Information Distortion in Physicians' Diagnostic Judgments

Olga Kostopoulou, PhD, J. Edward Russo, PhD, Greg Keenan, MSc, Brendan C. Delaney, MD, Abdel Douiri, PhD

Background: *Information distortion suggests that people change the evaluation of new information to support an emerging belief. The present study was designed to measure the extent to which physicians distort incoming medical information to support an emerging diagnosis.* ***Design:*** *Data were collected via an anonymous questionnaire. The experimental group (102 physicians) read 3 patient scenarios, each with 2 competing diagnoses. Physicians first read information that favored 1 of the 2 diagnoses (the "steer"). They then rated a series of neutral cues that favored neither diagnosis. At each cue presentation, respondents rated the extent to which cues favored either diagnosis and updated the strength of their diagnostic belief. After the neutral cues in the third scenario, respondents rated cues that opposed the initial steer. A control group (36 physicians) rated all the cues in random order and not within scenarios, thus providing unbiased baseline ratings for calculating distortion in the experimental group.* ***Results:*** *Distortion was statistically significant (*$P < 0.001$*) and was associated with the strength of belief in the leading diagnosis. Physicians with over 10 years in practice distorted less than their less experienced counterparts (*$\bar{x} = 1.04$ *v.* $\bar{x} = 1.78$*,* $P < 0.05$*). Having developed an initial diagnostic leaning consistent with the steer, 56% of physicians remained committed to it after receiving the conflicting cues. Distortion was strongly associated with commitment to the steer (odds ratio, 1.4; 95% confidence interval, 1.03–1.79;* $P = 0.03$*).* ***Limitations:*** *Physicians did not elicit information; therefore, the authors cannot estimate the size of distortion in tasks involving information search.* ***Conclusions:*** *Distortion could partly explain commitment of physicians to an early diagnosis. Both distortion and strength of initial diagnostic belief seem to decline after 10 years in family medicine.* ***Key words:*** *cognitive bias; diagnostic reasoning; information distortion; confidence; commitment; experience.* ***(Med Decis Making 2012;32: 831–839)***

People have been found to bias their interpretation of information to support a valued position[1] and consolidate a decision.[2] Furthermore, people have been shown to bias their interpretation of new information to support an *emerging* position during the decision-making process.[3,4] As one option emerges as the leader in overall preference or judgment, the evaluation of new information tends to be "distorted" toward supporting that option, before a final decision is declared. The implication of such predecisional distortion of information is that decision makers seem to be persuading themselves of the continued superiority of the leading option. This "emerging" version of the bias has been observed even when it is clearly undesirable and has demonstrably negative effects on the choice itself.[5] The bias is thought to be driven by the need for coherence, which dictates that judgments be consistent with the information on which they are based.[4,6]

Predecisional distortion of information has been found in auditors and salespersons,[7] prospective jurors,[8] professional boxing judges,[9] racetrack bettors[10] and the general population.[11] Wallsten found that physicians sometimes used irrelevant information about normal findings to reduce the strength of their diagnostic beliefs, when such information should not have been used at all.[12] However, his method was indirect, and he had access to only 6 physicians as study participants. Kostopoulou, Mousoulis, and Delaney[13] found that family physicians differed in the way that they evaluated

Received 11 November 2011 from King's College London, London, United Kingdom (OK, GK, BCD, AD), and Cornell University, Ithaca, New York (JER). Revision accepted for publication 15 February 2012.

Address correspondence to Olga Kostopoulou, PhD, King's College London, School of Medicine, Department of Primary Care and Public Health Sciences, Capital House, 42 Weston Street, London SE1 3QD, UK; e-mail: olga.kostopoulou@kcl.ac.uk.

DOI: 10.1177/0272989X12447241

information (as well as in the amount and type of information elicited), according to the diagnosis that they eventually gave, when faced with a realistic, computer-driven patient scenario. However, information evaluation by physicians was assessed on the basis of stimulated recall—that is, postdecisionally. The current study is the first to attempt to 1) use a direct method to measure distortion in physicians and to do so 2) during, rather than after, the diagnostic process.

There are several reasons why physicians might be less prone than other professional groups to the bias. First, diagnosis is a highly technical skill: a long period of medical education and training is devoted to teaching accurate diagnostic inference from the assembled medical evidence. Second, the evidence gathered for diagnosis has an objective diagnostic value in terms of which diagnosis it favors and by how much, which is derived from research. This may not be the case in other domains of professional judgment. Third, diagnostic errors can have serious implications for patients in terms of harm or death and can lead to litigation. Therefore, clinical reasoning ought to be driven by accuracy rather than coherence goals.[1] Given physicians' extensive training, pressure for evidence-based decisions, and accountability for good patient outcomes, we expected them to exhibit less information distortion than that of other professionals.

METHODS

Materials

Three patient scenarios were constructed using an extensive evidence review conducted for a previous study.[14] Each scenario had 2 competing diagnoses that are common in family medicine:

1. Chest pain scenario—musculoskeletal diagnosis v. cardiac diagnosis.
2. Dyspnea scenario—chronic obstructive pulmonary disease v. left ventricular failure.
3. Fatigue scenario—diabetes v. depression.

The information for each scenario (''cues'') consisted of symptoms and signs that could be obtained through history and physical examination. Two scenarios contained results of investigations (resting electrocardiogram and random blood glucose). Some cues were ''diagnostic''—that is, supporting 1 of the 2 competing diagnoses—while others were ''neutral'' and supported neither diagnosis. Neutral cues were relevant to the scenarios; specifically, they were cues that a physician deciding between the 2 diagnoses could well have requested. However, their values would not help to differentiate between the 2 diagnoses.

Each scenario contained 4 parts: a presenting problem and 3 sets of medical cues. The presenting problem consisted of a brief patient description (name, age, and sex) and the main symptom (chest pain, dyspnea, or fatigue). One set of cues contained neutral cues only, and the other 2 sets contained diagnostic cues only. One set of diagnostic cues favored 1 diagnosis, and the other set favored the opposing diagnosis.

Participants

In sum, 138 family physicians in the United Kingdom took part in the study: 67% women, 25% trainees, 25–68 years old ($\bar{x}$ = 40, s = 11), and 0–39 years of experience in clinical practice ($\bar{x}$ = 11, s = 11; *0* denotes trainees). Of the 138 participating physicians, 102 saw the 3 scenarios, and 36 served as controls, according to an established method of having a separate control group for the measurement of distortion (see Procedure section).[15,16]

Data were collected via an anonymous printed questionnaire distributed at conferences and educational meetings for family physicians. At conferences, respondents were compensated with a £10 gift certificate for a book. At educational meetings, data were collected from the audience at the start of the meeting, and feedback about the group results was offered at the end. The study received ethical approval by a research ethics committee in London (Wandsworth REC, approval No. 10/H0803/36).

Procedure

The experimental group. The physicians in the experimental group read the 3 patient scenarios, presented in a counterbalanced order. Each scenario started with the patient's presenting problem, which was immediately followed by a set of diagnostic cues (step 1 in Figure 1). This initial set of diagnostic cues intended to create a diagnostic leaning: ''the steer.'' Half the participating physicians received the steer favoring one diagnosis, and the other half received the steer favoring the competing diagnosis. Physicians were then asked to rate the likelihood of the cause of the patient's problem, assuming that there were only 2 mutually exclusive diagnoses. They were given a 21-point scale, with verbal anchors at 10 = *diagnosis A more likely*, 0 = *equally likely*, and 10 = *diagnosis B more likely.* This initial rating of diagnostic belief identified whether the steer was successful in creating the diagnostic leaning intended.

Step 1: Presenting problem and steer

Patricia Aldridge, 52, presents with fatigue for the past 2 months.

Patricia says: "I get up a couple of times a night to go to the loo but it's probably because I've been drinking more fruit juice than usual." Her BMI is 30. You measure her random blood glucose level, which is 7.5 mmol/l. *(Steer favoring diabetes)*

Assuming that there are only 2 mutually exclusive diagnoses, please rate your estimated likelihood of the cause of Patricia's fatigue. *(1st rating of diagnostic belief)*

⬇

Step 2: Neutral cues presented sequentially, in random order, one cue with its ratings per page.

- Patricia says: "I don't feel like doing very much in the evenings other than watching TV."

How much does this item of information favour either diagnosis?

Given all the information provided so far, please rate your estimated likelihood of the cause of Patricia's fatigue.

- Patricia says: "As far as I know, the only person with diabetes in my family is one of my cousins."

How much does this item of information favour either diagnosis?

Given all the information provided so far, please rate your estimated likelihood of the cause of Patricia's fatigue.

- Patricia says: "I've lost about half a stone over the last 3 months."

How much does this item of information favour either diagnosis?

Given all the information provided so far, please rate your estimated likelihood of the cause of Patricia's fatigue.

⬇

Step 3 (only in 3rd scenario in the sequence): Conflicting cues

- Patricia says: "Nothing seems like much fun at the moment, even seeing my friends."

How much does this item of information favour either diagnosis?

Given all the information provided so far, please rate your estimated likelihood of the cause of Patricia's fatigue.

- Patricia says: "I've been generally feeling a bit low about everything."

How much does this item of information favour either diagnosis?

Given all the information provided so far, please rate your estimated likelihood of the cause of Patricia's fatigue.

- Patricia says: "I find concentrating quite difficult and often my mind wanders."

How much does this item of information favour either diagnosis?

Given all the information provided so far, please rate your estimated likelihood of the cause of Patricia's fatigue.

Figure 1 Steps in cue presentation and elicitation of ratings, using the fatigue scenario as an example.

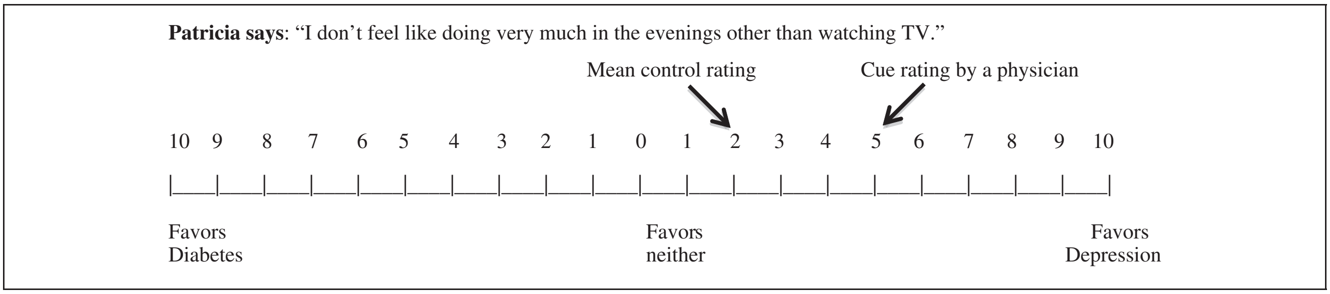

Figure 2 Calculation of information distortion of an individual cue rating by a physician who currently thinks that depression is more likely (5 – 2 = +3). If the physician thought that diabetes was more likely before seeing this cue, distortion would have been –3.

The neutral cues were then presented to the physicians, each on a separate page and in random order (step 2 in Figure 1). They were 4 or 5, depending on the scenario. Below each cue, there were 2 similar 21-point scales. The first asked participants to rate how much the cue favored one diagnosis over the other (10 = *favors diagnosis A*, 0 = *favors neither*, 10 = *favors diagnosis B*). The second asked them to rate the likelihood of the cause of the patient's problem, given all the information up to that point.

The first 2 scenarios ended with these neutral cues. In the third scenario only, the neutral cues were followed by a set of diagnostic cues that opposed the steer (step 3 in Figure 1). These too were presented on separate pages and in random order. This final set of cues made the total information for the third scenario neutral, favoring neither diagnosis. By confronting physicians with conflicting information late in the diagnostic process, we tested whether distortion would contribute to preserving the initial diagnostic leaning. Note that conflicting cues were presented at the end of the third scenario only, to avoid respondents detecting a pattern and expecting conflicting information at the end of every scenario.

The control group. The 36 physicians in the control group provided unbiased baseline ratings of all the available cues, neutral and diagnostic, presented in a random order. The patient's age, sex, and presenting problem (chest pain, dyspnea, or fatigue) were given for each cue. To prevent the respondents from spontaneously forming a coherent impression of a patient and distort cues thereafter, 3 precautions were taken: First, each cue referred to a different patient, denoted by a letter—for example, patient X, patient D. Second, age varied by up to 4 years, above or below the age in the corresponding scenario (the sex always matched that of the patient in the scenario). It was deemed that such small variations in age were not clinically significant. Finally, 3 decoy cues relating to 3 completely different pairs of diagnoses were interspersed in the questionnaire.

The 36 physicians in the control group provided ratings of each cue—that is, of how much the cue favored 1 of 2 diagnoses, using the same 21-point scale with the same diagnoses at either end, as seen by the experimental group in the scenarios. They were never asked to estimate diagnostic likelihood. Since they had no opportunity to build a working diagnosis, their mean cue ratings served as the "unbiased" diagnosticity of each cue, relative to which the distortion of the experimental group was calculated.

Computation of distortion. The distortion of each cue's evaluation by the experimental group was calculated as follows (Figure 2).[7,15] First, the absolute difference was computed between each individual cue rating and its mean control rating. Then, this difference was signed positively or negatively depending on whether it pointed toward or away from the leading diagnosis (i.e., the diagnosis indicated as being more likely immediately before the cue was seen). This measure of distortion has been termed *leader-signed distortion.*[16] If the immediately previous rating on the diagnostic likelihood scale was 0—that is, the diagnoses were deemed equally likely—then distortion could not be computed, as there was no currently leading diagnosis to distort toward.

ANALYSES

Since each physician provided several observations, we expected that these observations were not independent. Therefore, we report the mean distortion after it was averaged per physician, and we use these per-physician means in subsequent analyses. A 1-sample *t* test was used to determine whether distortion was significantly different from 0. A paired *t* test was used to test for differences in distortion between neutral and diagnostic cues. One-way

Table 1 Final Diagnoses in the Third Scenario and Mean Distortion by Steer

Third Scenario	Steer	Steer Chosen as the Final Diagnosis % (*n*)	Commitment to Steer % (*n*)	Mean Distortion (All Scenarios)
Chest pain (*n* = 32)	MSK	53 (8/15)	73 (8/11)	2.42
	Cardiac	53 (9/17)	47 (8/17)	0.64
Dyspnea (*n* = 35)	COPD	12 (2/17)	20 (2/10)	1.16
	LVF	78 (14/18)	93 (14/15)	2.30
Fatigue (*n* = 34)	Diabetes	29 (5/17)	31 (5/16)	0.43
	Depression	65 (11/17)	65 (11/17)	1.59
Total (*n* = 101)		49 (49/101)	56 (48/86)	

Note: "Steer chosen as the final diagnosis": the denominator is the number of times that a diagnosis was the scenario steer. "Commitment to steer": the denominator is the number of times that physicians accepted the steer. MSK, musculoskeletal; COPD, chronic obstructive pulmonary disease; LVF, left ventricular failure.

analysis of variance was used to test for differences in distortion between steers.

Physician experience was not normally distributed, due to 33% of the experimental group being trainees (0 years in family medicine). Following the "10-year rule" in the expertise literature,[17,18] we partitioned the sample into physicians with > 10 years in practice ($n = 40$) and physicians with ≤ 10 years in practice ($n = 62$). We employed independent-samples t tests to measure differences in distortion between the 2 groups. To ensure that any relationship with distortion did not depend on the specific cutoff point chosen to dichotomize the experience variable, we performed the same comparisons with 2 different cutoff points, above and below 10 years (8 and 12 years). Finally, Spearman rho was used to measure the correlation between experience (continuous variable) and distortion.

Previous research has found a linear relation between how strongly a decision maker is leaning toward one option and how much the subsequent cue is distorted to support that leaning.[15] That is, confidence in the current leader drives the size of the distortion of incoming information ("leader-driven distortion").[16] To test whether information distortion increased with increasing diagnostic belief, a multilevel linear random coefficients model was built, with diagnostic belief as the predictor, distortion as the dependent variable, and physician as the grouping variable.[16,19] The model paired the distortion of each cue with the immediately preceding diagnostic belief rating. Absolute values were used for the diagnostic belief ratings because the direction of the relationship—that is, whether distortion favored the currently leading diagnosis or not—was already contained in the sign of the distortion values.

We also investigated physician commitment to the steer, which we defined as leaning toward the steer from start to end. This required that, having accepted the steer, physicians never switched diagnoses—that is, never crossed to the other side of the diagnostic likelihood scale and never showed indifference (diagnoses equally likely) at any point during the third scenario. Commitment to a diagnosis favored early in the process has been linked to diagnostic error.[20–23]

Differences in physician commitment by experience group and steer were examined with χ^2 tests. Finally, logistic regression was employed to measure the influence of mean distortion on physician commitment.

RESULTS

Mean distortion was 1.5 ($s = 1.57$), statistically significant ($t = 9.56$, $df = 101$, $P < 0.001$), and equivalent to 7.6% of the scale. There were no differences in distortion between neutral and diagnostic cues ($\overline{x} = 1.51$ for neutral, $\overline{x} = 1.47$ for diagnostic, $P = 0.895$). However, we cannot firmly conclude that both types of cues were distorted equally, because of the unequal number of observations: each physician saw 14 neutral cues across scenarios and only 3 diagnostic cues at the end of the third scenario. Distortion differed significantly by steer, $F_{(5,298)} = 8.048$, $P < 0.001$: in each scenario, one steer was associated with higher distortion than the other steer (Table 1).

As expected, a relationship was found between strength of belief in the currently leading diagnosis and distortion of the next cue. The mean slope of this relationship in the multilevel model was 0.3 (95% confidence interval [CI]: 0.21–0.38, $P < 0.001$), indicating that for a 1-unit increase in diagnostic belief, distortion of the next cue increased by 0.30 units on average.

Experience-Related Differences

The highly experienced physicians (> 10 years in practice) exhibited significantly lower distortion than their less experienced counterparts ($\bar{x}$ = 1.04 v. $\bar{x}$ = 1.78, respectively, t = 2.37, df = 100, P = 0.020). When the experience variable was dichotomized at 2 different cutoffs (8 and 12 years), group differences remained significant (P = 0.017 and P = 0.024, respectively). A ranked correlation between experience and distortion was negative, though significance was borderline (Spearman rho = –0.17, P = 0.081, 2-tailed). The highly experienced physicians also gave significantly lower ratings to the initial diagnosis than the less experienced physicians ($\bar{x}$ = 5.10 v. $\bar{x}$ = 5.95, t = 2.10, df = 100, P = 0.038). When mean distortion was regressed on experience (binary variable) controlling for the strength of initial diagnostic belief, both predictors were found to exert independent influence on distortion (coefficient, –0.61, P = 0.053, for experience; coefficient, 0.15, P = 0.055, for belief), suggesting that, after 10 years of practicing family medicine, distortion decreases on average by 0.61 units.

To identify any physicians who were genuinely distortion-free, we applied 2 criteria. The first criterion was very low mean distortion (< 0.100). As this was based on only 17 observations per physician, a low mean could have occurred by chance. Thus, we applied a second criterion: each physician's slope, provided by the multilevel model. An equally low slope (< 0.100) ensured no systematic relation between strength of belief and distortion. We identified 15 physicians who had both mean distortion and slope < 0.100. Nine of these (60%) were highly experienced, compared to 39% of experienced physicians in the whole sample.

Commitment to Steer

Of the 102 physicians in the experimental group, one did not complete the third scenario. This resulted in a total of 305 responses across scenarios for further analysis (101 in the third scenario). The steer manipulation was successful in creating an initial leaning toward the intended diagnoses in 87% of the responses (265/305). Of the 40 occasions where the steer did not initially persuade a physician, most (29/40, 72.5%) suggested indifference (i.e., diagnoses equally likely), reflecting perhaps a wait-and-see approach in the knowledge that more information was coming. On 11 of 40 occasions (27.5%), the opposing diagnosis was considered more likely.

The steer was chosen as the final diagnosis in the third scenario by 49% of physicians (49/101). This proportion varied widely between the 2 opposing steers in 2 scenarios: in the dyspnea scenario, the chronic obstructive pulmonary disease steer was chosen by only 12% of physicians as the final diagnosis, while the competing steer (left ventricular failure) was chosen by 78% of the physicians. In the fatigue scenario, the diabetes steer was chosen by a third of the physicians, and the competing depression steer, by two thirds. Both these highly chosen steers (left ventricular failure and depression) were linked to higher mean distortions than their competing steers (Table 1).

In the third scenario, the steer was accepted as the initial diagnosis by 86 of the 101 physicians. Of these, 48 showed unwavering commitment to the steer from start to end (56%, Table 1). Commitment differed by steer (χ^2 = 20.03, df = 5, P = 0.001). The most-committed-to steers (musculoskeletal, left ventricular failure, and depression) were associated with higher mean distortions than their opposing steers. Commitment did not differ by physician experience (P = 0.47) but did differ by sex. In further exploratory analyses, female physicians were found to remain committed to the steer more frequently than male physicians (57% v. 34%, χ^2 = 4.95, df = 1, P = 0.026). It should be noted that there were no sex-related differences in distortion (P = 0.82), in the rate of accepting the steer as the initial diagnosis (P = 0.39), or in the strength of the initial diagnostic belief (P = 0.76). Furthermore, male and female physicians were represented equally frequently in the low and high experience groups (P = 0.15).

When commitment was regressed on mean distortion in the third scenario, a significant effect was found: the odds ratio for distortion was 1.4 (95% CI: 1.10–1.68, P = 0.005), suggesting a 40% increase in the odds of committing to the steer with a 1-unit increase in distortion. When the model was adjusted for steer (6-level categorical variable), prediction improved substantially: the area under the receiver operating characteristic curve increased from 0.69 (95% CI: 0.58–0.79) to 0.80 (95% CI: 0.72–0.89), while distortion remained significant: odds ratio = 1.4, 95% CI: 1.03–1.79 (P = 0.03).

In the third scenario, the total information was designed to be neutral—that is, to favor neither diagnosis. When the mean control ratings were averaged across cues in the chest pain scenario, the total information slightly favored the cardiac diagnosis (1.51 on the 0–10 scale). In the dyspnea scenario, the total information was neutral (0.25 on the 0–10 scale). In the fatigue scenario, the total information slightly favored depression (1.23 on the 0–10 scale). Physicians should therefore have rated the diagnoses as equally likely or indicated low diagnostic belief at the end of the scenario. Only 6 physicians considered the diagnoses equally likely. The mean final diagnostic belief across the sample was 4.60, which could be interpreted as substantial belief in either of the competing diagnoses on the 0–10 scale.

DISCUSSION

This is the first study to directly measure the size of predecisional information distortion in physicians, consolidating evidence from earlier studies that physicians do distort information to support an emerging diagnosis.[12,13] In agreement with previous research, distortion was associated with the strength of belief in the leading diagnosis. Furthermore, distortion was associated with physicians' tendency to stick with their initial diagnosis, even when disconfirming information was presented to them later in the process.[9]

Distortion was equivalent to 7.6% of the scale. This can be compared to other published studies of distortion that used similar methods. For example, information distortion by auditors was measured in a scenario that required them to decide which of 2 publicly traded companies to visit urgently for an audit.[7] Their distortion was equivalent to 7.8% of the scale. Higher values for mean distortion are usually reported in the literature—for example, 10% for representatives of a pharmaceutical company deciding which physician to visit for a sales call,[7] 15.5% for prospective jurors judging evidence from a mock trial,[8] and 11.1% for students choosing consumer products.[15] Auditors are, however, the most comparable group of professionals studied to physicians, and their distortion was of similar size. Being an auditor usually necessitates a university degree and additional certification, while the technical knowledge acquired during education and training is directly used in auditing work.

Within each scenario, both distortion size and final diagnostic choice differed by steer. The source of these differences is not apparent. They could be attributed either to different base rates of the competing diagnoses or to the total net information not being perfectly equivocal. Neither of these plausible hypotheses is satisfactory across scenarios. Both dyspnea and fatigue—scenarios with the largest differences between steers—were designed so that the competing diagnoses had similar base rates. In contrast, in the chest pain scenario, where there is an objective difference in the base rates (musculoskeletal causes of chest pain are more common than cardiac), each steer was chosen as the final diagnosis as equally frequent as the other.

The other possible explanation is that the total information in the third scenario was not equivocal but favored one of the competing diagnoses more than the other. When the mean ratings of the control group were averaged across all cues in each scenario, we found this to be the case in 2 scenarios: the cardiac diagnosis was slightly favored over musculoskeletal in the chest pain scenario, and depression was slightly favored over diabetes in the fatigue scenario. One would therefore expect these 2 diagnoses to have been chosen more frequently than their competitors. However, this happened only with depression. Furthermore, the total information was neutral in the scenario with the largest differences in the final diagnostic choice (dyspnea). Therefore, neither a difference in the base rates nor the possibility of the total information not being equivocal can explain the differences across scenarios. These differences speak of ''case specificity,'' a phenomenon that has been observed in previous studies of general physicians (primary care or internal medicine) that possibly reflects differences in the physicians' knowledge base. General physicians respond differently to cases from different areas of medicine; performance on one case does not predict performance on another.[14,22,24] This emphasizes the need to employ more than 1 patient case when studying general physicians' behavior.

There are several positive messages for clinicians in this study. First, mean distortion across participants was smaller than most other values of distortion reported in the literature. This could reflect greater objectivity in clinical judgments in comparison to other professionals. Second, even when distortion was very large, commitment to the diagnosis was less than 100%. Third, highly experienced physicians distorted less than their less experienced counterparts, and some physicians seemed completely free of the bias. We should nevertheless acknowledge a potential demand effect: physicians knew that they were being studied. As a consequence, they may have been more careful in their evaluation of the cues than in a practice setting.

The literature on the relationship among experience, distortion, and confidence (''strength of diagnostic belief'' in this study) is not consistent. One study found a positive relationship among all 3 variables,[10] while another found that experience was linked to less biased judgments and higher confidence.[9] These studies compared complete novices with experienced participants, while only qualified doctors participated in our study. Therefore, results between studies are not directly comparable. It could be argued that more experienced doctors should be more confident and hence distort more. However, it is also possible that confidence may increase early on but decline in later years, possibly as a result of recognizing uncertainty in diagnosis and having made mistakes. The relationship between experience and distortion requires further study. If more experienced clinicians evaluate information in a less biased way, it is important to understand the factors that underlie this, such as frequency of feedback, feelings of accountability, or even past errors.

The distortion of information to support a currently held belief could be considered as another manifestation of confirmation bias, a general phenomenon in human judgment that encompasses a number of biases in the search for and interpretation of information.[25] We would argue, however, that information distortion is not just another manifestation of confirmation bias but a way to create and maintain coherent judgments.[4,6] This represents an inherent need of the decision maker.

Study Limitations and Future Research

Study limitations concern mainly the task used to measure distortion. First, physicians were presented with information and did not have to search for it. Information distortion has never been measured in situations of active information search, which could increase physicians' commitment to the focal hypothesis and, in turn, increase distortion. Conversely, if physicians try to disconfirm a diagnosis by searching specifically for opposing evidence, they may distort information less than when encountering opposing information unexpectedly. The present study does not provide evidence for the effects of information search on distortion or the effects of distortion on the search itself.

During their everyday clinical practice, physicians do not explicitly evaluate each piece of evidence that they encounter, but they had to do this in the study, to enable direct measurement of distortion. It is possible that this changed their ratings of diagnostic belief, though it would be difficult to predict the direction of such change. Furthermore, physicians had to update their diagnostic belief at each cue presentation. This in itself could have influenced their final diagnostic belief. Hogarth and Einhorn's belief adjustment model compares end-of-sequence versus step-by-step judgments in belief updating.[26] It predicts a primacy effect for end-of-sequence judgments and a recency effect for step-by-step judgments. Therefore, if respondents rated their diagnostic belief only at the end, we would expect their estimates to be higher (primacy effect, because the steer was shown first) than if they updated their belief at each cue presentation (recency effect, because neutral cues were shown last). In scenarios with conflicting cues presented at the end, the model's predictions are less clear-cut. When information is inconsistent, respondents may have to update their judgments in a step-by-step fashion, even when their explicit judgments of probability are required only at the end of the sequence. In this case, the mode of response should not make a difference in the final judgment.

We steered physicians toward a diagnosis from the start, to ensure that a diagnostic leaning would develop in most of them. This does not always happen in clinical consultations; that is, physicians may not encounter diagnostic information from the start and may not immediately develop an initial diagnostic leaning. A future study could adopt a design where clinicians are left to develop a diagnostic leaning over a series of neutral cues, which would allow us to examine whether a diagnosis can emerge and be singled out with equivocal or weakly diagnostic information and whether confidence in this diagnosis can increase with such information.[15,27]

REFERENCES

1. Kunda Z. The case for motivated reasoning. Psychol Bull. 1990;108(3):480–98.

2. Svenson O. Values, affect and process in human decision making: a differentiation and consolidation theory perspective. In: Schneider SL, Shanteau J, eds. Emerging Perspectives on Judgment and Decision Research. Cambridge: Cambridge University Press; 2003. p 287–326.

3. Brownstein AL. Biased predecision processing. Psychol Bull. 2003;129(4):545–68.

4. Russo JE, Carlson KA, Meloy MG, Yong K. The goal of consistency as a cause of information distortion. J Exp Psychol Gen. 2008;137(3):456–70.

5. Russo JE, Carlson KA, Meloy MG. Choosing an inferior alternative. Psychol Sci. 2006;17(10):899–904.

6. Simon D, Snow CJ, Read SJ. The redux of cognitive consistency theories: evidence judgments by constraint satisfaction. J Pers Soc Psychol. 2004;86(6):814–37.

7. Russo JE, Meloy MG, Wilks TJ. Predecisional distortion of information by auditors and salespersons. Manag Sci. 2000;46(1):13–27.

8. Carlson KA, Russo JE. Biased interpretation of evidence by mock jurors. J Exp Psychol Appl. 2001;7(2):91–103.

9. Tyszka T, Wielochowski M. Must boxing verdicts be biased? J Behav Decis Mak. 1991;4:283–95.

10. Brownstein AL, Read SJ, Simon D. Bias at the racetrack: effects of individual expertise and task importance on predecision reevaluation of alternatives. Pers Soc Psychol Bull. 2004;30(7):891–904.

11. Levy AG, Hershey JC. Value-induced bias in medical decision making. Med Decis Making. 2008;28(2):269–76.

12. Wallsten TS. Physician and medical student bias in evaluating diagnostic information. Med Decis Making. 1981;1(2):145.

13. Kostopoulou O, Mousoulis C, Delaney BC. Information search and information distortion in the diagnosis of an ambiguous presentation. Judgm Decis Mak. 2009;4(5):408–18.

14. Kostopoulou O, Oudhoff J, Nath R, et al. Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in family medicine. Med Decis Making. 2008;28(5):668–80.

15. Russo JE, Meloy MG, Medvec VH. Predecisional distortion of product information. J Mark Res. 1998;35(4):438–52.

16. DeKay ML, Stone ER, Miller SA. Leader-driven distortion of probability and payoff information affects choices between risky prospects. J Behav Dec Making. 2011;24(4):394–411.

17. Simon HA, Chase WG. Skill in chess. Am Sci. 1973;61(4): 394–403.

18. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004;79(10):S70–80.

19. DeKay ML, Patiño-Echeverri D, Fischbeck PS. Distortion of probability and outcome information in risky decisions. Organ Behav Hum Decis Process. 2009;109(1):79–92.

20. Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning of randomly selected physicians in general medical practice. Clin Invest Med. 1982;5(1):49–55.

21. Dubeau CE, Voytovich AE, Rippey RM. Premature conclusions in the diagnosis of iron-deficiency anemia: cause and effect. Med Decis Making. 1986;6(3):169.

22. Sibbald M, Cavalcanti RB. The biasing effect of clinical history on physical examination diagnostic accuracy. Med Educ. 2011; 45(8):827–34.

23. Kostopoulou O, Devereaux-Walsh C, Delaney BC. Missing celiac disease in family medicine: the importance of hypothesis generation. Med Decis Making. 2009;29(3):282–90.

24. Elstein AS, Shulman LS, Sprafka SA. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge (MA): Harvard University Press; 1978.

25. Klayman J. Varieties of confirmation bias. In: Busemeyer J, Hastie R, Medin DL, eds. Decision Making From a Cognitive Perspective. San Diego (CA): Academic Press; 1995. p 385–418.

26. Hogarth RM, Einhorn HJ. Order effects in belief updating: the belief-adjustment model. Cogn Psychol. 1992;24(1):1–55.

27. Russo JE, Yong K. The distortion of information to support an emerging evaluation of risk. J Econometrics. 2011;162(1):132–9.