Role of knowledge and reasoning processes as predictors of resident physicians' susceptibility to anchoring bias in diagnostic reasoning: a randomised controlled experiment

Sílvia Mamede ,[1] Adrienne Zandbergen ,[2] Marco Antonio de Carvalho-Filho ,[3] Goda Choi,[4] Marco Goeijenbier,[5,6] Joost van Ginkel ,[7] Laura Zwaan ,[1] Fred Paas ,[8] Henk G Schmidt [8]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjqs-2023-016621).

For numbered affiliations see end of article.

Correspondence to
Dr Sílvia Mamede;
s.mamede@erasmusmc.nl

Received 8 August 2023
Accepted 26 January 2024
Published Online First
16 February 2024

ABSTRACT

Background Diagnostic errors have been attributed to reasoning flaws caused by cognitive biases. While experiments have shown bias to cause errors, physicians of similar expertise differed in susceptibility to bias. Resisting bias is often said to depend on engaging analytical reasoning, disregarding the influence of knowledge. We examined the role of knowledge and reasoning mode, indicated by diagnosis time and confidence, as predictors of susceptibility to anchoring bias. Anchoring bias occurs when physicians stick to an incorrect diagnosis triggered by early salient distracting features (SDF) despite subsequent conflicting information.

Methods Sixty-eight internal medicine residents from two Dutch university hospitals participated in a two-phase experiment. Phase 1: assessment of knowledge of discriminating features (ie, clinical findings that discriminate between lookalike diseases) for six diseases. Phase 2 (1 week later): diagnosis of six cases of these diseases. Each case had two versions differing exclusively in the presence/absence of SDF. Each participant diagnosed three cases with SDF (SDF+) and three without (SDF−). Participants were randomly allocated to case versions. Based on phase 1 assessment, participants were split into higher knowledge or lower knowledge groups. Main outcome measurements: frequency of diagnoses associated with SDF; time to diagnose; and confidence in diagnosis.

Results While both knowledge groups performed similarly on SDF- cases, higher knowledge physicians succumbed to anchoring bias less frequently than their lower knowledge counterparts on SDF+ cases ($p=0.02$). Overall, physicians spent more time ($p<0.001$) and had lower confidence ($p=0.02$) on SDF+ than SDF− cases ($p<0.001$). However, when diagnosing SDF+ cases, the groups did not differ in time ($p=0.88$) nor in confidence ($p=0.96$).

Conclusions Physicians apparently adopted a more analytical reasoning approach when presented with distracting features, indicated by increased time and lower confidence, trying to combat bias. Yet, extended deliberation alone did not explain the observed performance differences between knowledge groups. Success in mitigating anchoring bias was primarily predicted by knowledge of discriminating features of diagnoses.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Experimental research has shown cognitive biases to cause diagnostic error, but vulnerability to bias differs substantially even among physicians at similar level of training and clinical experience. Predictors of susceptibility to bias have been much debated.

WHAT THIS STUDY ADDS

⇒ Knowledge of features that discriminate between lookalike diseases was the major predictor of anchoring bias in diagnostic reasoning. Engaging in more analytical reasoning in itself was not sufficient to overcome bias, which instead depended on possessing the relevant diagnostic knowledge.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The quest for ways to prevent cognitive bias and diagnostic error, hitherto focused on improving physicians' reasoning processes, should be redirected for the development of effective strategies for learning the critical knowledge to avoid errors throughout undergraduate and graduate training.

Linked

► http://dx.doi.org/10.1136/bmjqs-2024-017141

Check for updates

To cite: Mamede S, Zandbergen A, de Carvalho-Filho MA, *et al*. *BMJ Qual Saf* 2024;**33**:563–572.

BACKGROUND

Five to fifteen per cent of all diagnoses are estimated to be wrong[1 2] often with serious consequences to patients.[3 4] Several strategies have been proposed to reduce diagnostic error but with limited success.[5 6] Most diagnostic errors involve faults in physicians' reasoning,[7 8] but the origin of these faults is unclear. They have often been attributed to the use of heuristics, shortcuts in reasoning used by physicians to make routine judgements.[9 10] Heuristics are typically correct and efficient but can induce biases.[3 11]

Some experimental evidence exists that bias can cause errors. For example, recent experiences with a disease led physicians to confuse a subsequent lookalike (but different) case with the disease seen before in a demonstration of 'availability bias'.[12 13] Bias was also induced by 'salient distracting features' (hereafter, SDF) encountered early in a clinical case.[14 15] SDFs are clinical findings that are irrelevant to the current problem but draw the physicians' attention, because they are strongly associated with a particular disease that seems a plausible diagnosis at first glance. SDF amplified diagnostic mistakes, as physicians often adhered to the (erroneous) diagnosis prompted by these features, failing to revise it when contradictory evidence emerged. This phenomenon, identified as 'anchoring bias' within the medical literature, has been associated with premature conclusion and is frequently cited as a major contributor to cognitive diagnostic errors.[10 16 17]

Noteworthy, while most physicians indeed fell prey to bias in these experimental studies, a substantial fraction of them did not.[12 13] Physicians at similar expertise level (as measured by training and years of clinical experience) differed in their ability to overcome the influence of the bias. Better understanding the sources of these differences may help develop strategies to counteract bias. A prominent view in the medical literature is that variation in resistance to bias is primarily determined by differences in diagnostic reasoning *processes*.[9] Consequently, physicians are taught reasoning strategies to avoiding cognitive biases.[9 18] Conversely, a different perspective assumes that what predicts susceptibility to bias is primarily disease knowledge, that is, knowledge of the associations between each disease and its signs and symptoms as stored in the physician's memory.[19–21] Particularly important is probably knowledge of features that help discriminate between the disease and other usual plausible diagnoses for a patient with a particular clinical presentation. Strong knowledge of these *discriminating features* would make a physician less likely to overlook them when irrelevant cues become salient in bias-inducing circumstances.[21]

This hypothesis is not easily investigated because it requires measurements of *specific* rather than general knowledge as well as indicators of reasoning processes. Nevertheless, it has some preliminary support. A recent study requested physicians to diagnose cases under conditions that induce availability bias and evaluated their knowledge of discriminating features.[22] Lower knowledge physicians succumbed to the availability bias considerably more frequently than their higher knowledge counterparts. This observation cannot be explained by increased engagement in analytical reasoning (which would necessitate more time), given that both groups dedicated similar timeframes to diagnosis. The level of knowledge apparently predicted susceptibility to bias. These findings, however, are preliminary. The knowledge evaluation task, conducted immediately after the diagnostic task, may have been influenced by it, and time was the sole metric of reasoning processes. Moreover, the study investigated availability bias, and other types of biases are also deemed relevant.

The present study examined the role of knowledge of discriminating features in counteracting anchoring bias induced by SDF. Disease knowledge was evaluated 1 week *before* the physicians diagnosed clinical cases with and without SDF that have been shown to induce anchoring bias.[14 15] Time spent in diagnosis and ratings of confidence in the diagnosis were obtained as indicators of reasoning mode. We hypothesised that knowledge of discriminating features would be the primary predictor of susceptibility to anchoring bias. Therefore, physicians who had more knowledge of discriminating features were expected to make fewer diagnostic errors related to anchoring bias without differences in the measures of reasoning mode.

METHODS

Study design

The study was a two-phase experiment consisting of: (1) evaluation of the physicians' knowledge about discriminating features of the diseases tested in the study, and (2) (1 week later) diagnosis of clinical cases of these diseases. The cases were manipulated so that SDFs were either present (SDF+ cases) or absent (SDF− cases). Two case sets were prepared alternating the manipulation across cases. Participants were randomly assigned to one of the sets in a counterbalanced within-subjects incomplete block design. We measured time required to diagnose and confidence in the diagnosis. In psychological research on reasoning, moving from intuitive towards more analytical reasoning is associated with increased response time[23–25] and decreased confidence.[26 27] Figure 1 presents the study design.

The study protocol was pre-registered on the Open Science Framework (online supplemental file 1). We followed the Consolidated Standards of Reporting Trials reporting guidelines as far as they applied to the present experiment.

Setting and participants

Participants were residents in internal medicine (In the country, residents in internal medicine are physicians

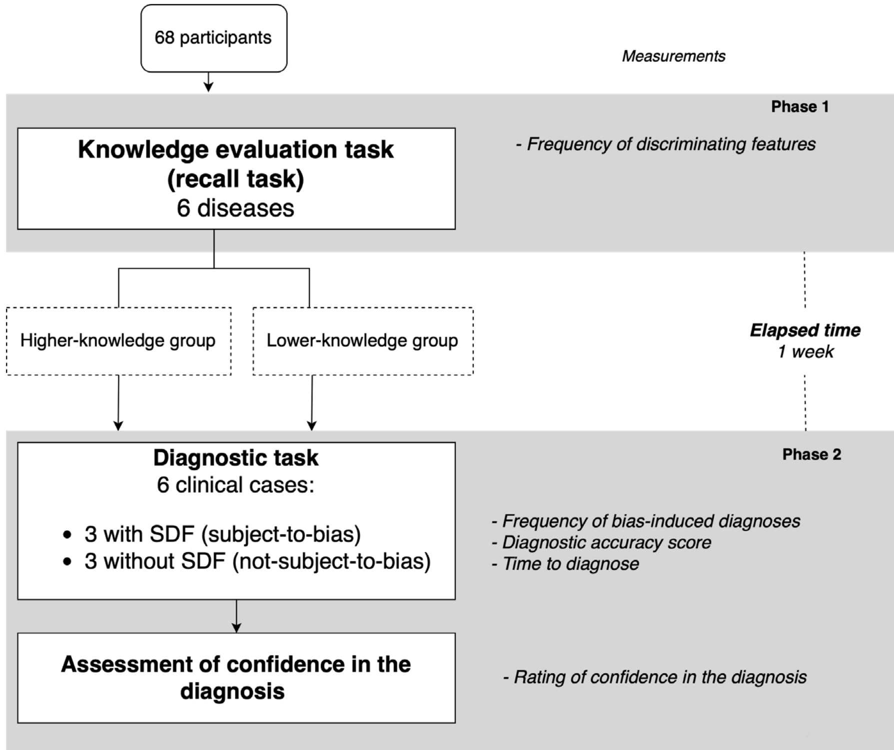

Figure 1 Diagram of the study design. SDF, salient distracting features.

in training to become internists after having graduated from medical schools which have a 6-year curriculum) at Erasmus Medical Center (Erasmus MC) and at the University Medical Centre Groningen in the Netherlands. Residents with at least 1 year of clinical experience were invited by the programme director to volunteer for the study. A €50/participant donation was made to the residents' association to organise extracurricular scientific activities. The study took place during regular educational days in 2021–2022, online at Erasmus MC and face to face in Groningen. (Because patterns of performance were similar regardless of session format, participants were aggregated for the analysis.) Written consent was obtained from all physicians.

Sample size estimation

A priori power analysis was performed, assuming a medium effect size (Cohen's f=0.30), as suggested by previous research,[15] and the standard alpha level of 0.05. It indicated that a sample size of 62 participants would be sufficient to achieve a power of 0.80.

Materials and procedure

In phase 1, the physicians performed a recall task to assess knowledge of findings associated with each of the six diseases included in the diagnostic task (phase 2). A recall task assesses mental representations of disease knowledge by measuring how much about it participants are able to recollect from memory under restricted time conditions.[28] The six tested diseases (see online supplemental file 2) were mixed with three unrelated diseases and presented one by one in random order on a computer screen. For each disease, the physician typed everything (s)he could remember about clinical findings (In a broad sense, including findings referring to medical history, complaints, results of physical examination and diagnostic tests) that were critical for the diagnosis of that particular disease and helped distinguish between the disease and other usual diagnoses. Based on a pilot with participants not involved in the study, a maximum time of 2.5 min was allocated for each disease.

In phase 2, six clinical cases were diagnosed (see online supplemental file 2). The cases were developed by board-certified internists and used in prior studies with internal medicine residents.[12 15] None of the cases involved real patients' information. Each case was prepared in two versions which differed exclusively on the presence or absence of SDF (see table 1 for an example) but had the same most likely diagnosis. This manipulation has been used in previous studies.[15]

Two sets of cases were prepared, alternating the manipulation across the cases. The cases were presented one by one in random order on the Qualtrics research suite (Qualtrics, Provo, Utah), which automatically randomly allocated each participant to one of the case sets and registered responses and response time. The participant was asked to read the case and type the most likely diagnosis as fast as possible without

Table 1 Example of a clinical case (diagnosis: vitamin B_{12} deficiency) used in the study with the version with and without salient distracting features (SDF)

Version without SDF	Version with SDF
A 62-year-old man with uncomplicated diabetes in his otherwise unremarkable medical history was brought to the clinic by his son. The patient has been referred to your clinic because of increasing changes in mental state associated with emotional lability for 4 months. He also had problems with walking and reduced vision during this period. The patient reports sexual contacts only with his wife. Medication: NPH insulin. Health-related behaviours: social alcohol use, not drugs. Physical examination: Patient is in a wheelchair, cries often and does not understand much. BP 140/90 mm Hg; pulse 88/min; temperature 36.2°C. No signs of dehydration. Neurological examination: Patient walks slowly and with long strides; ataxia; difficulties keeping balance with open and closed eyes. Muscle power and reflexes are normal; Babinsky negative. Fundoscopic examination: no abnormalities. Further physical examination: no abnormalities. Diagnostic tests: Hb 5.6 mmol/L; Ht 32.9%; MCV 112 µm³; macrocytosis; leucocytes 5.2×10^9/L with normal differentiation and hypersegmentation of the polymorphonuclear granulocytes; thrombocytes 154×10^9/L; reticulocytes <1%; glucose 7.27 mmol/L; urea 9.3 mmol/L; creatinine 79.5 µmol/L; electrolytes, liver functions and blood gases: normal; TSH 5 mU/L. Antibodies to syphilis and HIV: negative. Cerebrospinal fluid analysis: no abnormalities. CT scan and MRI skull: no abnormalities.	A 62-year-old man with uncomplicated diabetes in his otherwise unremarkable medical history was brought to the clinic by his son who reports that 4 months ago, before the onset of the present complaints, his father <u>fell from his own height</u>, without loss of conscience. His son was very worried because <u>his grandfather died of dementia at the age of 67</u>. The patient has been referred to your clinic because of increasing changes in mental state associated with emotional lability for 4 months. He also had problems with walking and a reduced vision during this period. The patient reports sexual contacts only with his wife. Medication: NPH insulin. Health-related behaviours: social alcohol use, not drugs. Physical examination: Patient is in a wheelchair, cries often and does not understand much. BP 140/90 mm Hg; pulse 88/min; temperature 36.2°C. No signs of dehydration. Neurological examination: Patient walks slowly and with long strides; ataxia; difficulties keeping balance with open and closed eyes. Muscle power and reflexes are normal; Babinsky negative. Fundoscopic examination: no abnormalities. Further physical examination: no abnormalities. Diagnostic tests: Hb 5.6 mmol/L; Ht 32.9%; MCV 112 µm³; macrocytosis; leucocytes 5.2×10^9/L with normal differentiation and hypersegmentation of the polymorphonuclear granulocytes; thrombocytes 154×10^9/L; reticulocytes <1%; glucose 7.27 mmol/L; urea 9.3 mmol/L; creatinine 79.5 µmol/L; electrolytes, liver functions and blood gases: normal; TSH 5 mU/L. Antibodies to syphilis and HIV: negative. Cerebrospinal fluid analysis: no abnormalities. CT scan and MRI skull: no abnormalities

SDFs are underlined.
BP, blood pressure; Hb, haemoglobin; Ht, haematocrit; MCV, Mean Corpuscular Volume; NPH, Neutral Protamine Hagedorn; TSH, Thyroid Stimulating Hormone.

compromising accuracy.[12 13] Each participant diagnosed three SDF+ cases and SDF− cases, but which case was diagnosed with and without SDF varied for each participant depending on the case set to which s/he had been assigned. After diagnosing all cases, each case was presented again, with only its initial sentence and the diagnosis given by the participant, who was asked to mark, on a 0–100% scale, how confident s/he was that the diagnosis was correct. Subsequently, the participants provided demographic information, answered probing questions about the study and received feedback on the correct diagnoses. (Online supplemental file 2 provides additional information.)

Outcome measurements

Knowledge was measured (phase 1) by the frequency of discriminating features mentioned by the participant in the recall task. Four board-certified internists (AZ, GC, MAdC-F, MG) worked independently to assign each clinical finding mentioned by the participants for each disease to one of three categories: 'discriminating feature', 'correct finding but not critical for the diagnosis' or 'incorrect finding'. The internists formed triads, and the most frequent category attributed to each finding was used. For each participant, we summed up the number of discriminating features mentioned, and descriptive statistics were computed. Similar to what happens when students' outcomes in tests are based on relative standard setting methods,[29] participants were assigned to either a lower knowledge or a higher knowledge group based on the median. (See online supplemental file 2 for further details.)

Our main outcomes of interest were: frequency of incorrectly giving the diagnosis associated with the SDF as the most likely diagnosis; diagnostic accuracy score; time to diagnose the case; and rating of confidence in the diagnosis. (Note that while the first outcome measures only errors *linked* to the SDF, the second outcome refers to any type of error.) These measurements were computed for SDF+ cases and for SDF− cases. First, the diagnoses given by the participants were categorised as either correct (accuracy score 1.0); partially correct (accuracy score 0.5); incorrect—associated with the SDF (accuracy score 0); and incorrect—not associated with the SDF (accuracy score 0). Triads of internists categorised the diagnoses, following a procedure similar to the one described for the knowledge evaluation task. The mean frequency of incorrect diagnoses associated with the SDF on SDF+ and SDF− cases was computed. Notice that comparing these two types of cases is critical to determine whether the incorrect diagnosis associated with the SDF can actually be attributed to bias. For example, dementia would be a plausible diagnosis in a patient with the clinical presentation displayed in table 1 even in the absence of the SDF inserted in one of the versions. Therefore, only if the frequency of dementia among the diagnoses *increased* when the SDF were present, anchoring bias was considered to have occurred.

Missing data

Several variables in the data had missing values. Missing data were handled using multiple imputation,[30] creating 100 imputed datasets, by using the mice package in R.[31] The dataset roughly contained three types of variables: numerical variables, numerical variables that were computed using other variables and categorical variables. Missing data on numerical variables were imputed using predictive mean matching,[32] numerical variables computed from other variables were imputed using passive imputation[33] and categorical variables were imputed using a (multinomial) logistic regression model. Both outcomes and predictors were imputed. In total, the data contained 80 variables, all of which were used in the imputation process, either to be imputed, to serve as predictors for the missing values on other variables or both. Detailed information on the choices and procedures for the imputation are provided in online supplemental file 2.

Statistical analysis

Separate mixed analyses of variance with knowledge group (lower knowledge vs higher knowledge) as between-subjects factor and exposure to SDF (SDF+ and SDF−) as within-subjects factor were performed on the mean frequency of incorrect diagnoses associated with the SDF, mean diagnostic accuracy scores, mean time spent in diagnosis and mean confidence in the diagnosis. These analyses were performed both on the multiple imputation dataset and on the observed data (complete case analysis). Online supplemental file 2 provides additional information on the statistical analysis and its rationale.

RESULTS

Seventy-five physicians attended the two phases of the study. Seven were removed from the analysis (five denied consent to use their data; two resumed the tasks repeatedly due to interruptions), resulting in 68 participants.

The complete case analysis is presented in online supplemental file 2. The number of years in clinical practice did not differ between lower knowledge and higher knowledge groups (mean (SD), respectively: 2.71 (1.21); 2.70 (1.39); p=0.98). The complete case analysis yielded results similar to the analyses with imputation for missing data, which are presented below. (In all figures error bars represent 95% CIs.)

The frequency with which the diagnosis associated with the SDF was incorrectly given as the most likely diagnosis in SDF+ cases and SDF− cases is presented in figure 2. Overall, as expected, the error occurred more frequently when the SDFs were present in the case (ie, the case was subject to bias) than when they were absent (p<0.001), and physicians with higher knowledge made fewer errors than physicians with lower knowledge (p=0.015). However, a significant interaction effect (p=0.020) was found, with the difference in performance showing only on SDF+ cases. While the two groups performed similarly in the absence of SDF, higher knowledge physicians fell to anchoring bias less frequently when SDFs were present than their colleagues with lower knowledge.

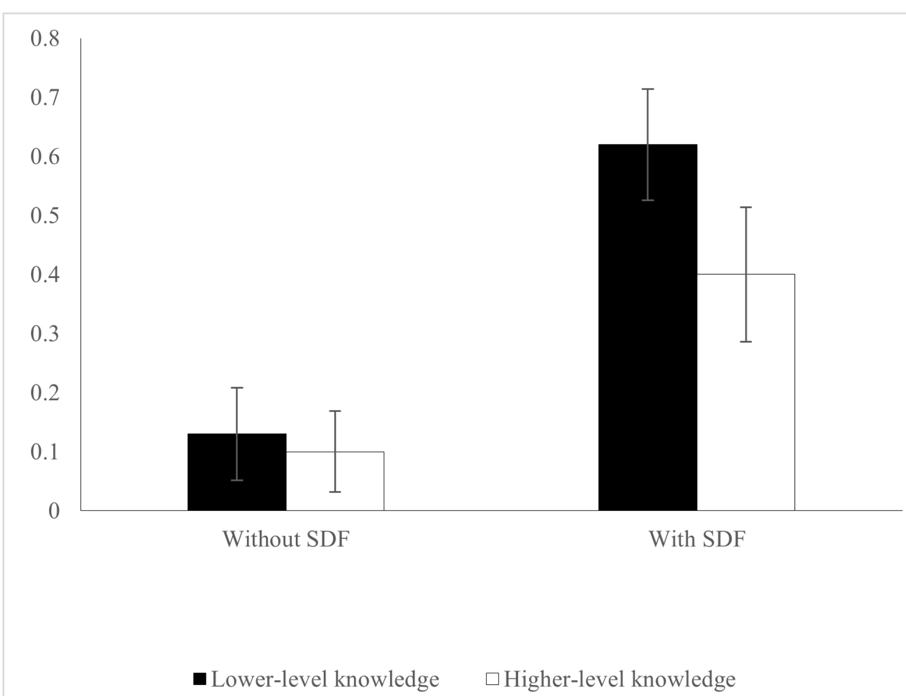

Figure 2 Incorrect diagnoses associated with salient distracting features (SDF) on cases with SDF and without SDF (mean frequency; range 0–1).

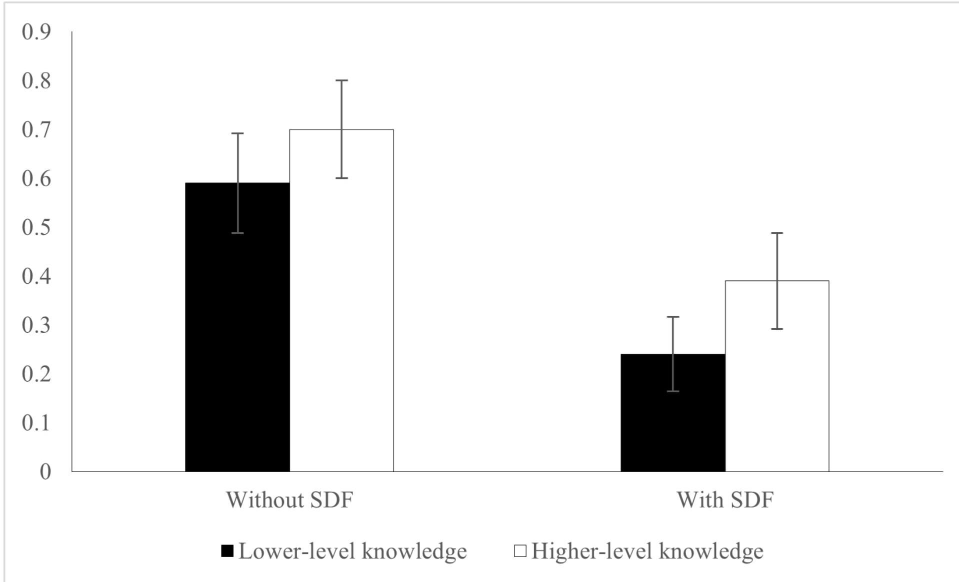

Figure 3 Diagnostic accuracy on cases with salient distracting features (SDF) and without SDF (mean diagnostic accuracy score; range 0–1).

Figure 3 presents the mean diagnostic accuracy scores when cases were subject to bias and not. Overall, higher knowledge physicians showed significantly higher accuracy relative to lower knowledge physicians (p=0.014), and the presence of SDF significantly decreased accuracy (p<0.001). Contrary to our expectation, the two groups did not significantly differ in how their diagnostic accuracy scores were hindered by the presence of SDF. Though the decrease in accuracy was lower in the higher knowledge than in the lower knowledge the interaction effect was not significant (p=0.667).

Figure 4 presents the indicators of reasoning mode—time spent in diagnosing and confidence in the diagnosis—in SDF+ and SDF− cases. Both knowledge groups invested more time to diagnose cases when SDFs were present than when they were not (p<0.001), but there was no overall significant difference between the two groups (p=0.334) and no interaction effect (p=0.138). Regarding rates of confidence, overall, the two knowledge groups did not significantly differ in their confidence ratings (p=0.087), and the presence of SDF decreased confidence (p=0.002). However, a significant interaction effect was observed (p=0.008), because while the confidence of lower knowledge physicians remained basically the same when cases were subject to bias and not (p=0.764), higher knowledge physicians' confidence decreased in the presence of SDF (p<0.001).

DISCUSSION

In this experimental study, SDF encountered early in clinical cases induced anchoring bias, contributing to diagnostic errors. Among physicians from the same training programme and with similar years of clinical experience, susceptibility to bias was predicted by knowledge of critical diagnostic features that differentiated between alternative diagnoses for the clinical presentations displayed in the cases. When the cases contained SDF, thereby being subject to anchoring bias, physicians with less knowledge of discriminating features gave the bias-induced diagnosis considerably

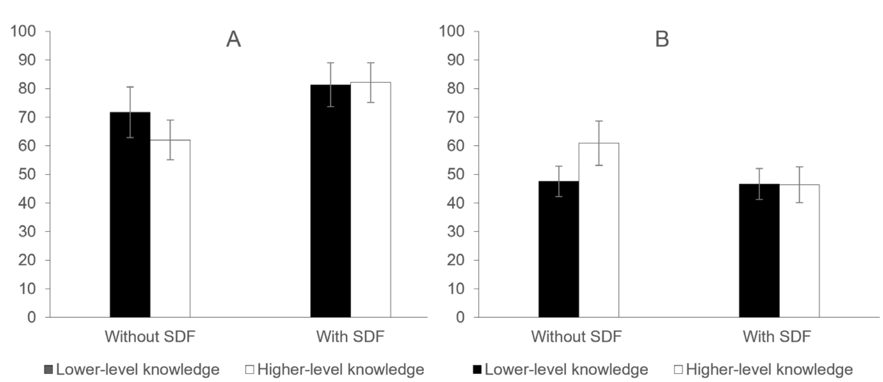

Figure 4 Indicators of reasoning mode: (A) time spent in diagnosis (mean time in seconds) and (B) confidence in the diagnosis (mean confidence; range 0–100) on cases with salient distracting features (SDF) and without SDF.

Mamede S, *et al*. *BMJ Qual Saf* 2024;**33**:563–572. doi:10.1136/bmjqs-2023-016621

more frequently than their counterparts who had more comprehensive knowledge of these specific features. Overall, physicians needed more time and had lower confidence in cases with SDF compared with the same cases without SDF. However, the higher knowledge and lower knowledge groups did not differ in time or confidence when diagnosing the SDF+ cases. In psychological research on bias, more time to respond[23–25] and lower confidence[26 34] have been associated with more engagement in analytical reasoning. Therefore, our findings suggest that physicians from both groups moved towards more analytical reasoning when the cases were subject to bias. Because the two groups did not differ in these measures, reasoning mode per se cannot explain the difference in susceptibility bias. Instead, susceptibility was predicted by differences in knowledge of discriminating features.

Anchoring bias has been often pointed in the medical literature as a common threat to diagnostic reasoning.[10 35] It appears frequently in retrospective investigations of actual diagnostic errors.[7 8 36] Experienced clinicians often attribute their own diagnostic errors to sticking to an early hypothesis triggered by salient information even after disproving information becomes available.[16] These studies suggest that repairing an incorrect initial hypothesis, though possible, is difficult. A hypothesis triggered by salient information comes to mind quickly and seems reasonable. After being generated, the hypothesis influences subsequent reasoning, potentially hindering gathering and/or correctly interpreting critical diagnostic information.[37 38] In the present study, few modest SDFs were added to the cases, but even this little information was enough to distract the physicians.

An appealing (and widely discussed) idea is that clinicians should learn to detect when they are at risk of being biased, when they should stop and think further.[39 40] Based on this idea, courses on clinical reasoning and cognitive debiasing strategies have been tried.[18] Results so far do not seem promising, at least when error reduction was evaluated.[20 41 42] An obstacle may be that detecting an error requires a sort of 'diagnosis' of one's own diagnostic reasoning, which depends on cognitive processes that are as fallible as the ones involved in the clinical diagnosis. Another difficulty is that engaging in analytical reasoning per se does not guarantee that the bias-induced response is over-ridden.[43] There may always be physicians who fail to acknowledge that a problem requires further deliberation, proceeding instead with their intuitive (and potentially biased) response. However, in our study, the overall increase in diagnosis time on the SDF+ cases suggests that most physicians somehow recognised that the problem required further consideration. Indeed, psychological research shows that problems containing salient potentially bias-triggering information tend to lead people to think further before responding.[43–45] Even those who end up giving the biased response seem to engage in more analytical reasoning. Nevertheless, analytical reasoning can focus either on 'justifying' the initial response by searching for evidence that ends up strengthening it or on 'overriding' the initial response by reconsidering other findings and generating alternatives. The latter is more likely to arrive at the correct response.[43–45] Physicians with stronger knowledge of discriminating features may be better equipped for the 'overriding-format' of analytical reasoning because they are more likely to recognise these features as relevant, which may raise alternative hypotheses and restructure initial reasoning. Less knowledgeable physicians who are not aware, for example, that hypersegmentation of the polymorphonuclear granulocytes would strongly support vitamin B_{12} deficiency in the patient described in table 1 could think about the case for a long time but still overlook these key features. Though apparently reasonable, these assumptions are conjectures demanding investigation.

Despite the seemly obvious importance of knowledge to improve diagnosis, whether knowledge helps counteract bias is far from clear. In psychological research, studies comparing experts and novices' susceptibility to bias have led to conflicting results.[46–48] Furthermore, earlier experiments with physicians showed that difficulty in restructuring initial clinical reasoning increased with experience, making more experienced (most often older) physicians more vulnerable to anchoring bias.[49–51] Noteworthy, these studies considered time in professional practice but did not assess physicians' knowledge. However, even among physicians with similar training and years in clinical practice, differences in disease knowledge are unavoidable. If these differences are examined, as our study did, we may have a better picture of the role of knowledge in bias.

Educational strategies to reduce clinicians' susceptibility to bias have hitherto focused on improving physicians' reasoning *processes*, teaching about reasoning and biases. The role of knowledge in counteracting bias has been largely neglected.[19] However, in a previous study we showed an intervention which enhanced relevant disease knowledge to reduce physicians' vulnerability to bias in future cases.[21] Consistently, a recent study associated deficiencies in diagnostic knowledge with adverse outcomes of primary care visits.[52] Our findings reinforce the need to redirect our efforts to develop effective educational strategies for strengthening knowledge of discriminating features.

Intriguingly, the different susceptibility to bias in the two knowledge groups did not result in significantly different diagnostic accuracy. With regard to the frequency of the diagnosis associated with the SDF, the two knowledge groups did not significantly differ in how often (scarcely, as expected) they mentioned this diagnosis on SDF− cases. However, when the presence of SDF made the cases subject to bias, the lower knowledge physicians incorrectly gave the bias-induced diagnosis considerably more frequently than the higher knowledge physicians. The results for diagnostic accuracy seem to show a similar pattern. The difference in diagnostic accuracy scores between the two knowledge groups was lower in SDF− cases than when the cases were subject to bias. Nevertheless, the interaction effect was not significant. The large variation within groups may have contributed to that. Furthermore, we only measured whether the physicians *possessed* the critical diagnostic knowledge but not if this knowledge was actually *used* during the diagnosis, which increases the noise. There may also be other reasons. The presence of SDF made the cases more ambiguous and, whereas the higher knowledge physicians could more easily exclude the bias-induced diagnosis than the lower knowledge physicians, higher knowledge physicians made other mistakes. Indeed, while the bias-induced diagnosis accounted for around 81% of the incorrect diagnoses given by the lower knowledge group, this proportion dropped to 65% in the higher knowledge group. Future research should examine this phenomenon.

The findings referring to confidence in the diagnosis are relevant to the current debate on diagnostic calibration. The two groups did not differ in confidence in SDF+ cases. Nonetheless, while the higher knowledge group reported lower confidence in SDF+ than in SDF− cases, the confidence reported by the lower knowledge group did not decrease when the cases were subject to bias. Knowledge seems to play a key role in susceptibility to bias and in diagnostic calibration.

This study has several limitations, some of them informing future research. We measured time and confidence, which though common in psychological research on bias[24 26] are *indirect* indicators of reasoning. Research on the knowledge–bias interaction has just started, and these easy-to-obtain measurements sufficed as a first step. Future research should examine what takes place when participants think further about the case to determine features of analytical reasoning that help counteract bias. On a related issue (and limitation), we measured knowledge of discriminating features that participants *possessed* but not the extent to which they actually *used* this knowledge while diagnosing the cases. This requires research with other methodological tools. Another limitation is that knowledge level was classified based on performance aggregating all phase 1 diseases, which is not ideal considering content specificity. Nevertheless, case-level knowledge was highly correlated across all diseases (online supplemental file 2), suggesting that aggregated performance sufficiently captured knowledge differences. The use of written clinical cases leaves out important steps of the authentic clinical process. Noteworthy, an effect found with cases that provide physicians with all relevant information would possibly be larger rather than smaller when physicians have to gather this information themselves. The influence of an incorrect initial hypothesis, for example, hinders gathering of critical information,[38] which would increase the risk of anchoring bias. Although accessing other resources is possible in real settings, physicians often fail to recognise this need.[53] Moreover, written cases allow for control and have proved an acceptable proxy for group differences in performance in practice.[54] The limited clinical experience of our participants makes it unclear whether the findings apply to more experienced physicians. Previous studies have shown that more experienced physicians tend to be more rather than less susceptible to bias; however, this susceptibility might also depend on specific clinical knowledge in this group.

In conclusion, in this study incorrect diagnoses induced by anchoring bias were more frequent among physicians with less knowledge of features that discriminate between lookalike diseases relative to more knowledgeable physicians. Regardless of their knowledge level, the physicians spent more time and reported lower confidence in the diagnosis of cases that contained SDF than in those without SDF. This suggests that the physicians realised, even if intuitively, that the case required further thought when it was subject to bias and moved towards more analytical reasoning. The lower susceptibility to bias cannot therefore be explained by differences in the degree of engagement in analytical reasoning. Differences in knowledge of discriminating features seem to have predicted it. These findings may help place refinement of diagnostic knowledge in the core of strategies to make clinicians less vulnerable to bias.

Author affiliations
[1]Institute of Medical Education Research Rotterdam, Erasmus Medical Center, Rotterdam, The Netherlands
[2]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
[3]Wenckebach Institute (WIOO), University Medical Centre Groningen, Groningen, The Netherlands
[4]Department of Hematology, University Medical Centre Groningen, Groningen, The Netherlands
[5]Department of Intensive Care, Spaarne Gasthuis, Haarlem, The Netherlands
[6]Department of Intensive Care, Erasmus MC, Rotterdam, The Netherlands
[7]Department of Psychology, Methodology and Statistics, Leiden University, Leiden, The Netherlands
[8]Department of Psychology, Education and Child Studies, Erasmus Universiteit Rotterdam, Rotterdam, The Netherlands

Acknowledgements The authors are grateful to the residents who dedicated their scarce time to participate in the study and to Liona Ionescu for her help to format the manuscript.

Contributors All authors had full access to all the study data and take responsibility for the integrity of the data and the accuracy of the data analysis. Study conception and design: SM, HGS. Development of study materials: SM, AZ, MAdC-F. Acquisition of data: SM, AZ, MAdC-F, MG, GC, LZ. Statistical analysis: SM, HGS, JvG. Analysis or interpretation of data: all authors. Drafting of the manuscript: SM. Critical revision of the manuscript for important intellectual content: all authors. Administrative, technical or material support: SM, AZ, GC. Supervision: SM, HGS. Guarantor: SM.

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

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Ethics Review Committee of the Department of Psychology, Education and Child Studies (reference number: 20-026 ErC DPECS), Erasmus University Rotterdam. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are available upon reasonable request and subject to institutional regulations.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Sílvia Mamede http://orcid.org/0000-0003-1187-2392
Adrienne Zandbergen http://orcid.org/0000-0001-5056-5921
Marco Antonio de Carvalho-Filho http://orcid.org/0000-0001-7008-4092
Joost van Ginkel http://orcid.org/0000-0002-4137-0943
Laura Zwaan http://orcid.org/0000-0003-3940-1699
Fred Paas http://orcid.org/0000-0002-1647-5305
Henk G Schmidt http://orcid.org/0000-0001-8706-0978

REFERENCES

1 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med* 2008;121(5 Suppl):S2–23.
2 Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf* 2014;23:727–31.
3 Balogh EP, Miller BT, Ball JR, *et al*. Improving diagnosis in health care. In: *Committee on Diagnostic Error in Health Care. Improving diagnosis in health care*. Washington, D.C: The National Academies Press, 29 December 2015.
4 Wallace E, Lowry J, Smith SM, *et al*. The epidemiology of malpractice claims in primary care: a systematic review. *BMJ Open* 2013;3:e002929.
5 Dave N, Bui S, Morgan C, *et al*. Interventions targeted at reducing diagnostic error: systematic review. *BMJ Qual Saf* 2022;31:297–307.
6 Prakash S, Sladek RM, Schuwirth L. Interventions to improve diagnostic decision making: a systematic review and meta-analysis on reflective strategies. *Medical Teacher* 2019;41:517–24.
7 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
8 Kachalia A, Gandhi TK, Puopolo AL, *et al*. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. *Ann Emerg Med* 2007;49:196–205.
9 Croskerry P. From mindless to mindful practice — cognitive bias and clinical decision making. *N Engl J Med* 2013;368:2445–8.
10 Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115–20.
11 Evans Js. The heuristic-analytic theory of reasoning: extension and evaluation. *Psychon Bull Rev* 2006;13:378–95.
12 Mamede S, van Gog T, van den Berge K, *et al*. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;304:1198–203.
13 Schmidt HG, Mamede S, van den Berge K, *et al*. Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. *Acad Med* 2014;89:285–91.
14 Mamede S, Splinter TAW, van Gog T, *et al*. Exploring the role of salient distracting clinical features in the emergence of diagnostic errors and the mechanisms through which reflection counteracts mistakes. *BMJ Qual Saf* 2012;21:295–300.
15 Mamede S, van Gog T, van den Berge K, *et al*. Why do doctors make mistakes? A study of the role of salient distracting clinical features. *Academic Medicine* 2014;89:114–20.
16 Balla J, Heneghan C, Goyder C, *et al*. Identifying early warning signs for diagnostic errors in primary care: a qualitative study. *BMJ Open* 2012;2:e001539.
17 Restrepo D, Armstrong KA, Metlay JP. *Annals* clinical decision making: avoiding cognitive errors in clinical decision making. *Ann Intern Med* 2020;172:747–51.
18 Reilly JB, Ogdie AR, Von Feldt JM, *et al*. Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ Qual Saf* 2013;22:1044–50.
19 Dhaliwal G. Premature closure? Not so fast. *BMJ Qual Saf* 2017;26:87–9.
20 Norman GR, Monteiro SD, Sherbino J, *et al*. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine* 2017;92:23–30.
21 Mamede S, de Carvalho-Filho MA, de Faria RMD, *et al*. 'Immunising' physicians against availability bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Qual Saf* 2020;29:550–9.
22 Mamede S, Goeijenbier M, Schuit SCE, *et al*. Specific disease knowledge as predictor of susceptibility to availability bias in diagnostic reasoning: a randomized controlled experiment. *J Gen Intern Med* 2021;36:640–6.
23 Bonner C, Newell BR. In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition* 2010;38:186–96.
24 De Neys W, Glumicic T. Conflict monitoring in dual process theories of thinking. *Cognition* 2008;106:1248–99.
25 Rotgans JI, Schmidt HG, Rosby LV, *et al*. Evidence supporting dual-process theory of medical diagnosis: a functional near-infrared spectroscopy study. *Med Educ* 2019;53:143–52.
26 Gangemi A, Bourgeois-Gironde S, Mancini F. Feelings of error in reasoning—in search of a phenomenon. *Thinking & Reasoning* 2015;21:383–96.
27 Vartanian O, Beatty EL, Smith I, *et al*. The reflective mind: examining individual differences in susceptibility to base rate neglect with fMRI. *J Cogn Neurosci* 2018;30:1011–22.
28 Westerman DL, Payne DG. Research methods in human memory. In: Davis SF, ed. *Handbook of research methods in experimental psychology*. Wiley-Blackwell, 2005: 346–65.

29 Norcini J, Guille R. Combining tests and setting standards. In: Van der Norman GR, Vleuten CPM, Newble DI, eds. *International handbook of research in medical education*. Dordrecht: Kluwer Academic Publishers, 2002.
30 Rubin DB. Multiple imputation for nonresponse in surveys. New York Wiley; 1987.
31 Van Buuren S, Groothuis-Oudshoorn CGM. mice: Multivariate imputation by chained equations in R. *J Stat Softw* 2011;45:1–67.
32 Rubin DB. Statistical matching using file concatenation with adjusted weights and multiple imputations. *J Bus Econ Stat* 1986;4:87–94.
33 van Buuren S. *Flexible imputation of missing data, 2nd ed*. Boca Raton, Fl: Chapman& Hall/CRC Press, 2018.
34 Vartanian O, Lam TK, Maceda E, *et al*. Can a fast thinker be a good thinker? The neural correlates of base-rate neglect measured using a two-response paradigm. *Cogn Neuropsychol* 2021;38:365–86.
35 Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med* 2003;78:775–80.
36 Saposnik G, Redelmeier D, Ruff CC, *et al*. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak* 2016;16:138.
37 Barrows HS, Norman GR, Neufeld VR, *et al*. The clinical reasoning of randomly selected physicians in general medical practice. *Clin Invest Med* 1982;5:49–55.
38 Kostopoulou O, Russo JE, Keenan G, *et al*. Information distortion in physicians' diagnostic judgments. *Med Decis Making* 2012;32:831–9.
39 Croskerry P. Cognitive forcing strategies in clinical decisionmaking. *Ann Emerg Med* 2003;41:110–20.
40 Croskerry P. When I say… cognitive debiasing. *Med Educ* 2015;49:656–7.
41 Sherbino J, Kulasegaram K, Howey E, *et al*. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. *CJEM* 2014;16:34–40.
42 Staal J, Hooftman J, Gunput STG, *et al*. Effect on diagnostic accuracy of cognitive reasoning tools for the workplace setting: systematic review and meta-analysis. *BMJ Qual Saf* 2022;31:899–910.
43 Pennycook G, Fugelsang JA, Koehler DJ. What makes us think? A three-stage dual-process model of analytic engagement. *Cogn Psychol* 2015;80:34–72.
44 De Neys W. Bias and conflict: a case for logical intuitions. *Perspect Psychol Sci* 2012;7:28–38.
45 De Neys W, Bonnefon J-F. The "whys" and "whens" of individual differences in thinking biases. *Trends Cogn Sci* 2013;17:172–8.
46 Brewer NT, Chapman GB, Schwartz JA, *et al*. The influence of irrelevant anchors on the judgments and choices of doctors and patients. *Med Decis Making* 2007;27:203–11.
47 Mussweiler T, Englich B. Adapting to the Euro: evidence from bias reduction. *J Econ Psychol* 2003;24:285–92.
48 Schmittat SM, Englich B. If you judge, investigate! Responsibility reduces confirmatory information processing in legal experts. *Psychology, Public Policy, and Law* 2016;22:386–400.
49 Eva KW, Cunnington JPW. The difficulty with experience: does practice increase susceptibility to premature closure? *J Contin Educ Health Prof* 2006;26:192–8.
50 Eva KW, Link CL, Lutfey KE, *et al*. Swapping horses midstream: factors related to physicians' changing their minds about a diagnosis. *Academic Medicine* 2010;85:1112–7.
51 St-Onge C, Landry M, Xhignesse M, *et al*. Age-related decline and diagnostic performance of more and less prevalent clinical cases. *Adv in Health Sci Educ* 2016;21:561–70.
52 Vandergrift JL, Weng WF, Gray BM. The association between physician knowledge and inappropriate medications for older populations. *J Am Geriatr Soc* 2021;69:3584–94.
53 Meyer AND, Payne VL, Meeks DW, *et al*. Physicians' diagnostic accuracy, confidence, and resource requests. *JAMA Intern Med* 2013;173:1952.
54 Peabody JW, Luck J, Glassman P, *et al*. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715–22.