Disrupting Diagnostic Reasoning: Do Interruptions, Instructions, and Experience Affect the Diagnostic Accuracy and Response Time of Residents and Emergency Physicians?

Sandra D. Monteiro, PhD, Jonathan David Sherbino, MD, Jonathan Seth Ilgen, MD, Kelly L. Dore, PhD, Timothy J. Wood, PhD, Meredith E. Young, PhD, Glen Bandiera, MD, Danielle Blouin, MD, Wolfgang Gaissmaier, PhD, Geoffrey R. Norman, PhD, and Elizabeth Howey

Abstract

Purpose

Others have suggested that increased time pressure, sometimes caused by interruptions, may result in increased diagnostic errors. The authors previously found, however, that increased time pressure alone does not result in increased errors, but they did not test the effect of interruptions. It is unclear whether experience modulates the combined effects of time pressure and interruptions. This study investigated whether increased time pressure, interruptions, and experience level affect diagnostic accuracy and response time.

Method

In October 2012, 152 residents were recruited at five Medical Council of Canada Qualifying Examination Part II test sites. Forty-six emergency physicians were recruited from one Canadian and one U.S. academic health center. Participants diagnosed 20 written general medicine cases. They were randomly assigned to receive fast (time pressure) or slow condition instructions. Visual and auditory case interruptions were manipulated as a within-subject factor.

Results

Diagnostic accuracy was not affected by interruptions or time pressure but was related to experience level: Emergency physicians were more accurate (71%) than residents (43%) ($F = 234.0$, $P < .0001$) and responded more quickly (54 seconds) than residents (65 seconds) ($F = 9.0$, $P < .005$). Response time was shorter for participants in the fast condition (55 seconds) than in the slow condition (73 seconds) ($F = 22.2$, $P < .0001$). Interruptions added about 8 seconds to response time.

Conclusions

Experienced emergency physicians were both faster and more accurate than residents. Instructions to proceed quickly and interruptions had a small effect on response time but no effect on accuracy.

Diagnostic errors are a significant problem in medicine. In the emergency department (ED), the prevalence may as high as 20%.[1] Perhaps not surprisingly, then, much of the literature on diagnostic errors examines the ED. As Croskerry[2] explains:

> In this milieu, decision-making is often naked and raw, with its flaws highly visible. Nowhere in medicine is rationality more bounded by relatively poor access to information and with limited time to process it, all within a milieu renowned for its error-producing conditions.

One characteristic feature of the ED is multiple interruptions, estimated at 10 to 20 per hour, which may increase errors.[3–5] Chisholm and colleagues[3] argue that to deal with disruptions, providers should employ cognitive forcing strategies, which are generalized techniques for increasing self-awareness:

> There is evidence supporting the negative effect of interruptions on task performance and subject perception of stress and … interruptions likely cause emergency medicine providers to compensate through task short cuts or failure to reengage in the task. This suggests that teaching cognitive forcing strategies to reorient after an interruption … may be beneficial to ED providers.[3(p121)]

Chisholm is one of a number of authors[6–8] who have argued that diagnostic errors arise from cognitive biases and can be reduced through explicit instruction on the nature of cognitive biases. Allied with the explicit identification of cognitive bias is the recommendation to slow down the decision-making process to allow more time for deliberation and reflection. Kahneman[9(p417)] writes:

> The way to block errors that originate in System 1 is simple in principle: recognize the signs that you are in a conceptual minefield, *slow down*, and ask for reinforcement from System 2. [emphasis added]

Both of these strategies are derived from a "default-interventionist dual processing" model of reasoning.[10,11] In this model, System 1 engages fast, nonanalytical processes that require few cognitive resources and is the source of all cognitive biases. System 2 engages slower, deliberate analytical processes that function to override the biases and correct errors through application of logical deduction.

However, despite the appeal of attributing diagnostic errors to cognitive biases related to rapid reasoning, the evidence is not persuasive. First, although some studies have associated diagnostic errors with cognitive biases,[12] other research has found errors to be primarily a consequence of mistakes, defining "mistake" as "an intended act, but the

Please see the end of this article for information about the authors.

Correspondence should be addressed to Dr. Monteiro, McMaster University, Program for Educational Research and Development, MDCL 3519, 1280 Main St. W., Hamilton, Ontario, L8S 4L1, Canada; telephone: (905) 525-9140, ext. 22614; e-mail: monteisd@mcmaster.ca.

Acad Med. 2015;90:511–517.
First published online January 6, 2015
doi: 10.1097/ACM.0000000000000614

physician does not *know* it is incorrect"[13] [emphasis added]. Second, if diagnostic errors are caused by cognitive biases, then debiasing strategies should reduce errors. However, as Graber and colleagues[14] have noted, there is evidence that debiasing strategies have no effect on errors.[15,16] Finally, contrary to the view that increasing reliance on nonanalytical reasoning should increase diagnostic errors, experienced (and older) physicians rely more on nonanalytical System 1 than analytical System 2 reasoning,[17,18] yet they are clearly more accomplished diagnosticians than less experienced (and younger) physicians.

Lloyd and Reyna[19,20] argue that diagnostic expertise is well represented by "fuzzy-trace theory," a different dual-process model which posits that humans rely on two kinds of memory—verbatim and gist. Increasing expertise is associated with greater reliance on gist memory. This differs from the default-interventionist dual-process model in at least two fundamental ways. First, "fuzzy-trace theorists place intuition at the apex of development rather than at the nadir."[20] Second, the dual processes of the fuzzy-trace model refer more to the nature of knowledge in memory—gist versus verbatim—than to the process of thinking.

The implications of these two opposing viewpoints are profound. If the default-interventionist model is correct, medical educators should devise educational strategies to discourage intuitive thinking and to ensure that clinicians understand the kinds of cognitive biases that arise so they can consciously and deliberately override them. Conversely, if fuzzy-trace theory is correct, that would be exactly the *wrong* thing to do. Medical educators should instead rely on clinical experiences to gradually induce intuition and automaticity as hallmarks of expertise. Moreover, the locus of intervention changes. In the default-interventionist perspective, the focus is on improving clinicians' thinking strategies by overcoming general biases; in the fuzzy-trace view, the focus is on seeking to enhance clinicians' experiential "gist" knowledge.

However, any attempt to test these two perspectives by estimating error rates arising from System 1 and System 2 processes is likely to be frustrated. It appears unlikely that the usual diagnostic task, which may occur over minutes or hours, could be viewed as purely System 1 or System 2. The issue is not whether the two systems exist in isolation but, rather, whether they should be viewed as competitors, with the errors of System 1 overridden by System 2. If System 2 has a central monitoring role in reducing System 1 errors, then increasing the analytical resources devoted to the problem by, for example, increasing the time available to complete a task, should reduce diagnostic errors. Conversely, decreasing the effectiveness of System 2 thinking by restricting the time available to complete a task or by introducing distractions should increase diagnostic errors.

One recent observational study was not consistent with this hypothesis: Sherbino and colleagues[21] found that diagnostic accuracy was strongly correlated with reduced time—a correlation of −0.55. In two recent experimental studies contrasting a "fast" cohort (told to go as fast as possible without making mistakes) and a "slow" cohort (told to be careful to examine all the data), instructions had an influence on the time taken in a diagnostic task yet had no impact on error rates.[22,23] However, those studies took place in an interruption-free environment, so they did not provide a direct test of whether strategies designed to encourage slow, deliberate thinking will directly reduce the impact of interruptions on the accuracy of clinical reasoning. Moreover, these studies involved residents only. It may be that junior physicians, who rely more on analytical processing than experienced physicians do,[23] may be more vulnerable to the effect of interruptions.

To address these issues, in this study we investigated the effects of time pressure, experience, and interruptions on accuracy in a diagnostic task. Similar to previous studies,[21,22] participants were asked to diagnose a series of general medicine cases and were randomly assigned to receive instructions that encouraged either a faster or slower approach. The critical additions of this study's design were (1) the inclusion of experienced emergency physicians and (2) the requirement that participants manage interruptions.

We addressed the following research questions: What is the effect of increasing time pressure, through interruptions and instructions to proceed quickly, on response time and diagnostic accuracy? What is the effect of experience on response time and diagnostic accuracy?

Method

This study was approved by the McMaster research ethics board (REB; approval 10-409). It also received REB approval at McGill University, Queen's University, the University of Ottawa, the University of Toronto, and the University of Washington.

Design

Emergency physicians and junior residents (primarily in their second postgraduate year) were recruited to complete a diagnostic task on a computer. Participants were randomly assigned to receive instructions to diagnose more quickly and within a shorter time limit (30 minutes; "fast" condition, with time pressure) or more carefully within a longer time limit (45 minutes; "slow" condition). In both conditions, interruptions were presented during a random half of the cases in one of two counterbalanced orders. Therefore, we used a randomized mixed-model design with two between-subject factors, instruction (fast or slow) and experience level (emergency physician or resident); and one within-subject factor, interruptions (present or absent).

Participants

Emergency physicians: Recruitment and setting. In October 2012, 46 emergency physicians were recruited from two academic health centers: McMaster University Medical Centre in Hamilton, Ontario (n = 21 of 25 approached) and University of Washington Medical Center in Seattle, Washington (n = 25 of 47 approached). They received an e-mail inviting them to participate by author J.S. (in Hamilton) or J.I. (in Seattle). The invitation provided a Web link through which the computer-based experiment could be downloaded directly to a personal computer. The emergency physicians participated in the study outside of work hours and in a quiet location of their choice. Upon following the study link, these participants were asked to provide electronic consent, and then to diagnose a series of cases. Upon completion of the study, they were offered an honorarium of $50.

Residents: Recruitment and setting. In October 2012, 152 residents with a minimum of two years of postgraduate training or the equivalent were recruited across Ontario and Quebec, Canada, following their completion of the Medical Council of Canada Qualifying Examination Part II (MCCQE Part II). Residents were recruited at the following five testing sites, where the examination was administered only in English: McGill University Faculty of Medicine (n = 55; 36%), Queen's University Faculty of Health Sciences (n = 21; 14%), University of Toronto (n = 30; 20%), University of Ottawa Faculty of Medicine (n = 23; 15%), and McMaster University Michael G. DeGroote School of Medicine (n = 23; 15%). Because the MCCQE Part II is a requirement for licensure, the residents were from a variety of specialties. Residents were recruited by a personal approach during a sequestering period following completion of the examination. They were signed up on a "first come, first served" basis until the maximum dictated by availability of computers at the site was reached.

These participants provided written consent prior to the task. Upon completion of the study, they were offered an honorarium of $30.

Diagnostic task

Instructions. All participants were randomly assigned to receive one of two sets of instructions (described below) and were asked to diagnose 20 general medicine cases presented as brief, written clinical vignettes. After reading a case, participants advanced to the next screen and entered a diagnosis of five words or less; once they advanced, they could not go back. They then went to the next case. A list of correct diagnoses was provided at the end of the experiment.

Slow condition. Participants in the slow condition received these oral and written instructions:

> For the next 45 minutes you will be presented with information about 20 general medical cases. We are asking you to provide a single best diagnosis for each one of them. Make sure you consider all the data before you arrive at your diagnosis.

As they worked through the cases, these participants were shown a progress bar in the corner of the screen that indicated how many cases had elapsed.

Fast condition. Participants in the fast condition received these oral and written instructions:

> Imagine that you are in a busy emergency department. As usual there is a large backlog of patients. You are about to see a series of 20 general medicine cases, and you have a limited time (that is about 30 minutes) to see them all. It's possible you won't be able to complete all the cases, but work as quickly as you can without sacrificing accuracy.

As they worked through the cases, these participants were shown a digital timer in the corner of the computer screen, indicating the decreasing amount of time available for the case.

Interruptions. Participants were informed they would have to attend to occasional interruptions. Two kinds of interruptions occurred: auditory and visual. Each interrupted case contained both interruptions. The auditory interruption occurred first; after a delay, the visual interruption occurred. The auditory interruption was a simulated page, asking the participant to remember a four-digit number; these pages were presented via the computer's audio system. Residents, in a computer lab, were provided with headphones. Participants were asked to remember the number, without writing it down, and to enter it after they diagnosed the interrupted case. The visual interruption was a multiple-choice question (MCQ) about general medical knowledge that was unrelated to the case. When presented, the MCQ replaced the case on the screen until the participant selected the correct answer.

Interruptions occurred in exactly half the cases (n = 10) in a counterbalanced design, so that the cases that were interrupted for one group of participants (residents or emergency physicians) were not interrupted for the other group of participants. Interruptions always occurred about one-third of the way into the encounter with the case, based on average reading times for a single case from prior studies.[15,16] On the basis of this estimate, we set the fast condition time at 30 minutes. If participants completed all 20 cases, with two interruptions on 10 cases (half of the cases), there would be 20 interruptions during the 30 to 45 minutes allotted to complete the task. As noted above, observational studies[3,4] have indicated that an emergency physician can anticipate 10 to 20 interruptions per hour, which aligns well with our design.

Materials. The general medicine cases used in a previous study[21] were presented on the participant's computer screen. One-half of the screen displayed written descriptions of the primary complaint, the relevant patient history, and the relevant test results. The other half of the screen displayed an image presenting findings of the physical exam and/or investigation (e.g., CT scan, x-ray). The measurement units were appropriate to the country of practice of the participant. (All residents were presented with Canadian units of measurement.)

Measurement of performance

Diagnostic task. Using a previously developed rubric,[21] we scored free-text responses to the diagnostic cases on a three-point scale, where 0 = incorrect, 1 = partially correct, and 2 = completely correct.

Interruptions. The percentage of correct pager reports was calculated. The time taken to respond to each MCQ was recorded.

Analysis

Case response times were computed for each participant, excluding the time taken to manage the visual (MCQ) interruptions. The auditory (pager) interruptions lasted four seconds, but the case was still visible and could be reviewed. Therefore, in the primary analyses, the time required to listen to auditory interruptions was not subtracted from the time taken to complete a case.

An average accuracy score was computed for each participant based on the sum of the completed case scores (maximum possible = 40), divided by the number of cases completed. In this way, scores were adjusted for participants who did not complete all 20 cases. Average accuracy scores and average response times were calculated separately for interrupted and uninterrupted cases.

Data were analyzed with repeated-measure ANOVAs with one within-subject factor (interruptions, present or absent) and two between-subject factors (experience level, emergency physician or resident; instruction, fast or slow) on response time and accuracy. Data were analyzed with IBM SPSS version 19 (IBM Corp., Chicago, Illinois). The alpha level was 0.05 for all analyses.

Average accuracy scores for emergency physicians were submitted to a preliminary one-way ANOVA to test the effect of country of practice (Canada versus the United States). No statistically significant differences were found ($P > .5$), so subsequent analyses were collapsed across this factor.

Finally, to explore the relationship between case difficulty and the independent factors, the accuracy and response time data were analyzed for simple versus complex cases (categorized on the basis of a median accuracy of 45%).

Results

Forty-six emergency physicians and 152 residents participated in the study. As residents were candidates for the MCCQE Part II, they represented a variety of disciplines, with about half the sample from medicine, surgery, and family medicine. Because the MCCQE Part II exam is typically taken after one year of residency, most of the residents (n = 107; 70%) were second-year residents. The few first-year residents (n = 4; 3%) had additional qualifications, so they were able to take the exam sooner. Residents' specialties and training levels are shown in Table 1.

Table 1

Specialties and Training Levels of 152 Residents Participating in the Study of Diagnostic Reasoning at Five Medical Council of Canada Qualifying Examination Part II Sites in Ontario and Quebec, 2012[a]

Characteristic	Residents, no. (%)[b]
Specialty	
Medicine and subspecialties	50 (33)
Surgery and subspecialties	12 (8)
Pediatrics	17 (11)
Family medicine	11 (7)
Emergency medicine	6 (4)
Obstetrics–gynecology	9 (6)
Radiology	10 (7)
Psychiatry	7 (5)
Pathology	4 (3)
Anesthesiology	12 (8)
Other	14 (9)
Level of training	
PGY 1	4 (3)
PGY 2	107 (70)
PGY 3	6 (4)
PGY 4+	3 (2)
International medical graduate	26 (17)
Other	6 (4)

Abbreviations: PGY indicates postgraduate year; other includes participants identified as postcertification in program year.

[a]The sites were McGill University Faculty of Medicine, Queen's University Faculty of Health Sciences, University of Toronto, University of Ottawa Faculty of Medicine, and McMaster University Michael G. DeGroote School of Medicine.

[b]Total may exceed 100% because of rounding.

The emergency physicians had a mean age of 40 years and had been in practice an average of 8 years. They reported working an average of 12.6 shifts per month with an average of 23.8 patients seen per shift.

Effect of experience level on diagnostic accuracy and response time

Forty-five (98%) of the 46 physicians, and 136 (89%) of the 152 residents, completed all 20 cases. All participants contributed to the analysis, however, because average accuracy was computed for cases completed. The average accuracy scores for physicians and residents by instruction condition are displayed in Figure 1. Overall diagnostic accuracy for emergency physicians (mean = 71%, SD = 8.8%) was significantly higher than for residents (mean = 43%, SD = 10.6%) ($F = 234.0$, $df = 1, 194$, $P < .0001$). This analysis was repeated for the subsample of 73 residents from medicine, family medicine, and surgery, and the results were essentially the same, with a mean of 44%.

Average response times for emergency physicians and residents for each instruction condition are presented in Figure 2. On average, residents' processing times to diagnose cases (mean = 65 seconds, SD = 23 seconds) were longer than experienced physicians' (mean = 54 seconds, SD = 17 seconds) ($F = 9.0$, $df = 1, 194$, $P < .005$).

Effect of instructions on diagnostic accuracy, response time, and cases completed

There was no overall effect of instructions on accuracy and no interaction with experience level, as shown in Figure 1. For emergency physicians, average accuracy was 70% (SD = 9.5%) under the fast condition and 71% (SD = 8.0%) under the slow condition ($F < 0.2$, $df = 1, 194$, $P = .71$); for residents, average accuracy was 43% (SD = 10%) under both conditions ($F < 0.05$, $df = 1, 194$, $P = .84$), which is consistent with a previous study using similar materials and residents.[21] A power calculation, based on an alpha of 0.05 and a beta of 0.20, indicated that the study would be able to detect a difference in accuracy of 4%.

Participants in the fast condition had significantly shorter case response times (mean = 55 seconds, SD = 15 seconds) than participants in the slow condition (mean = 73 seconds, SD = 26 seconds) ($F = 22.2$, $df = 1, 194$, $P < .0001$). As a result of taking longer to diagnose cases, fewer participants in the slow condition (93%) completed all cases than participants in the fast condition (97%), despite having an additional 15 minutes available.

Effect of interruptions on diagnostic accuracy and response time

There was no effect of interruptions on accuracy in either participant group (residents: interruptions = 43%, no interruptions = 44%; emergency physicians: interruptions = 71%, no interruptions = 71%) ($F = 0.01$; $df = 1, 194$, $P = .99$). Participants took longer to report a diagnosis in cases with interruptions (mean = 71 seconds, SD = 27) compared with cases without interruptions (mean = 63 seconds, SD = 24) ($F = 51.8$, $df = 1, 197$, $P < .01$). Adjusting the average case response time by subtracting the time required to listen to the auditory interruption still revealed a significant three-second slowing for cases with interruptions ($F = 6.14$, $df = 1, 194$, $P < .01$).

The overall accuracy for reporting the pager number was 84% for emergency physicians and 74% for residents ($t = 2.6$, $df = 45$, $P < .01$). Emergency physicians were faster (mean = 6 seconds, SD = 1) than residents (mean = 9 seconds, SD = 3) at identifying the correct answer to MCQs ($t = -7.1$, $df = 173$, $P < .001$, equal variance not assumed).

Effect of interruptions and instructions on diagnostic accuracy by case difficulty

One plausible explanation for the lack of effect of interruptions and instructions on accuracy may be that the cases were relatively straightforward and

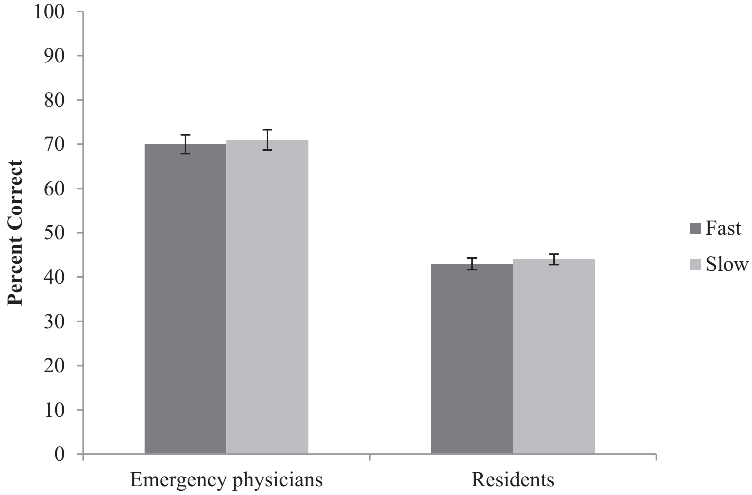

Figure 1 Average accuracy score (percent correct) for Canadian and U.S. emergency physicians (n = 46) and Canadian residents (n = 152) under instructions to proceed quickly (fast condition) or slowly and systematically (slow condition) in diagnosing 20 general medicine cases presented as computer-based, written clinical vignettes, 2012 study of diagnostic reasoning.

so could be solved without extensive analytic processing. (This appears unlikely, however, as the mean accuracy for residents was only 43%.) Another possibility may be that interruptions or instructions to proceed quickly had a greater effect on accuracy for difficult cases, which was masked in the overall means. To examine relation to case difficulty, we used a median split of case accuracy of 45% to divide cases into difficult and easy categories. We then computed average accuracy under fast and slow, interrupted and noninterrupted conditions for each case.

The average accuracy for each case for emergency physicians and for residents is presented in Table 2. In the overall analysis of accuracy, the only significant effect was for case difficulty. There was no effect on accuracy, and no interaction with case difficulty, of instructions or interruptions: While cases analyzed under the fast condition were processed significantly more quickly than cases analyzed under the slow condition (as shown above), and easy cases were processed marginally more quickly than difficult cases (62 seconds versus 89 seconds, F = 2.83, df = 1, 18, P = .10), there was no interaction between case difficulty and interruptions on case response time (F = 0.93, df = 1, 18, not significant).

Discussion

Not surprisingly, we found a clear effect of experience: Emergency physicians were both substantially more accurate and had faster response times than residents. With respect to the effect of interruptions, it has been suggested that multiple interruptions during case management can have serious consequences, as physicians may be unable to automatically reengage in previous tasks.[3–5] In this study, we found no evidence that interruptions adversely affected diagnostic accuracy. Interruptions had only a minor effect on response time both for experienced emergency physicians and for residents. The pager interruption was designed to load on working memory, but participants at both experience levels were able to use strategies that reduced this load. The MCQ interruption required accessing long-term memory, but again had no deleterious effect. The main effect of each interruption, therefore, was to temporarily disengage attention from the current case, but this had no effect on diagnostic accuracy. Although our results appear counterintuitive, they are consistent with research in psychology. Typically, studies of approaches to simultaneous tasks demonstrate that people "task switch." When tasks are confusable, such as requiring similar calculations[24,25] or involving similar types of patients and symptoms, accuracy may be affected. However, when attention is switched between two unrelated tasks, such as diagnosing a clinical case and answering an MCQ, the typical result is an impact on response time but not on accuracy,[24,25] as we observed in this study.

In addition to the practical consequences related to the impact of interruptions, our results have theoretical implications for a default-interventionist dual-processing model. A prevalent view among proponents of this theory is that slowing down, and thereby allowing for more extensive analytical (System 2) processing, will lead to detection and mitigation of errors. We found no evidence that this is so; neither the absence of interruptions nor instructions to consider all the data had any measurable effect on accuracy, although the latter did result in significantly slower mean response time. This is consistent with other studies showing that rapid processing does not affect diagnostic accuracy. Ilgen and colleagues[23] used a fast/slow intervention similar to ours in a study with 393 participants: Accuracy under the "first impression" condition was 64% and under the "directed search"

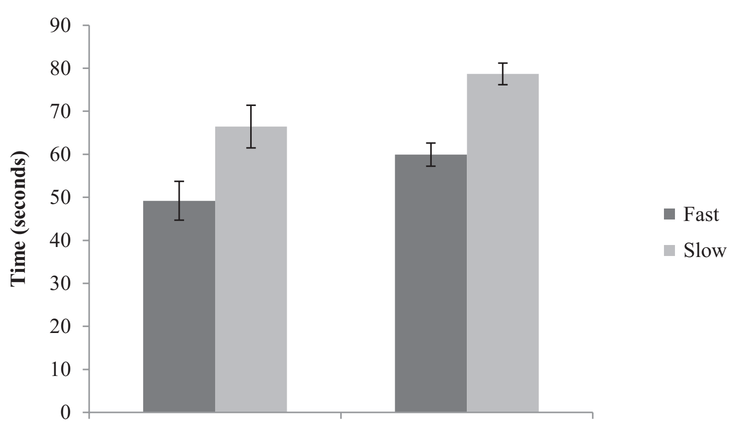

Figure 2 Average case response times for Canadian and U.S. emergency physicians (n = 46) and Canadian residents (n = 152) under instructions to proceed quickly (fast condition) or slowly and systematically (slow condition) in diagnosing 20 general medicine cases presented as computer-based, written clinical vignettes, 2012 study of diagnostic reasoning.

Table 2

Diagnostic Accuracy by Case for 46 Emergency Physicians and 152 Residents on 20 Written General Medicine Cases, 2012 Study of Diagnostic Reasoning[a]

		% Accuracy	
Case	Diagnosis	Emergency physicians	Residents
1	Epidural hematoma	94	82
2	Iritis	61	21
3	Malignant otitis externa	80	29
4	Addison's disease	54	27
5	Posterior STEMI	63	38
6	Torsades de pointes	80	56
7	Pulmonary abscess	78	51
8	Glycoside-induced hypoglycemia	10	5
9	Digoxin-induced toxicity	72	56
10	Bronchitis	31	32
11	Lisfranc fracture	86	38
12	Subconjunctival hemorrhage	99	45
13	CO poisoning	91	46
14	Hemorrhagic cystitis	36	25
15	Perforated viscus	76	53
16	Thoracic aortic dissection	56	47
17	Measles	86	67
18	Lingular pneumonia	54	44
19	Distal radius fracture	98	58
20	Peritonsillar abscess	100	43
Mean		71	43

Abbreviations: STEMI indicates ST elevation myocardial infarction; CO, carbon monoxide.

[a]Participating physicians were from McMaster University Medical Centre in Hamilton, Ontario, Canada, and the University of Washington Medical Center in Seattle, Washington. Participating residents were from a variety of specialties and participated in the study at five Medical Council of Canada Qualifying Examination Part II testing sites across Ontario and Quebec, Canada. Cases were presented as computer-based, written clinical vignettes.

condition was 61%. In a previous study involving 204 residents, Norman and colleagues[22] found accuracy of 46% in a "speed" group versus 45% in a "reflect" group. However, our finding that slower reflective thinking had no advantage for difficult cases is counter to other reported results.[26]

This study has some limitations. First, the findings are based on diagnosis of written cases, which may not be as rich as cases in the "real world" of the ED. However, recent studies have shown that (a) learning from written cases leads to clinical performance equivalent to learning from video cases or standardized patients[27] and (b) assessment based on low-fidelity simulations (e.g., a laptop presentation of an ECG) is equivalent to evaluation with real patients.[28]

Second, the conditions of the experiment were slightly different for the two participant groups; in particular, the experienced emergency physicians diagnosed the cases on their own time and in an environment of their choosing. Several concerns might arise, such as (1) the environment of the experts may have contained more distractions, because the residents were in a completely quiet and invigilated setting; or (2) experts could have "cheated" and consulted resources. However, the first issue would bias against detecting an effect of experience, which was not apparent. The second might be revealed by excessive response time, but the emergency physicians responded more quickly than residents.

Third, the interruptions may not have been sufficiently powerful. However, they did approximate in kind and frequency the situation in the ED. Alternatively, it remains possible that the sample was too small to detect a clinically important effect, although, as indicated in the Results, we did have sufficient power to identify a 4% difference in accuracy.

Finally, there is evidence that with increasing experience, physicians rely more on nonanalytic System 1 than analytic System 2 reasoning.[16] According to some authors,[6–8] such reliance on cognitive shortcuts or heuristics is associated with increased error rates. In this study, we found no loss of diagnostic accuracy for the experts. Rather, the emergency physicians were faster, more accurate overall, and more efficient at dealing with interruptions than were the residents.

Conclusion

Dual-processing models of reasoning, as described by Kahneman,[9] suggest that slowing down and consciously engaging analytical (System 2) processes will help reduce the influence of cognitive biases and thereby improve diagnostic accuracy. Our results run counter to this hypothesis by demonstrating that increased resources devoted to thinking about the case—whether induced by instructions or absence of interruptions—do not improve accuracy. Future research must aim to understand the role of other factors that influence diagnostic accuracy.

Acknowledgments: Dr. Monteiro managed, collated, and analyzed all data in this study and takes responsibility for the integrity and accuracy of the analyses. The authors thank David Keane, PhD, a research associate at McMaster University who assisted with the development of the online format of the experiment and managed the storage and security of online data from emergency physicians.

Funding/Support: The study was supported by grants from the Medical Council of Canada. The funding organization was not directly involved in the preparation of this manuscript or in any part of the design or conduct of the study. The funding organization was not involved in data collection, management, analysis, or interpretation of the data and did not review this manuscript prior to submission.

Other disclosures: None reported.

Ethical approval: The study received approval from research ethics boards at McMaster University (approval 10-409, September 18, 2012), McGill University, the University of Toronto, Queen's University, the University of Ottawa, and the University of Washington.

Dr. Monteiro is assistant professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

Dr. Sherbino is associate professor, Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

Dr. Ilgen is assistant professor, Division of Emergency Medicine, School of Medicine, University of Washington, Seattle, Washington.

Dr. Dore is assistant professor, Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

Dr. Wood is associate professor, Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

Dr. Young is assistant professor, Department of Medicine, McGill University, Montreal, Quebec, Canada.

Dr. Bandiera is director of postgraduate programs, Department of Medicine and Postgraduate Medical Education, University of Toronto, Toronto, Ontario, Canada.

Dr. Blouin is director of faculty development for health sciences, Department of Emergency Medicine, Queen's University, Kingston, Ontario, Canada.

Dr. Gaissmaier is professor, Department of Psychology, University of Konstanz, Konstanz, Germany.

Dr. Norman is professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

Ms. Howey is a research associate, Program for Educational Research and Development, McMaster University, Hamilton, Ontario, Canada.

References

1 Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf. 2013;22(suppl 2):ii21–ii27.
2 Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78:775–780.
3 Chisholm CD, Weaver CS, Whenmouth L, Giles B. A task analysis of emergency physician activities in academic and community settings. Ann Emerg Med. 2011;58:117–122.
4 Chisholm CD, Dornfeld AM, Nelson DR, Cordell WH. Work interrupted: A comparison of workplace interruptions in emergency departments and primary care offices. Ann Emerg Med. 2001;38:146–151.
5 Westbrook JI, Coiera E, Dunsmuir WT, et al. The impact of interruptions on clinical task completion. Qual Saf Health Care. 2010;19:284–289.
6 Croskerry P. Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. Adv Health Sci Educ. 2009;14:27–35.
7 Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. Ann Intern Med. 2005;142:115–120.
8 Elstein AS, Schwartz A, Schwarz A. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. BMJ. 2002;324:729–732.
9 Kahneman D. Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux; 2011.
10 Evans JS. Dual-processing accounts of reasoning, judgment and social cognition. Annu Rev Psychol. 2008;59:255–278.
11 Stanovich KE, West RF. On the relative independence of thinking biases and cognitive ability. J Pers Soc Psychol. 2008;94:672–695.
12 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165:1493–1499.
13 Zwaan L, Thijs L, Timmermans DR, Wagner C. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. Acad Med. 2012;87:149–156.
14 Graber ML, Kissam S, Payne VL, et al. Cognitive interventions to reduce diagnostic error: A narrative review. BMJ Qual Saf. 2012;21:535–557.
15 Sherbino J, Yip S, Dore KL, Siu E, Norman GR. The effectiveness of cognitive forcing strategies to decrease diagnostic error: An exploratory study. Teach Learn Med. 2011;23:78–84.
16 Sherbino J, Kulasegaram K, Howey E, Norman G. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: A controlled trial. CJEM. 2014;16:34–40.
17 Eva KW, Link CL, Lutfey KE, McKinlay JB. Swapping horses midstream: Factors related to physicians' changing their minds about a diagnosis. Acad Med. 2010;85:1112–1117.
18 Eva KW. The aging physician: Changes in cognitive processing and their impact on medical practice. Acad Med. 2002;77(10 suppl):S1–S6.
19 Lloyd FJ, Reyna VF. Clinical gist and medical education: Connecting the dots. JAMA. 2009;302:1332–1333.
20 Reyna V. How people make decisions that involve risk: A dual systems approach. Current Dir Psychol Sci. 2004;30:60–66.
21 Sherbino J, Dore KL, Wood TJ, et al. The relationship between response time and diagnostic accuracy. Acad Med. 2012;87:785–791.
22 Norman G, Sherbino J, Dore K, et al. The etiology of diagnostic errors: A controlled trial of system 1 versus system 2 reasoning. Acad Med. 2014;89:277–284.
23 Ilgen JS, Bowen JL, McIntyre LA, et al. Comparing diagnostic performance and the utility of clinical vignette-based assessment under testing conditions designed to encourage either automatic or analytic thought. Acad Med. 2013;88:1545–1551.
24 Monsell S. Task switching. Trends Cogn Sci. 2003;7:134–140.
25 Pashler H. Dual-task interference in simple tasks: Data and theory. Psychol Bull. 1994;116:220–244.
26 Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ. 2008;42:468–475.
27 La Rochelle JS, Durning SJ, Pangaro LN, Artino AR, van der Vleuten CP, Schuwirth L. Authenticity of instruction and student performance: A prospective randomised trial. Med Educ. 2011;45:807–817.
28 Hatala R, Issenberg SB, Kassen B, Cole G, Bacchus CM, Scalese RJ. Assessing cardiac physical examination skills using simulation technology and real patients: A comparison study. Med Educ. 2008;42:628–636.