Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents

Silvia Mamede; Tamara van Gog; Kees van den Berge; et al.


http://jama.ama-assn.org/cgi/content/full/304/11/1198
Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents

Silvia Mamede, MD, PhD
Tamara van Gog, MSc, PhD
Kees M. J. P. Berge, MD
Remy M. J. P. Rikers, MSc, PhD
Jan L. C. M. van Saase, MD, PhD
Coen van Guldener, MD, PhD
Henk G. Schmidt, MSc, PhD

A major aim of every clinical teacher is to foster the quality of students’ and residents’ clinical reasoning, one of the most important factors affecting individual physicians’ performance. Diagnostic errors constitute a substantial portion of preventable medical mistakes, and they have been attributed to a large extent to faulty clinical reasoning. The development of educational strategies to minimize flaws in clinical reasoning depends on a better understanding of their underlying cognitive mechanisms.

Cognitive biases are a source of flaws in reasoning processes. At least 40 types of biases that may affect clinical reasoning have been described. A prime example is a biased use of the availability heuristic (the tendency to weigh likelihood of things by how easily they are recalled), which may erroneously lead a physician to consider a diagnosis more frequently and judge it as more likely if it comes to mind more easily. Relying on availability is often helpful during reasoning because things that come to mind easily generally do occur more frequently. However, a serious problem may arise when this first impression is wrong, because physicians often become anchored in their initial hypothesis, looking for confirming evidence to support their initial diagnosis, underestimating evidence against it, and therefore failing to adjust their initial impression in light of all available information.

Context Diagnostic errors have been associated with bias in clinical reasoning. Empirical evidence on the cognitive mechanisms underlying biases and effectiveness of educational strategies to counteract them is lacking.

Objectives To investigate whether recent experience with clinical problems provokes availability bias (overestimation of the likelihood of a diagnosis based on the ease with which it comes to mind) resulting in diagnostic errors and whether reflection (structured reanalysis of the case findings) counteracts this bias.

Design, Setting, and Participants Experimental study conducted in 2009 at the Erasmus Medical Centre, Rotterdam, with 18 first-year and 18 second-year internal medicine residents. Participants first evaluated diagnoses of 6 clinical cases (phase 1). Subsequently, they diagnosed 8 different cases through nonanalytical reasoning, 4 of which had findings similar to previously evaluated cases but different diagnoses (phase 2). These 4 cases were subsequently diagnosed again through reflective reasoning (phase 3).

Main Outcome Measures Mean diagnostic accuracy scores (perfect score, 4.0) on cases solved with or without previous exposure to similar problems through nonanalytical (phase 2) or reflective (phase 3) reasoning and frequency that a potentially biased (ie, phase 1) diagnosis was given.

Results There were no main effects, but there was a significant interaction effect between “years of training” and “recent experiences with similar problems.” Results consistent with an availability bias occurred for the second-year residents, who scored lower on the cases similar to those previously encountered (1.55; 95% confidence interval [CI], 1.15-1.96) than on the other cases (2.19; 95% CI, 1.73-2.66; P=.03). This pattern was not seen among the first-year residents (2.03; 95% CI, 1.55-2.51 vs 1.42; 95% CI, 0.92-1.92; P=.046). Second-year residents provided the phase 1 diagnosis more frequently for phase 2 cases they had previously encountered than for those they had not (mean frequency per resident, 1.44; 95% CI, 0.93-1.96 vs 0.72; 95% CI, 0.28-1.17; P=.04). A significant main effect of reasoning mode was found: reflection improved the diagnoses of the similar cases compared with nonanalytical reasoning for the second-year residents (2.03; 95% CI, 1.49-2.57) and the first-year residents (2.31; 95% CI, 1.89-2.73; P=.006).

Conclusion When faced with cases similar to previous ones and using nonanalytic reasoning, second-year residents made errors consistent with the availability bias. Subsequent application of diagnostic reflection tended to counter this bias; it improved diagnostic accuracy in both first- and second-year residents.

JAMA. 2010;304(11):1198-1203 www.jama.com

©2010 American Medical Association. All rights reserved.
The scientific literature on the availability bias in medicine is mainly descriptive. Some correlational studies suggest that it occurs, but these do not allow causal inferences to be made. Experimental research is required to provide direct evidence for availability bias in medical diagnosis but, to the best of our knowledge, it is lacking. Moreover, if documented, it is perhaps even more important to medical education and practice to investigate ways in which availability bias can be counteracted.

Expertise might play a role in bias. Experienced physicians tend to rely more on nonanalytical (or System 1) reasoning based on pattern recognition to diagnose routine problems; this is a rapid, largely unconscious diagnostic approach. Although effective and highly efficient in most cases, it might be more easily affected by biases. One way to counteract biases, suggested by studies in psychology, is to induce physicians to adopt more reflective (or analytical, also referred to as System 2) reasoning, which comprises careful, effortful consideration of findings in a case, or to combine nonanalytical and analytical reasoning.

Therefore, we investigated whether availability bias occurs when physicians diagnose cases that have clinical manifestations similar to those of recently encountered cases and if so, whether reflection could counteract this bias. Because nonanalytical reasoning develops in association with clinical experience, we also investigated whether there would be a difference in degree of bias between residents in the first and second year of a residency program. We hypothesized that (1) recent experiences with clinical problems would generate an availability bias when physicians nonanalytically diagnose subsequent cases of similar diseases; (2) more experienced residents would be more prone to this bias; and (3) reflective reasoning would counteract this bias and improve diagnostic accuracy.

**METHODS**

**Overview**

This experiment consisted of 3 phases conducted sequentially in a single session (Table 1). *Phase 1*, exposure, required participants to evaluate the accuracy of a diagnosis provided for 6 different cases. *Phase 2*, nonanalytical diagnosis, required participants to diagnose 8 new cases, 4 of which had clinical manifestations that were similar to 2 of the diseases encountered in phase 1. This was expected to induce an availability bias for those 4 cases and reduce diagnostic accuracy. *Phase 3*, reflective diagnosis, required participants to reflect on the diagnosis of the 4 cases that could have been influenced by an availability bias in phase 2. This was expected to override the bias and lead to more accurate diagnoses.

**Participants**

Thirty-six out of 42 eligible internal medicine residents (participation rate, 85.7%) from the Erasmus Medical Centre, Faculty of Medicine, Erasmus University Rotterdam (mean [SD] age, 29.5 [2.1] years) volunteered to participate in this study. Eighteen were in their first and 18 were in their second year of the residency program. The study took place during an educational meeting held in September 2009; the academic year starts in January for the majority of the residents. Participants did not receive any compensation or other incentives. The nonparticipants were either doing shifts or on holidays. The ethics review committee from the Department of Psychology, Erasmus University Rotterdam provided approval for this study. Because the nature of the study prevented prior disclosure of its objectives, oral consent was obtained after informing participants about their tasks. Debriefing was provided later.

**Procedure**

In total, 16 written clinical cases were used in this study (Table 1). Cases consisted of a brief description of a patient’s medical history, signs and symp-
Box. Example of a Case (Diagnosis: Celiac Disease)

A 27-year-old woman presented with 11-month duration of complaints of diarrhea, flatulence, and episodes of abdominal cramps. She has had stools 5 to 6 times a day, and has often woken up during the night for defecation. The feces are voluminous and soft without mucus, blood, or pus. The abdominal cramps are more severe just before defecation, after which they become less painful. The patient is fatigued and has experienced a 5-kg weight loss over the past 11 months. She also noticed red spots on her skin. She says that she has not had fever or joint pains. The patient consulted a physician 4 months ago as well. The physician prescribed ferrous sulfate for anemia, which she has been using until now. Family history: her father was treated for lung tuberculosis 20 years ago.

Physical Examination
Young, somewhat emaciated woman of otherwise healthy appearance. BP: 110/70; pulse: 80/min; temperature: 36°C. Mucocutaneous paleness (+/4). No other abnormalities.

Laboratory Tests
Hemoglobin: 9 g/dL; hematocrit: 34%; Mean corpuscular volume: 74 fl; serum iron: 45 mg/dL (normal, 50-170 mg/dL); calcium: 8.1 mg/dL (normal, 8.6-10 mg/dL); albumin: 3.2 g/dL (normal, 3.4-4.8 g/dL); Alanine aminotransferase test: 38 U/L; Aspartate aminotransferase test: 25 U/L; Prothrombin time 24 seconds (normal, 12-22 seconds). Feces revealed no worm eggs, no parasites, no white cells; stool fat level was 12g/24 h (normal, <7g/24 h), D-xylose test was positive. Human immunodeficiency virus antibodies: negative. Tuberculosis skin test (PPD): 5 mm.

Imaging Tests
Chest x-ray: no abnormalities; colonoscopy: no abnormalities.

To convert alanine aminotransferase from U/L to µkat(kat)/L, multiply by 0.0167; aspartate aminotransferase from U/L to µkat(kat)/L, multiply by 0.0167; iron from µg/dL to µmol/L, multiply by 0.179.
vided. The diagnoses were evaluated as fully correct, partially correct, or incorrect, and scored as 1, 0.5, or 0 points, respectively. A diagnosis was considered fully correct whenever the core diagnosis was cited by the participant and partially correct when the core diagnosis was not mentioned but a constituent element of the diagnosis was cited. For example, in the case in the Box, “celiac disease” was scored as correct, and “malabsorption” as partially correct.

For each participant, we separately summed the scores obtained in phase 2 on the 4 cases that had similarities to the cases encountered in phase 1 and the 4 cases that did not. For phase 3, the diagnostic scores obtained on the 4 cases were summed for each participant.

For phase 2, an analysis of variance (ANOVA) with years of training as a between-subjects factor (first vs second year) and recent experiences with similar cases as a within-subjects factor (with vs without) was conducted on the mean diagnostic performance scores obtained through nonanalytical reasoning on both types of cases (similar to cases seen in phase 1 or not). This analysis tested the hypothesis that recent experiences with similar cases would generate an availability bias and that this bias would be larger for more experienced, second-year residents. Post hoc paired t tests were performed to compare the diagnostic performance of first- and second-year residents under the 2 experimental conditions. To assess whether the diagnoses of the cases encountered in phase 1 were indeed provided as diagnosis of the similar cases in phase 2, we computed the number of times the diagnoses of cases in phase 1 were mentioned by participants in phase 2 who had seen similar cases in phase 1 vs those who had not and conducted paired t tests on these data for the first- and second-year residents.

A second ANOVA with years of training as a between-subjects factor (first year vs second year) and type of reasoning as a within-subjects factor (nonanalytical vs reflective) was conducted on the mean diagnostic performance scores in phase 2 and phase 3. This analysis tested the hypothesis that reflection (phase 3) could counteract the availability bias by improving the diagnostic performance scores compared with those of the first- and second-year residents who had encountered the similar cases through nonanalytical reasoning (phase 2).

Significance was set at P<.05 for all comparisons (2-tailed). SPSS version 15.0 (SPSS Inc, Chicago, Illinois) for Windows was used for the statistical analyses.

## RESULTS

### Nonanalytic Reasoning (Phase 2)

Table 2 presents the mean diagnostic accuracy scores obtained by first-year and second-year residents when cases were solved through nonanalytical reasoning (phase 2). The ANOVA showed no significant main effects, but there was a significant interaction effect between years of training and recent experiences with similar cases (F[1,34]=10.35, mean square error (MSE)=0.68, P=.003, n²=0.23). Mean scores for the second-year residents were consistent with an availability bias. They obtained significantly lower diagnostic scores on the cases similar to those encountered in phase 1 than the other cases (0.4 scale, 1.55; 95% confidence interval [CI], 1.15-1.96 vs 2.19; 95% CI, 1.73-2.66; P=.03).

Among the 8 phase 2 cases potentially similar to phase 1, second-year residents more frequently gave the phase 1 diagnosis when they had encountered the cases in phase 1 than when they had not (mean frequency per resident, 1.44; 95% CI, 0.93-1.96 vs 0.72; 95% CI, 0.28-1.17; P=.04) (Table 3). Even when the participants had not encountered the similar cases in phase 1, they sometimes incorrectly provided the phase 1 diagnosis to the related cases but this occurred less frequently than when they had been previously exposed to the phase 1 cases.

### Table 2. Mean Diagnostic Accuracy Scores in Phase 2 (Perfect Score, 4.0) as a Function of Previous Exposure to Similar Cases in Phase 1

<table>
<thead>
<tr>
<th>Training Year</th>
<th>Mean Score (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First (n = 18)</td>
<td>2.03 (1.55-2.51)</td>
</tr>
<tr>
<td>Second (n = 18)</td>
<td>1.55 (1.15-1.96)</td>
</tr>
</tbody>
</table>

### Table 3. Persistence of Phase 1 Diagnoses Among 8 Potentially Similar Cases in Phase 2 and 4 Similar Cases in Phase 3

<table>
<thead>
<tr>
<th>Training Year</th>
<th>First (n = 18)</th>
<th>Second (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of phase 1 diagnosis given for phase 2 cases, mean (95% CI) [ % of all wrong diagnoses]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Having encountered similar cases in phase 1</td>
<td>0.78 (0.34-1.26) [42]</td>
<td>1.44 (0.93-1.96) [63]</td>
</tr>
<tr>
<td>Not having encountered similar cases in phase 1</td>
<td>0.89 (0.47-1.30) [36]</td>
<td>0.72 (0.28-1.17) [42]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correction of phase 1 diagnoses in phase 3</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected after reflection in phase 3</td>
<td>14 (36)</td>
</tr>
<tr>
<td>Adhered to after reflection in phase 3</td>
<td>7 (50)</td>
</tr>
<tr>
<td>Changed to a different incorrect diagnosis</td>
<td>2 (14)</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.²Comparison of phase 1 diagnoses with vs without having encountered similar cases in phase 1, 2-sided t test.
In contrast, this pattern was not seen for the first-year residents, who had a higher score on the cases similar to those encountered in phase 1 than on the other cases (Table 2). Having encountered a similar case in phase 1 did not lead to more frequently giving this diagnosis in phase 2 than when they had not seen a similar case (mean frequency per resident, 0.78; 95% CI, 0.34-1.26 vs 0.89; 95% CI, 0.47-1.30; P = .67) (Table 3).

Reflective Reasoning (Phase 3 vs Phase 2)
The diagnostic scores obtained through reflective reasoning (phase 3) on the cases similar to the diseases that had been encountered in phase 1 (those cases subject to an availability bias in phase 2) are presented in Table 4. A significant main effect of “type of reasoning” was found in the ANOVA (F[1,34] = 8.46, MSE = 0.30, P = .006, η² = 0.20) indicating that reflection improved all participants’ diagnoses compared with nonanalytical reasoning. The percentage of phase 1 diagnoses that were corrected or adhered to after reflection is shown in Table 3.

Encountering only one case of a disease was sufficient to make the second-year residents more prone to incorrectly giving that diagnosis to subsequent cases of different, though similar, diseases. In emergency departments and outpatient clinics, physicians are likely to see (often close in time) several patients with similar symptoms caused by different diseases. In many clinical settings, therefore, conditions propitious for the occurrence of the availability bias prevail.

Moreover, because reliance on nonanalytical reasoning tends to increase with experience, it is possible that physicians with many years of clinical practice may be even more susceptible to availability bias than second-year residents, and this should be investigated. In real-life situations, an initial incorrect hypothesis might be spontaneously revised before expensive or time-consuming tests are ordered. However, the effects of anchoring by an early incorrect diagnosis may still lead to inaccurate judgment and inappropriate decisions. More experienced clinicians appear to be more subject to an anchoring effect, which makes it less likely that they will spontaneously overrule an incorrect initial diagnosis.

These findings contribute some insight into cognitive mechanisms underlying errors, which are the object of ongoing scientific debate. Evidence of the availability bias emerged in phase 2, when participants diagnosed the cases through a nonanalytical reasoning mode, and this was in part repaired in phase 3 by reflective reasoning. This suggests that mistakes made in phase 2 did not derive from lack of knowledge. Residents who failed to correctly diagnose the cases through nonanalytical reasoning may have arrived at the correct diagnoses after reflecting on the same cases by activating existing knowledge. Therefore, errors in phase 2 were more likely to have been provoked by bias in the reasoning processes.

We had expected the availability bias to be larger for the more experienced residents because the tendency to diagnose cases through pattern recognition increases with clinical experience. However, we had not expected to find an opposite pattern for the first-year residents, who had better performance on similar cases. It is possible to speculate on reasons for this finding, such as that these novice residents might have already used a more reflective mode of reasoning during the exposure phase (phase 1), being less self-confident than their more experienced colleagues and, therefore, perhaps less reliant on immediate decisions. They may not have had a sufficient amount of clinical experience to make extensive use of pattern-recognition and had to rely on a more analytic approach that could have been activated by phase 1 cases. However, as a post hoc analysis yielding an unexpected finding, these are speculations that should only be interpreted as hypothesis generating.

Although reliance on nonanalytical reasoning and heuristics such as availability work well in many situations, reducing the time and effort involved in decision making and allowing physicians to make accurate diagnoses in routine situations, it may open the door to cognitive bias. Reflection has been shown to improve diagnosis when problems are complex or nonroutine, and this study indicates that reflection may also be a mechanism to counteract cognitive biases.

With respect to medical education, this study suggests that a relatively simple instructional procedure can be used to induce reflective reasoning and improve diagnostic accuracy. This procedure for reflective reasoning can be implemented relatively easily in educational situations. Further research should investigate the effects.

### Table 4. Mean Diagnostic Accuracy Scores in Phase 2 (Perfect Score, 4.0) and Phase 3 for the Cases Similar to Those Encountered in Phase 1

<table>
<thead>
<tr>
<th>Training Year</th>
<th>First (n = 18)</th>
<th>Second (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonanalytical diagnostic reasoning (Phase 2)</td>
<td>Mean Score (95% Confidence Interval)</td>
<td>2.03 (1.55-2.51)</td>
</tr>
<tr>
<td>Reflective diagnostic reasoning (Phase 3)</td>
<td>Mean Score (95% Confidence Interval)</td>
<td>2.31 (1.89-2.73)</td>
</tr>
</tbody>
</table>

*There was a significant main effect of reflection, which improved diagnoses compared to nonanalytical reasoning for both first and second year residents (P = .008 by ANOVA).
REFERENCES


Author Contributions: Dr Mamede had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Mamede, Berge, Rikers, Schmidt.

Acquisition of data: Berge, Saase, Guldener.

Analysis and interpretation of data: Mamede, Gog, Berge, Rikers, Saase, Guldener, Schmidt.

Drafting of the manuscript: Mamede, Gog.

Critical revision of the manuscript for important intellectual content: Mamede, Gog, Berge, Rikers, Saase, Guldener, Schmidt.

Statistical analysis: Mamede, Gog.

Administrative, technical, or material support: Rikers, Saase, Guldener.

Study supervision: Schmidt.

Financial Disclosures: None reported.

Funding/Support: None.

Additional Contributions: We thank Júlio César Penaforte, MD, MSc (Hospital Geral de Fortaleza, Brazil) and João Macedo Coelho Filho, MD, PhD (Faculty of Medicine, Federal University of Ceará, Brazil) for their permission to use the clinical cases that they prepared for previous studies, without compensation.

©2010 American Medical Association. All rights reserved.