was a main effect of Expertise level when analyzed for Diagnosis B, ($F_{(2, 172)}=5.53$, $p=0.03$). There was no Expertise×Condition interaction, for either Diagnosis A, $p>0.8$ or Diagnosis B, $p>0.3$.

Upon inspection, there was no evidence of exposure for residents or staff. The difference between conditions was from 0 to 3% and was not significant ($F_{(3,240)}=1.25$, $p=0.30$ and $F_{(3,248)}=2.04$, $p=0.11$). However, for medical students, any prior exposure to either Diagnosis A or B resulted in a gain of performance from 11 to 20%, ($F_{(3,224)}=3.76$, $p<0.02$). As cases were specifically designed to reference related diagnosis pairs, a secondary analysis was conducted. Although the cases were designed to facilitate recall based on specific similarity, the benefit measured for medical students, of exposure to either Diagnosis A or B, suggested that the cases may have had a category priming effect [26, 27].

The concept of category priming is well established within cognitive psychology and is typically described as the effect of responding faster in a recognition task to an item that was presented recently, (i.e., old) compared to another item (i.e., new) [26–28]. In cognitive psychology, items that are presented within the context of a single experiment are considered 'old', while other items, even though the participants might be familiar with them, are considered 'new';
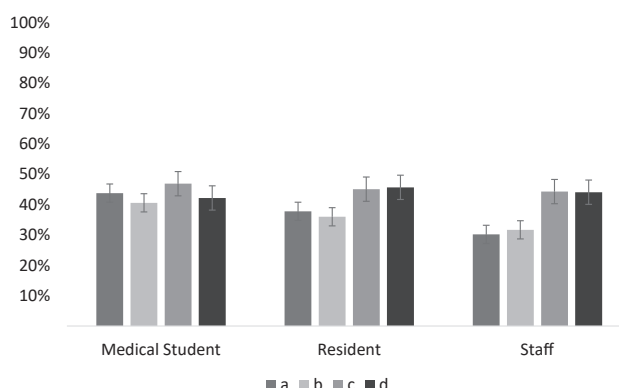

Figure 2: Specific Similarity – Diagnostic accuracy results for *conditions a to d*, determined by examining the influence of condition on Diagnosis A only.
Similar results are seen when analyzed for Diagnosis B. Error bars are standard error for condition.

they are new within the context of the experiment. Despite including items (i.e., word lists or images) that are common to everyday speech or experience, participants consistently respond faster when correctly identifying 'old' items than 'new' ones [26–28]. Explanations of category priming effects rely on the associative memory model [29, 30]. The basic premise of the human associate memory model is that information from the environment, such as images, words or people, can activate, or retrieve, associated information held in memory [30]. The relevance of this theory to the current study became apparent when we examined the results of the medical students; they seemed to gain an advantage for being exposed to one diagnosis (whether Diagnosis A or B) in the *exposure* phase, which may have activated associated knowledge of the other diagnosis [5, 27, 29, 30]. Indeed, the test cases were all designed with clinical features that could be interpreted as Diagnosis A or B. Therefore, we conducted a separate scoring where each response was correct if it was either diagnosis A or B, but incorrect only if it was something else. We then re-analyzed the data to measure the influence of prior exposure on identifying Diagnosis A *or* B.

Post-hoc analysis

Analyses to explore broad category priming

An alternative to a specific retrieval of a particular similar case is a category priming mechanism, where any previous exposure leads to retrieval of knowledge with some relationship to the test diagnosis. Under these circumstances, because of the design of the test cases, responses of either Diagnosis A *or* Diagnosis B would be considered correct, since previous priming would activate knowledge relevant to both diagnoses. This is tested by the difference between test cases in condition a, b, and c versus condition d (no prior similar case), with experience level as a between subjects factor. A mixed model ANOVA was conducted to evaluate the dependent variable of diagnostic accuracy for detecting either Diagnosis A or Diagnosis B and the independent variable of condition (a to d).

Post-hoc results

Broad category priming

Figure 3 shows the analysis by level of expertise and condition a–c, when either Diagnosis A or B is given, as well as *condition d* (unbiased 50/50 cases). There was an effect of experience on diagnostic accuracy; medical students

achieved an accuracy of 85%, medical residents achieved an accuracy of 98%, and faculty achieved an accuracy of 98%, and (F(2,712)=35.6, p<0.0001). There was a main effect of condition, F (3,712)=3.91, p<0.01, with *condition d* (no bias) having a lower accuracy. There was also an interaction between condition and experience level, F(6, 712)=3.55, p<0.01.

Discussion

This study was designed to critically examine the role of prior examples; specifically, to test whether prior examples are beneficial or harmful (availability bias), and whether a prior example affects all levels of expertise equally. In a post-hoc analysis we also examined whether the effect of similarity operates at a specific case feature level or at a broader category priming level.

The study found a small, but significant, main effect of prior examples. On more detailed examination, the effect of a single example on downstream performance was restricted to novice medical students. In this population, the effect was fairly large, ranging from 11 to 20% higher accuracy. Conversely, there was no detectable impact on more experienced clinicians, perhaps in part because there was a ceiling effect with mean accuracy scores approaching 100%. Indeed, for the more experienced physicians, accuracy for filler cases was 98%. In the literature, availability is typically framed as a cognitive bias leading to increased errors. However, this may be an artifact of the experimental conditions. Most of the studies that demonstrate negative effects of availability use priming cases that resemble the test case but are from a different, incorrect,
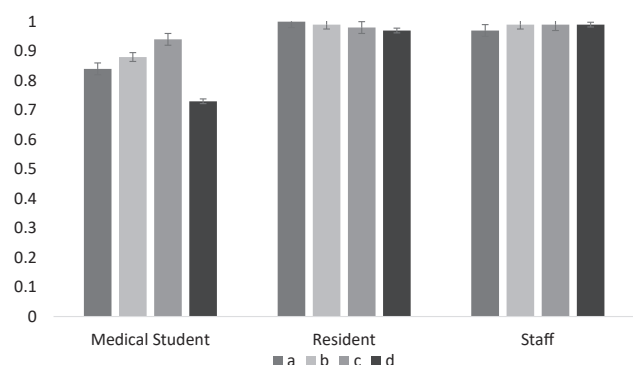

Figure 3: Category priming – diagnostic accuracy results determined by examining the influence of *conditions a to d* on Diagnosis A and B.

The near ceiling effect for staff and residents indicates that Diagnosis A and B were entered with equal frequency and were not influenced by the exposure phase. An influence of the exposure phase is only seen with medical students. Error bars are standard error for condition.

diagnostic category [12]. Under these circumstances, availability leads to error. However in the present study and some others [7] availability contributes to an increase in accuracy.

In prior work, the effect of expertise on cognitive bias is not well studied. Kahneman [18] has claimed that cognitive biases, such as availability, are not amenable to instruction or related to experience:

> *"What can be done about biases? … How can we improve judgments and decisions … ? The short answer is that little can be achieved without a considerable investment of effort. …System 1 is not readily educable…." [18], p. 416.*

However, he does not present any evidence to substantiate this claim. By contrast, some studies of cognitive biases in clinical reasoning have shown that biases diminish with expertise [21, 23]. Indeed, the present study showed no effect of availability (i.e., recent exposure) in the two experienced cohorts. Our findings suggest that the effect of a *single*, recent case diminishes with expertise. As acquisition of expertise in part reflects a larger mental inventory of exemplars, it is reasonable that the effect of any one exemplar, particularly a very recent one, diminishes, which is compatible with an exemplar-based model of reasoning.

While availability was present in novices, it emerged with two critical characteristics: First, there was a significant and substantial *benefit* arising from exposure to a prior similar case. Second, the benefit appeared to be a consequence of category priming [26, 27]. This suggests that the prior experience facilitated retrieval of a broader differential diagnosis.

It is challenging to integrate the current findings with those of prior work. Mamede [12] used a very broad definition of availability, assuming that any diagnosis that may be on the differential diagnosis for the test case may be more available from prior exposure. In that respect, their conclusion is similar to the present study. However Hatala [8] showed specific similarity effects in ECG interpretation derived from matching of contextual features only. Similarly, a number of studies in dermatology [31, 32] have shown specific effects based on similarity in appearance.

One explanation for the variability may be that studies that show effects of specific similarity are based on cognitively-focused tasks; interpretation of ECGs or skin lesions, or maintaining a clinical interview. In a cognitively-focused task, participants may have limited attention for multiple contextual features or limited ability to consider multiple sources of information. Hence, any effect of prior examples may result primarily from matches to a limited set of information. It may be that the manipulations in the present study simplified the task, encouraging broader connections.

Limitations

Residents' and faculty members' accuracy including both diagnoses approached 100%, raising concern for a ceiling effect. This may be a limitation arising from the use of written cases. While some studies [33] have established that written cases are as valid as simulations of clinical reasoning, written cases may have suppressed exemplar effects. Specifically, we recognized that our test cases (with two equiprobable diagnoses) were ambiguous in a very particular way. There is no perceptual ambiguity: the cases are written and designed to be clear. They are truly 50/50 in that the data suggest that the case is equally likely to be Diagnosis A or diagnosis B. Thus, the correct answer is "50% likelihood that it is A and 50% likelihood that it is B". For more experienced physicians, any evidence of Diagnosis A should also lead to a differential that contains Diagnosis B. Rather than creating cases that were objectively too easy, we inadvertently ensured that more experienced physicians would think of Diagnosis A or B.

While the present study was carefully designed to systematically explore the factors that may influence diagnostic accuracy, the extreme care used in carefully creating balanced cases by feature counts may have ultimately encouraged focus on the test case and extinguished the very effect we were attempting to explore.

Conclusions

In this study, novices showed an effect of a previous example on diagnostic accuracy. Moreover, the effect had two specific characteristics. First, it improved accuracy, in contrast to the prevailing view that availability leads to greater errors. Second, the phenomenon arose at the level of category priming, increasing the potential breadth of the differential diagnosis. The diagnoses provided by intermediate and expert clinicians did not appear to be influenced by the experimental manipulation of the availability bias.

References

1. Elstein AS, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. BMJ 2002;324:729–32.
2. Gruppen LD, Palchik NS, Wolf FM, Laing TJ, Oh MS, Davis WK. Medical student use of history and physical information in diagnostic reasoning. Arthritis Care Res (Hoboken) 1993;6: 64–70.
3. Gruppen LD, Woolliscroft JO, Wolf FM, editors. The contribution of different components of the clinical encounter in generating and eliminating diagnostic hypotheses. Conference on research in medical education; 1988.
4. Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning of randomly selected physicians in general medical practice. Clin Invest Med 1982;5:49–56.
5. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. Acad Med 1991;66:S70–2.
6. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. NEJM 2006;355:2217–25.
7. Allen SW, Norman GR, Brooks LR. Effects of prior examples on rule-based diagnostic performance. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans; 1988.
8. Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. JGIM 1999;14: 126–9.
9. Brooks LR, Norman GR, Allen SW. Role of specific similarity in a medical diagnostic task. J Exp Psychol Gen 1991;120:278.
10. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ 2008;42: 468–75.
11. Mamede S, Schmidt HG, Rikers RM, Penaforte JC, Coelho-Filho JM. Influence of perceived difficulty of cases on physicians' diagnostic reasoning. Acad Med 2008;83:1210–6.
12. Mamede S, van Gog T, van den Berge K, Rikers RM, van Saase JL, van Guldener C, et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. J Am Med Assoc 2010;304:1198–203.
13. Croskerry P. Cognitive forcing strategies in clinical decisionmaking. Ann Emerg Med 2003;41:110–20.
14. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. BMJ Qual Saf 2013; 22:ii58–64.
15. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. BMJ Qual Saf 2013;22: ii65–72.
16. Chapman GB, Elstein AS. Cognitive processes and biases in medical decision making. Decision making in health care: theory, psychology, and applications. Cambridge, UK: Cambridge University Press; 2000. p. 183–210.
17. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med 2005;165:1493–9.
18. Kahneman D. Thinking, fast and slow. MacMillan; 2011.
19. Schmidt HG, Mamede S, Van Den Berge K, Van Gog T, Van Saase JL, Rikers RM. Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. Acad Med 2014;89:285–91.
20. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ 2005;39:98–106.
21. Weber EU, Böckenholt U, Hilton DJ, Wallace B. Determinants of diagnostic hypothesis generation: effects of information, base rates, and experience. J Exp Psychol Learn Mem Cognit 1993;19:1151.
22. Lopes LL. Three misleading assumptions in the customary rhetoric of the bias literature. Theor Psychol 1992;2:231–6.
23. Christensen C, Heckerung P, Mackesy-Amiti ME, Bernstein LM, Elstein AS. Pervasiveness of framing effects among physicians and medical students. J Behav Decis Making 1995; 8:169–80.
24. Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. Med Educ 2007;41: 1140–5.
25. Young M, Brooks L, Norman G. Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. Med Educ 2007;41:1146–51.
26. Collins M. Differences in semantic category priming in the left and right cerebral hemispheres under automatic and controlled processing conditions. Neuropsychologia 1999;37:1071–85.
27. Lucas M. Semantic priming without association: a meta–analytic review. Psychon Bull Rev 2000;7:618–30.
28. Voss A, Rothermund K, Gast A, Wentura D. Cognitive processes in associative and categorical priming: a diffusion model analysis. J Exp Psychol Gen 2013;142:536. .
29. Collins AM, Loftus EF. A spreading-activation theory of semantic processing. Psychol Rev 1975;82:407.
30. Anderson JR, Bower GH. Human associative memory. Psychology Press; 2014.
31. Kulatunga-Moruzi C, Brooks LR, Norman GR. Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. Teach Learn Med 2001;13:110–6.
32. Kulatunga-Moruzi C, Brooks LR, Norman GR. Using comprehensive feature lists to bias medical diagnosis. J Exp Psychol Learn Mem Cognit 2004;30:563.
33. Dong T, Saguil A, Artino , Jr. AR, Gilliland WR, Waechter DM, Lopreaito J, et al. Relationship between OSCE scores and other typical medical school performance indicators: a 5-year cohort study. Mil Med 2012;177:44–6.