Judgment and Decision Making, Vol. 4, No. 5, August 2009, pp. 408–418

Information search and information distortion in the diagnosis of an ambiguous presentation

Olga Kostopoulou*[1], Christos Mousoulis[2], and Brendan Delaney[1]
[1] Division of Health and Social Care Research, King's College London
[2] College of Medicine and Dental Sciences, University of Birmingham

Abstract

Physicians often encounter diagnostic problems with ambiguous and conflicting features. What are they likely to do in such situations? We presented a diagnostic scenario to 84 family physicians and traced their information gathering, diagnoses and management. The scenario contained an ambiguous feature, while the other features supported either a cardiac or a musculoskeletal diagnosis. Due to the risk of death, the cardiac diagnosis should be considered and managed appropriately. Forty-seven participants (56%) gave only a musculoskeletal diagnosis and 45 of them managed the patient inappropriately (sent him home with painkillers). They elicited less information and spent less time on the scenario than those who diagnosed a cardiac cause. No feedback was provided to participants. Stimulated recall with 52 of the physicians revealed differences in the way that the same information was interpreted as a function of the final diagnosis. The musculoskeletal group denigrated important cues, making them coherent with their representation of a pulled muscle, whilst the cardiac group saw them as evidence for a cardiac problem. Most physicians indicated that they were fairly or very certain about their diagnosis. The observed behaviours can be described as coherence-based reasoning, whereby an emerging judgment influences the evaluation of incoming information, so that confident judgments can be achieved even with ambiguous, uncertain and conflicting information. The role of coherence-based reasoning in medical diagnosis and diagnostic error needs to be systematically examined.

Keywords: coherence-based reasoning, cognitive consistency, stimulated recall, diagnostic error, ambiguous information

1 Introduction

The diagnosis of chest pain in primary care is notoriously difficult. The cause of the pain is usually not life threatening. However, conditions such as acute coronary syndrome and pulmonary embolism may be fatal if unrecognised and not promptly treated. There is evidence that myocardial infarction (MI) may be missed in primary care (Kentsch et al., 2002; Sequist, Marshall, Lampert, Buechler, & Lee, 2006). A missed MI is the commonest cause of negligence claims against primary care physicians in the USA and relates to more deaths than other conditions in claims against family physicians in the UK (Esmail et al., 2004; Phillips et al., 2004).

Epidemiological studies have found that atypical presentations as well as patients attributing their symptoms to a non-cardiac cause are significant factors in missed diagnoses of MI (Bleeker et al., 1995; Bouma et al., 1999; Kentsch et al., 2002; Zarling, Sexton, & Milnor, 1983). We studied the diagnosis of atypical chest pain using a simulated patient who attributes his pain to a pulled muscle. The scenario, described in detail later, included conflicting features, i.e., some features pointing to a cardiac diagnosis and others pointing to a musculoskeletal diagnosis. Accounts of coherence-based reasoning describe how people deal with such situations and what they do in order to resolve the conflict.

Faced with equivocal, i.e., non-diagnostic, information about two alternatives, people will spontaneously form a preference for one alternative and will distort information received subsequently so that it favors the preferred alternative (Holyoak & Simon, 1999; Russo, Meloy, & Medvec, 1998). They may even distort information that clearly favors one alternative, in order to preserve an earlier preference for the other alternative, though distortion generally declines as diagnosticity of information increases (Russo et al., 1998, study 2). If preferences or judgments change as a result of new incoming information, distortion continues to operate, now in favor of the new preference or judgment (Russo et al., 1998; Simon,

*We gratefully acknowledge the contribution of Craig Munro, Jurriaan Oudhoff and Radhika Nath in building the scenario and collecting the data, as part of a larger study funded by the Department of Health, UK (grant PS/027). The coding, analyses, and interpretation of the data presented here were done by the authors with funding from the NIHR School for Primary Care Research, UK. Correspondence to: Olga Kostopoulou, Division of Health and Social Care Research, Kings College London, 7th Floor, Capital House, 42, Weston Street, London SE1 3QD, UK. E-mail: olga.kostopoulou@kcl.ac.uk.

Snow, & Read, 2004). It seems that distortion of information is a psychological necessity that helps individuals form and maintain coherent judgments.

Parallels with medical diagnosis can be drawn from studies of legal decision making. Simon and colleagues gave participants a legal case with conflicting information, i.e., where each position was supported by equal amounts of probabilistic evidence, and asked them to render a verdict and indicate their confidence in it (Simon et al., 2004). They pre-disposed participants to either a guilt verdict or an innocence verdict by presenting them with DNA evidence that either incriminated or exonerated the defendant (odds of 1 to 7 million of the DNA evidence being wrong). Before seeing the case, participants had been asked to rate the diagnosticity of each piece of evidence, presented to them in unrelated vignettes describing various social situations. After they read and rendered their verdict on the legal case, participants were asked to rate the diagnosticity of each piece of evidence again, this time presented to them in the context of the case. These new ratings were consistent with the verdict and significantly differed from the ones made earlier in the unrelated vignettes. It seems that whilst the evidence was initially perceived as truly equivocal and non-diagnostic, it was subsequently contextualised and became part of the mental model of the legal case. Its perceived diagnosticity thus increased in line with the verdict. The authors argue that a DNA match should bear no relationship to the other pieces of evidence in the case, such as the reliability of an eyewitness' identification. Nevertheless, these are now perceived as related and their evaluation shifts to support the final verdict — what the authors call "coherence shifts." This basic experimental procedure has been repeated with several variations in order to test whether different factors affected coherence-based reasoning. Coherence shifts were found even when participants were asked simply to memorise an ambiguous legal case. When they were unexpectedly asked to render a verdict and rate their agreement with the various arguments in the case, agreement had shifted from baseline, as a function of the final verdict (Simon, Pham, Le, & Holyoak, 2001). The largest shifts however were observed between baseline and the interim phase (after memorisation but before being asked to render a verdict or indicate a leaning towards a verdict). These findings suggest that coherence shifts happen pre-decisionally and are not simply the result of decision justification. Processing a complex legal case that contained conflicting arguments, with the intention either to make a decision later or receive further information or memorise it in order to communicate it to someone else, resulted in comparable coherence shifts (Simon et al., 2001). This was attributed to an intense attempt to comprehend the complex material and represent it mentally in a coherent way.

Pre-decisional distortion of information has been investigated mostly with college students and with tasks that required no prior experience or technical skill. Professionals or people with experience in a specific domain making judgments within their domain of expertise are rarely studied. In one such study, participants experienced in horse race betting evaluated task-related information (chances of horses winning a simulated race) in accordance with their final choice, either in anticipation of having to place a bet later or as a consequence of simply processing task information. They distorted information to a greater extent than participants without experience on betting on horses (Brownstein, Read, & Simon, 2004).

We wanted to investigate how physicians deal with difficult diagnostic problems that contain conflicting information, where diagnoses cannot be confirmed or excluded with certainty. The diagnosis of chest pain in primary care lends itself to this type of enquiry. The chest pain scenario was one of seven scenarios that we built for a larger study that investigated the relationship between experience, information search and diagnostic accuracy in difficult problems in Family Medicine (Kostopoulou et al., 2008). We found that, across scenarios, experience was not related to diagnostic accuracy, which was predicted only by the number of critical cues elicited, i.e., cues with diagnostic value for the relevant differential diagnoses. The study presented here analysed concurrent process-tracing data and stimulated recall data from one scenario, to identify why critical cues were requested, how they were interpreted and how the two competing diagnoses and the respective management decisions were arrived at.

2 Methods

2.1 Participants

Eighty-four family physicians participated: 21 residents, 21 family physicians with 1–3 years in practice and 42 family physicians with $\geq$10 years in practice. The most experienced group included 21 family physicians that trained residents and 21 that did not, matched for years in practice. The mean age was 31 years for the residents (range 26–42) and 44 for the family physicians (range 28–65), with 60% being males.

2.2 The scenario

A description of the patient and his presenting complaint and a brief description of all scenario cues available for elicitation are presented in the appendix. The scenario describes a 60-year old male, ex-smoker and without any significant past medical history. The patient complains

Table 1: Scenario features that could provide support for each of the two differential diagnoses.

Cardiac diagnosis	Musculoskeletal diagnosis
1. Risk factors: male, 60 years old, ex-smoker	1. Higher prevalence than cardiac (not given but presumed to be known)
2. Chest pain relieved by rest	
3. Worsening chest pain	2. Onset of pain during lifting
4. Family history of ischemic heart disease	3. Patient thinks that he pulled a muscle
5. Chest pain radiates to right arm (atypical feature, less likely to be known to participants)	4. Tenderness over the intercostal muscles
Pain on gardening (ambiguous feature)	

of 7 days of intermittent chest pain, first felt while lifting a washing machine. Onset of chest pain during lifting a heavy object suggests a musculoskeletal cause, a common cause of chest pain (1:150 consultations for 60-year old males) (McCormick, Fleming, & Charlton, 1995). The patient emphasizes this further: "I thought I'd pulled a muscle in my chest or something." Nevertheless, the patient's age, sex and smoking status (ex-smoker) constitute risk factors for heart disease. Further information search can reveal the following positive clinical features:

1. The pain comes on when the patient is working in the garden.[1]
2. The pain stops when the patient rests.
3. The pain is getting worse both in terms of duration and intensity.[2]
4. The pain is in the middle of the chest but yesterday was also going down his right arm.
5. The patient's father had an MI at age 65.
6. There is some tenderness over the intercostal muscles on the right side of the chest.

Features 2–5 suggest a cardiac cause, feature 6 suggests a musculoskeletal cause, while feature 1 is somewhat ambiguous. Chest pain that comes on with exertion is indicative of cardiac pain. This patient gets pain when he works in the garden but does no other exertional activity, due to his osteoarthritis: "The only exertion I ever do is working in the garden. My knees aren't up to anything more than that." Feature 4, radiation to the right arm, is atypical, with left-arm radiation being part of the "textbook" description of cardiac chest pain. Right-arm radiation has been found to be specific to MI, but radiation to the right arm alone is uncommon (Berger, Buclin, Haller, Van Melle, & Yersin, 1990). We expected that most family physicians would find this feature difficult to interpret.

The scenario was not designed to contain strictly equivocal information for the two competing diagnoses, but with one ambiguous feature and one atypical feature, evidence was relatively balanced (Table 1). Other relevant differential diagnoses, for example, gastroesophageal reflux disease, pulmonary embolism, or pneumonia could be excluded fairly easily during information gathering, so that the two remaining possibilities were cardiac and musculoskeletal. These are not mutually exclusive and the patient may well be suffering from both. The potential consequences of a missed cardiac diagnosis, however, are very serious. Given that a cardiac cause cannot be excluded in the scenario, it should be managed, even if the physician thinks that a pulled muscle is more likely.[3]

We reviewed the literature to identify how scenario features related to the differential diagnoses of musculoskeletal chest pain and acute coronary syndrome (we needed to specify a heart condition, since "angina," "heart disease" or "cardiac chest pain" are too general). We considered features with likelihood ratios >1.5 or <0.67 as "critical." We also conducted a separate web-based study of expert opinion, to identify any further cue-diagnosis relationships not explored in the literature (for details see Kostopoulou et al., 2008). Table 2 presents the combined results from the evidence review and the study of expert opinion. Three further features, all with negative or normal values (no sweating, pain not influenced by movement, normal resting ECG) could provide differential support for the two diagnoses, without significantly shifting the balance of evidence.

2.3 Procedure

A researcher administered the scenario to the 84 participating family physicians individually, via a laptop computer using a computer program specifically written for

[1]The equivalent term in the US would be "yard."

[2]This feature was intended to suggest Acute Coronary Syndrome (ACS), i.e., heart disease that is deteriorating, possibly leading to MI.

[3]If acute coronary syndrome is suspected, the patient should be sent to hospital urgently. If the physician does not think that an MI is imminent, he/she should refer the patient for further investigations, i.e., referral for an exercise tolerance electrocardiogram (ECG) or referral to a Rapid Access Chest Pain Clinic where the patient would be seen by a cardiologist and investigated within 2 weeks.

Table 2: The critical cues and their relationship, consistent (+) or inconsistent (-), with the two differential diagnoses. The shading of the cells indicates the provenance of the information.

Critical cues	Acute coronary syndrome	Musculo-skeletal pain
Chest pain on gardening	+	
Chest pain relieved by rest	+	
Worsening chest pain	+	
Chest pain radiates to right arm	+	
No sweating with chest pain	–	
Family history of ischemic heart disease	+	
Chest pain not influenced by movement		–
Tenderness over intercostal muscles	–	+
Normal resting ECG	–	+
Evidence review: cues with LR+ >1.5 or LR- <0.67		
Study of experts presented with scenario cues sequentially and asked to rate likelihood for each diagnosis after each cue: cues that shifted likelihood ratings significantly ($p \leq 0.1$) from the previous cue in a hierarchical linear model.		
Evidence review & study of experts		

the study. The laptop was connected to an additional monitor that was used to display information to the participants. Initially, participants read the patient description and presenting complaint, displayed on the screen and could request further information. The researcher selected the cue relevant to each information request from a drop-down menu, and displayed the answer on the participant's screen. The computer program recorded the information requests in order and the time that each cue was displayed to the participant. When the participant decided to end the consultation, the researcher recorded his or her diagnoses and management decisions. Participants who gave more than one diagnosis were asked to put them in order of likelihood. No feedback was provided at the time.

A subset of participants was subsequently interviewed about the scenario following a stimulated recall methodology.[4] Stimulated recall takes place after performance on a task and involves participants trying to recall their thoughts during task performance. Prompts are provided to aid recall usually in the form of a video recording of the earlier performance (Lyle, 2003). The methodology has been used to study the hypotheses considered and inferences made during the diagnosis of simulated patients and is an alternative approach to concurrent verbal reporting (think aloud), which may not be suitable in interactive decision situations (Barrows, Norman, Neufeld, & Feightner, 1982; Elstein, Shulman, & Sprafka, 1978). Being retrospective, stimulated recall data may suffer from hindsight bias, new inferences, and decision justification and may not reflect accurately the thinking processes during task performance. In fact, Ericsson and Simon (1993) argue that retrospective reports are more likely to be the product of "regeneration memory" (where participants mentally repeat the same task and actively generate a response anew) rather than of episodic memory for the earlier task performance. This does not mean that the data are not valid for answering specific research questions.

The aim of the stimulated recall in this study was to assess the participants' "states of knowledge," namely, the perceived links between the cues requested and the competing diagnoses, in order to understand how a diagnosis was arrived at. The researcher replayed to each physician the scenario and his/her information gathering process as recorded by the computer. For each cue requested, physicians were asked why they had requested it. For example, "you asked how long his chest pain lasts, why did you ask this?" Subsequently, the cue value was shown and they were asked what it had told them. For example, "the patient said that … for the past couple of days, it's been worse and going on for longer. What did this tell you?" It is entirely possible that physicians were actively constructing the answers to some of our questions a) if they did not consciously consider reasons for requesting cues and did not consciously interpret the information obtained during diagnosis and b) if they felt it easier to construct an answer than recall it. This does not

[4]Following diagnosis of the seven scenarios used in the larger study, all 84 physicians took part in stimulated recall in relation to the last three scenarios that each had diagnosed (order of scenario presentation differed).

invalidate their answers. We took care to give them no feedback on the appropriateness of their cue requests and the accuracy of their diagnosis and no new information that would have led them to perform new analyses and make new inferences. We therefore assume that the reasons and interpretations given were valid, reflecting their states of knowledge during diagnosis. This process was repeated for all cue requests in sequence.

At the end of this process, physicians were asked to give their final diagnosis. This was in order to check whether they had changed their mind having seen the scenario again, and having had the opportunity to re-appraise the information. Finally, they were asked to mark their certainty in their top diagnosis on a 100 mm visual analogue scale anchored at 0 ("not at all certain") and 100 ("absolutely certain"), with verbal labels ("somewhat certain," "fairly certain," "very certain"). The stimulated recall interviews were audio-recorded on minidisks and transcribed verbatim.

2.4 Data coding and analyses

Chi-square tests were used to test for experience-related differences in diagnosis (cardiac vs. musculoskeletal). Independent samples t-tests were used to test for differences in the amount of information gathered and time taken between cardiac and musculoskeletal diagnoses. Chi-square tests were used to test for differences in the elicitation of specific critical cues, with Bonferroni adjustment for multiple comparisons, and to test for the relationship between diagnosis and management.

Interpretations of critical cues from the stimulated recall transcripts were compared to the cue-diagnosis table (table 2) and were scored as either **correct interpretations** (if the cue-diagnosis link was interpreted as in the table); **misinterpretations** (if the cue-diagnosis link was interpreted as opposite to that in the table); **overinterpretations** (if cue-diagnosis links were perceived that were absent from the table); or **no information gained** (if the physician could not differentiate between cardiac and musculoskeletal causes on the basis of a cue, whilst the cue was differentiating in the table). Two raters (OK and CM) read and coded all transcripts independently. Inter-rater agreement was measured using the Kappa index. Differences in critical cue interpretations were tested with chi-square tests. Certainty ratings were compared using independent samples t-tests and one-way ANOVA.

3 Results

Overall, 37 family physicians (44%) diagnosed a cardiac cause (angina or acute coronary syndrome) either as a differential diagnosis or as the only diagnosis (the "cardiac group"). The remaining 47 gave only a musculoskeletal diagnosis (the "musculoskeletal group"). Management depended on diagnosis (χ^2=61.44, df=1, p=0.0001). Two physicians from the musculoskeletal group managed appropriately for the possibility of a cardiac problem, therefore, they were added to the cardiac group for all subsequent analyses. The 45 remaining physicians in the musculoskeletal group sent the patient home with a prescription of anti-inflammatory medication. Thirty-three did not arrange for a follow up (33/45, 73%) and 20 of them asked the patient to come back only if his pain did not improve. Twelve arranged for a follow up (12/45, 27%) and the interval ranged from 2 days to 6 weeks, with most choosing an interval of 1–2 weeks. Four of those also advised the patient to call for an ambulance if his pain got worse or his symptoms changed.

No differences were found in diagnosis between the three experience groups (p=0.52) in keeping with the findings of the larger study. The cardiac group elicited more cues overall (means 20 vs. 16, t=2.96, df=82, p=0.004), both critical (p=0.025) and non-critical (p=0.009), and spent more time on the scenario than the musculoskeletal group (means 9 vs. 6.5 mins, t=3.39, df=82, p=0.001). Furthermore, they elicited more critical cues that could provide evidence for a cardiac cause: a resting ECG (82% vs. 37%, χ^2=18.44, df=1, p=0.0001), pain relieved by rest (74% vs. 38%, χ^2=11.29, df=1, p=0.001), and family history of heart disease (62% vs. 31%, χ^2=7.81, df=1, p=0.005). The musculoskeletal group asked to palpate the chest more frequently than the cardiac group (89% vs. 64%, χ^2=7.33, df=1, p=0.007). These differences suggest differential emphasis in the testing of the two competing diagnoses.

Following diagnosis, 52 physicians took part in stimulated recall about the scenario (52/84, 62%): 20 from the cardiac group and 32 from the musculoskeletal group. These 32 could be divided into three management subgroups: 10 physicians who never mentioned a follow up, 14 who asked the patient to come back if his pain did not improve, and 8 who arranged for a follow up. In terms of experience, the 52 physicians consisted of 11 residents, 11 family physicians with 1–3 years in practice and 30 family physicians with $\geq$10 years in practice.

No physician changed his/her earlier diagnosis during stimulated recall. Inter-rater agreement on the coding of critical cue interpretations was high (Kappa=0.80). All disagreements were resolved with discussion. There was an overall significant difference in interpretations between the cardiac and musculoskeletal groups (χ^2=38.23, df=3, p=0.0001) and no difference between the three management subgroups (p=0.91). The musculoskeletal group interpreted critical cues correctly less frequently (40.1% vs. 75.9%, χ^2=32.61, df=1, p=0.0001), gained less information (no information gained: 34.3% vs.

Table 3: Frequencies and percentages of interpretations by physician group.

		Critical cues			
Physician group	Interpretation	Pain on gardening P=0.0001	Pain relief at rest P=0.01	Worsening pain P=0.003	Radiation P=0.001
Cardiac	Correct	14 (74%)	12 (80%)	9 (90%)	6 (38%)
	No info	5 (26%)	3 (20%)	1 (10%)	8 (50%)
	Misinterpretation	–	–	–	2 (13%)
	Overinterpretation	–	–	–	–
Times requested		19	15	10	16
Musculo-skeletal	Correct	2 (9%)	3 (23%)	2 (17%)	1 (4%)
	No info	11 (50%)	9 (69%)	8 (67%)	7 (29%)
	Misinterpretation	1 (5%)	–	–	16 (67%)
	Overinterpretation	8 (36%)	1 (8%)	2 (17%)	–
Times requested		22	13	12	24

20.7%, χ^2=5.77, df=1, p=0.017), misinterpreted more frequently (12.4% vs. 1.7%, χ^2=10.32, df=1, p=0.001) and overinterpreted more frequently (13.1% vs. 1.7%, χ^2=5.79, df=1, p=0.001) than the cardiac group. Table 3 provides a breakdown of interpretations for the critical cues where significant differences were observed between the cardiac and musculoskeletal groups.

All participants considered and tested for the possibility of a cardiac diagnosis. All knew in principle that chest pain that comes on with exertion and that is relieved by rest are typical features of cardiac pain. However, most physicians in the musculoskeletal group failed to gain information from these two cues. They did not consider gardening an activity sufficiently strenuous to induce cardiac pain. They thus inferred that the pain felt when gardening was due to an injured muscle being used; the pain stopped when the patient stopped gardening, hence stopped moving the muscle.

Physician 9: Working in the garden... That's difficult because it's kind of... work at rest. And I would say that, that would make me think more of musculoskeletal. Because it's not, in my opinion, taking heart rate to a point whereby he could be getting ischaemic type pain. So that's what it was, actually moving his arms about when he's working in the garden, thus making his musculoskeletal pain worse. (Cue shown — relieving factors: "If I take a rest then it usually goes away.") **Physician 9:** Yeah, once again that doesn't really help me differentiate between the two really. It just tells me that when he's not exerting himself, the pain is not there.

Physician 12: So he's doing the garden, that's not particularly helpful. If it's heavy digging or running around, then I would be a bit worried, but if he is lifting something, I think it may be that he has injured something, he's exacerbating the same muscle.

Physician 16: So I was looking for: "Yes, it came on when I walked down the garden and it stopped when I got to... and rested for five minutes." That would be cardiac. **Interviewer:** He says, "it's definitely worse when I'm working in the garden." **Physician 16:** Yeah, garden. I never got the feeling he was working terribly hard... Does it stop you? Always a good [question]. If a pain is bad enough it will stop you doing something. **Interviewer:** And he says, "if I take a rest then it usually goes away." **Physician 16:** It puts back the cardiovascular a little bit but it didn't stop him altogether and I would like to think, if he got really good-going cardiac pain, you wouldn't go on doing anything more.

Physician 36: And he says, "if I take a rest then it usually goes away," but then that could easily be musculoskeletal, because, like a sprain anywhere else, if you rest it, it will get better.

Interviewer: You wanted to know what makes it better. Why did you want to know this? **Physician 83:** Because if it's going to be muscular, it's going to get better when you rest or take something, but if it's a real chest pain, it's going to be there, if it's a cardiac cause.

Most physicians in the cardiac group interpreted pain on gardening that is relieved at rest as indicative of a cardiac cause.

Interviewer: You asked about things that bring on his chest pain. **Physician 29:** Yeah, because if he gives me the slightest hint that exercise induces this pain, I'm going for angina. **Interviewer:** And this is what he said. **Physician 29:** Yes, so until proved otherwise, he's got angina in my mind

Physician 79: Stable angina. Cardiac pain is associated with exertion and resting. If it is coming on regularly with exertion, it points more towards that. If it's regularly relieved completely by rest, again it comes towards a stable angina picture. You can also confuse it with movement, with musculoskeletal pain, but musculoskeletal pain is sometimes persistent once at rest as well after exertion, so...[5] **Interviewer:** So he said, "it's definitely worse when I'm working in the garden." **Physician 79:** That to me said, because it's exertional chest pain, I have to put cardiac pain a bit higher on my list and consider that rather than dismiss it.

Physician 39: It sounds like it comes on when he exerts himself.

The patient's chest pain is worsening: "when it started, it was just lasting for a few minutes, but the past couple of days, it's been worse and going on for longer." Worsening severity and duration of pain is characteristic of unstable or crescendo angina, a very serious condition due to the possibility of an impeding MI. Most physicians in the musculoskeletal group did not interpret this cue as evidence for angina, even when the link with unstable angina was considered, as illustrated by the following quotes.

Physician 16: Crescendo angina – "it's been getting worse and going on longer," but there was no other systemic upset with it, so he was a bit vague about it. It didn't sound to me like a particularly severe, cardiac sort of pain.

Physician 57: It is getting worse. But again, because he's been doing things, because he was lifting washing machine and then he was doing gardening, so he is aggravating it. It looks more like a muscular problem.

Physician 62: If he's torn a few intercostal muscles, that can get worse before it gets better.

Physicians in the cardiac group considered this cue as evidence for a cardiac cause.

Physician 11: That would suggest he has got crescendo or developing crescendo angina, possibly going to infarct.

Physician 33: Yeah, you see that is not consistent with a pulled muscle. A pulled muscle would tend to get better or stay static and the fact it is getting worse is more slightly worrying.[6] And the few minutes is a sort of cardiac thing.

Physician 80: Yes, so what it is telling me was that obviously this isn't musculoskeletal, so that was telling me that this is most likely going to be something like angina. It's going to be a cardiac underlying reason for his symptoms.

Forty physicians asked if the pain radiated anywhere. Only seven of them interpreted radiation to the right arm correctly, 6 in the cardiac group and 1 in the musculoskeletal group. The rest either gained no information from the cue or considered it as evidence against a cardiac cause. Some thought that it was the muscular injury affecting the right arm or that it was due to osteoarthritis (the patient has osteoarthritis of the knee). The following quotes illustrate the range of interpretations of this cue.

Physician 43, musculoskeletal group: He says it goes down to his right arm, which is reassuring... it's probably unlikely to be cardiac.

Physician 56, musculoskeletal group: It's unlikely to be angina unless he's got dextrocardia, heart on the right side of his body... Which is very rare.

Physician 65, musculoskeletal group: I thought maybe he has got musculoskeletal and this could be cervical spondylosis down the right arm. So I was happy with that, I think it's more musculoskeletal.

Physician 84, musculoskeletal group: I didn't like that. I didn't know what that means to be honest, but it's reassuring that it's not his left arm, so I thought.

Physician 69, musculoskeletal group: It didn't score anything in terms of the differential diagnosis because you can get cardiac pain down your right arm, but you can also get musculoskeletal, but it's important to know because you're still suspicious of both.

Physician 8, cardiac group: It can be either of the arms. It's a misconception it goes to left arm only. Yeah, so it was going to the right arm. It was significant to me. It could be cardiac.

Physician 79, cardiac group: Cardiac chest pain, generally central, can radiate to your left arm or your right arm. If it's a radiating pain, it's more likely to be a cardiac cause.

Physician 88, cardiac group: It doesn't help much, but it does take it away from angina a little bit and points more towards musculoskeletal. I am still not sure at this stage. It's not a typical area of radiation, although it can, but it's not very typical.

Most physicians (78%) indicated certainty of >50 for their top diagnosis (i.e., "fairly certain," "very certain" or "absolutely certain") and over half (52%) indicated certainty of $\geq$68. No significant differences were found between the two groups of physicians (p=0.11), between angina and musculoskeletal top diagnoses in the cardiac group (p=0.89) and between types of management in the musculoskeletal group (p=0.33). Nevertheless, physicians in the musculoskeletal group and especially those who did not mention a follow up tended to be the most certain. Table 4 presents descriptive statistics for certainty ratings by group.

[5]Compare with physician 83 above.
[6]Compare with physician 62 above.

Table 4: Mean (standard error of the mean), median and range of certainty ratings (0-100) for the top diagnosis by physician group.

Physician group		Mean (SE)	Median	Range
Musculoskeletal group (n=32)		71.09 (3.34)	72.50	20–100
Cardiac group (n=20)		61.37 (5.31)	60.00	20–96
Musculoskeletal group	Follow up not mentioned (n=10)	78.20 (5.88)	85.00	51–100
	Follow up if no improvement (n=14)	69.36 (5.61)	69.50	20–96
	Follow up arranged (n=8)	65.25 (4.95)	66.50	48–80
Cardiac group	Angina top diagnosis (n=11)	62.10 (7.75)	64.00	20–96
	Musculoskeletal top diagnosis (n=9)	60.56 (7.67)	60.00	24–90

4 Discussion

We investigated how family physicians diagnose a realistic and difficult case of chest pain, where diagnoses cannot be confirmed or excluded with certainty. The patients' risk factors and most of the clinical features present suggested a cardiac cause. However, the onset of pain (during lifting), the patient's own explanation and a single clinical feature (pain on palpation) suggested the more common diagnosis of musculoskeletal chest pain. The case contained an atypical feature of cardiac chest pain (right arm radiation) that most physicians either misinterpreted or failed to gain any information from. It also contained a clinical feature that could indicate a cardiac problem but was fairly ambiguous and therefore open to interpretation (pain on gardening). On balance, the two competing diagnoses were supported by similar amounts of evidence but the seriousness of the cardiac diagnosis necessitated that it was investigated further.

Nevertheless, more than half of the participating family physicians diagnosed musculoskeletal chest pain and decided not to investigate further. There were clear differences between them and the cardiac group, both in the search for information and the meaning attributed to the information elicited. The cardiac group elicited more information overall, requested more cues that were relevant to a cardiac diagnosis and took more time to diagnose than the musculoskeletal group. This could suggest that they were trying to exclude a cardiac diagnosis but were unable to. The musculoskeletal group was more focused on confirming the musculoskeletal diagnosis — they rarely failed to palpate the patient's chest — and needed less information before deciding how to manage the patient — they sent him home with painkillers. Some physicians in the musculoskeletal group advised the patient to come back if his pain did not improve, which can be considered standard clinical advice. Others arranged for a follow up and a minority advised the patient to seek help if his pain got worse. We cannot know to what extent this advice was due to a serious consideration of the possibility of a cardiac problem or part of their usual safety netting.

Sending the patient home is a riskier decision than referring him for further investigations and those who made it tended to be more confident in their diagnosis but not significantly so. Furthermore, those who never mentioned following up the patient tended to be more confident than those who did, but not significantly so, and were no different in their interpretation of critical cues, suggesting that they all built up a case for the musculoskeletal diagnosis. Most physicians were somewhere between "fairly certain" and "very certain," irrespective of what their diagnosis was. Moderate to high confidence in a final decision is a common finding in studies of coherence-based reasoning and is seen as integral to the process of making coherent decisions in ambiguous situations (Holyoak & Simon, 1999; Simon et al., 2001; Simon et al., 2004).

We found striking differences in the interpretation of information. The musculoskeletal group denigrated important information that the cardiac group took as evidence of heart disease. They interpreted information within the context of musculoskeletal chest pain, often employing simple, causal explanations that involved the movement of muscles. The scenario was unclear as to whether there was "pain on exertion," an undisputed indicator of cardiac pain. Pain came during gardening, an activity that the musculoskeletal group did not consider "exertion." Information diagnostic of cardiac pain that was elicited subsequently, such as pain relieved at rest and worsening pain, was fit into the mental representation of a pulled muscle.

Pre-decisional distortion of information is seen as a way to achieve coherent judgments (Russo, Carlson, Meloy, & Yong, 2008). According to theories of cog-

nitive consistency, reasoning is bi-directional: the evidence available helps form a judgment, and as a judgment emerges, it influences the evaluation of evidence (Simon et al., 2004). We are not passive recipients of information but proactively process the incoming information in light of our emerging judgments and conclusions. Cognitive consistency theories suggest that there is an inherent pressure to achieve coherent judgments. Therefore, faced with multiple and often conflicting pieces of probabilistic information, we may suppress, reject, or decrease the importance of inconsistent evidence, while bolstering consistent evidence (Brownstein et al., 2004; Dahlstrand & Montgomery, 1984; Russo, Medvec, & Meloy, 1996). This process is thought to happen without awareness (Holyoak & Simon, 1999; Russo, Meloy, & Wilks, 2000).

"Confirmation bias" has been discussed in relation to clinical diagnosis. Physicians generate diagnostic hypotheses within the first few minutes, possibly seconds, of the encounter (Elstein et al., 1978). This helps them to structure the subsequent search for information by reducing a very large problem space to a manageable size. A hypothesis may be singled out as the most promising early on for a variety of reasons, such as its prevalence, the patient's suggestion, the salience of a single cue or a collection of cues leading to rapid recognition of a disease or recall of similar patients (Schmidt, Norman, & Boshuizen, 1990). This selection can result in failing to collect information about alternatives, over-interpreting non-diagnostic information as supporting the leading diagnosis (Elstein et al., 1978), while ignoring or explaining away inconsistent evidence (Groopman, 2008; Kostopoulou, Devereaux-Walsh, & Delaney, 2009).

Our findings shed further light on this process, and show how diagnostic information (not just ambiguous information or non- diagnostic information) can be denigrated and its meaning can get distorted to fit the leading diagnosis. We suggest that this can be done through the construction of a mental model of the situation, a conditional reference frame (Koehler, 1991), supported by causal reasoning ("so that's what it was, actually moving his arms about when he's working in the garden, thus making his musculoskeletal pain worse"; "like a sprain anywhere else, if you rest it, it will get better"; "because he's been doing things, because he was lifting washing machine and then he was doing gardening, so he is aggravating it."), by denigrating or failing to integrate disconfirming evidence ([worsening pain] "but there was no other systemic upset with it, so he was a bit vague about it. It didn't sound to me like a particularly severe, cardiac sort of pain"; "doesn't really help me differentiate between the two really. It just tells me that when he's not exerting himself, the pain is not there."), and by making unwarranted assumptions ("[gardening] it's kind of... work at rest"; "because [gardening] it's not, in my opinion, taking heart rate to a point whereby he could be getting ischemic type pain"; "yeah, garden. I never got the feeling he was working terribly hard"; "it [the pain] didn't stop him altogether and I would like to think, if he got really good-going cardiac pain, you wouldn't go on doing anything more"). During information distortion, decision makers can change the meaning of incoming information, not only its perceived credibility or weight, to fit it with their current mental model of the situation. Information distortion is not usually considered as such in the confirmation bias literature (Klayman, 1995; Nickerson, 1998).

Coherence-based reasoning has been investigated using different methodologies, for example, pre-decisional vs. post-decisional comparisons of beliefs and tracing the development of preferences during the sequential presentation of information. The process tracing methodologies that we employed (active information search coupled with stimulated recall) help illustrate how the meaning and credibility of information can get distorted as a case is built in favor of the leading diagnosis. However, our data are retrospective and cannot be taken as an exact account of how physicians thought when they first diagnosed the scenario. It is unclear to what extent they were faithfully recalling their earlier thoughts, were recreating a process that more or less reflected how they reached a diagnosis in the first place or were justifying this diagnosis. In the latter case, information distortion would serve to reduce any post-decisional cognitive dissonance. All three possibilities are plausible and could have occurred to different degrees during stimulated recall. Nevertheless, there is evidence that pre-decisional distortion of information can occur and is of greater magnitude than post-decisional distortion (more than twice) suggesting that the drive to achieve a coherent judgment may often be greater than the drive to reduce cognitive dissonance following a decision (Russo et al., 1996; Russo et al., 1998). Furthermore, no physician changed his/her earlier diagnosis during stimulated recall, suggesting that reviewing the case by going systematically over the information elicited did not lead to novel insights and analyses.

Although coherence-based reasoning purports to provide "a general model of judgment and decision making in conditions of complexity" (Simon, 2004), it has been studied mainly in the formation of attitudes, beliefs and preferences, and has been used to explain performance on tasks with little or no diagnostic information and no apparent normative answer (Brownstein, 2003). Tasks traditionally associated with accuracy and a gold standard, like medical diagnosis, are not studied within the paradigm. Nevertheless, doctors see patients, not cues, and interpret diagnostic information in context. Even when a significant cue-diagnosis link is acknowledged, it

may not seem relevant to the specific patient ("Crescendo angina... but there was no other systemic upset with it, so he was a bit vague about it. It didn't sound to me like a particularly severe, cardiac sort of pain.")

The processes that cause information distortion and their contribution to diagnostic error clearly require investigation. For example, it would be useful to determine the types of clinical information that tend to get distorted, any moderators, e.g., mood and experience, and potential ways of reducing distortion. Information with diagnostic value, i.e., non-equivocal, has been found to reduce but not necessarily eliminate distortion once this has start building in favor of an option (Russo et al., 1998). Furthermore, participants can be manipulated to choose an option that is inferior for them by being presented with its best characteristics first and last (Russo, Carlson, & Meloy, 2006). Coherence researchers have indeed started to wonder about decision making errors. "Although it remains to be demonstrated empirically, we believe that predecisional distortion is not so harmless. When one alternative emerges as the leading one, possibly for reasons that are capricious or irrelevant to the alternative's relative value, distortion builds support for the leader. This is not to say that clearly disfavorable information will not dislodge the leading alternative. It is to note, however, that it may require more strongly disconfirming information to reverse a preference than should be necessary. Further, as confidence is built in the leader, it becomes increasingly difficult for new information to receive an unbiased evaluation" (Russo et al., 2008, p. 25).

References

Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1982). The clinical reasoning of randomly selected physicians in general medical practice. *Clinical & Investigative Medicine, 5*, 49–55.

Berger, J. P., Buclin, T., Haller, E., Van Melle, G., & Yersin, B. (1990). Right arm involvement and pain extension can help to differentiate coronary diseases from chest pain of other origin: A prospective emergency ward study of 278 consecutive patients admitted for chest pain. *Journal of Internal Medicine, 227*, 165–172.

Bleeker, J. K., Simoons, M. L., Erdman, R. A., Leenders, C. M., Kruyssen, H. A., Lamers, L. M., et al. (1995). Patient and doctor delay in acute myocardial infarction: A study in Rotterdam, The Netherlands. *British Journal of General Practice, 45*, 181–184.

Bouma, J., Broer, J., Bleeker, J., van Sonderen, E., Meyboom-de Jong, B., & DeJongste, M. J. (1999). Longer pre-hospital delay in acute myocardial infarction in women because of longer doctor decision time. *Journal of Epidemiology and Community Health, 53*, 459–464.

Brownstein, A. L. (2003). Biased predecision processing. *Psychological Bulletin, 129*, 545–568.

Brownstein, A. L., Read, S. J., & Simon, D. (2004). Bias at the racetrack: Effects of individual expertise and task importance on predecision reevaluation of alternatives. *Personality and Social Psychology Bulletin, 30*, 891–904.

Dahlstrand, U., & Montgomery, H. (1984). Information search and evaluative processes in decision-making - a computer-based process tracing study. *Acta Psychologica, 56*(1–3), 113–123.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data — Revised edition*. Cambridge, MA: The MIT press.

Esmail, A., Neale, G., Elstein, M., Firth-Cozens, J., Davy, C., & Vincent, C. (2004). *Case studies in litigation: Claims reviews in four specialties*: Manchester Centre for Healthcare Management, University of Manchester. Available from: http://www.pcpoh.bham.ac.uk/publichealth/psrp/PS006_Project_Summary.shtml

Groopman, J. (2008). *How doctors think*. Boston: Houghton Mifflin.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology-General, 128*, 3–31.

Kentsch, M., Rodemerk, U., Munzel, T., Muller-Esch, G., Ittel, T. H., & Mitusch, R. (2002). Factors predisposing to a nonadmission of patients with acute myocardial infarction. *Cardiology, 98*, 75–80.

Klayman, J. (1995). Varieties of Confirmation Bias. In J. Busemeyer, R. Hastie & D. L. Medin (Eds.), *Decision making from a cognitive perspective* (Vol. 32, pp. 385–418). San Diego, CA: Academic Press.

Koehler, D. J. (1991). Explanation, Imagination, and Confidence in Judgment. *Psychological Bulletin, 110*, 499–519.

Kostopoulou, O., Devereaux-Walsh, C., & Delaney, B. C. (2009). Missing celiac disease in family medicine: the importance of hypothesis generation. *Medical Decision Making, 29*, 282–290.

Kostopoulou, O., Oudhoff, J., Nath, R., Delaney, B. C., Munro, C. W., Harries, C., et al. (2008). Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in Family Medicine. *Medical Decision Making, 28*, 668–680.

Lyle, J. (2003). Stimulated recall: a report on its use in naturalistic research. *British Educational Research*

Journal, 29, 861.

McCormick, A., Fleming, D., & Charlton, J. (1995). *Morbidity Statistics from General Practice: Fourth National Study 1991–1992*. London: HMSO.

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*, 175–220.

Phillips, R. L., Jr., Bartholomew, L. A., Dovey, S. M., Fryer, G. E., Jr., Miyoshi, T. J., & Green, L. A. (2004). Learning from malpractice claims about negligent, adverse events in primary care in the United States. *Quality and Safety in Health Care, 13*, 121–126.

Russo, J. E., Carlson, K. A., & Meloy, M. G. (2006). Choosing an inferior alternative. *Psychological Science, 17*, 899–904.

Russo, J. E., Carlson, K. A., Meloy, M. G., & Yong, K. (2008). The goal of consistency as a cause of information distortion. *Journal of Experimental Psychology-General, 137*, 456–470.

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes, 66*, 102–110.

Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research, 35*, 438–452.

Russo, J. E., Meloy, M. G., & Wilks, T. J. (2000). Predecisional distortion of information by auditors and salespersons. *Management Science, 46*, 13–27.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*, 611–621.

Sequist, T. D., Marshall, R., Lampert, S., Buechler, E. J., & Lee, T. H. (2006). Missed opportunities in the primary care management of early acute ischemic heart disease. *Archives of Internal Medicine, 166*, 2237–2243.

Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *University of Chicago Law Review, 71*, 511–586.

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology-Learning Memory and Cognition, 27*, 1250–1260.

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology, 86*, 814–837.

Zarling, E. J., Sexton, H., & Milnor, P. (1983). Failure to diagnose acute myocardial infarction - the clinicopathologic experience at a large community hospital. *JAMA — The Journal of the American Medical Association, 250*, 1177–1181.

Appendix

1. The patient description and presenting complaint.

NAME: Roger Knox
AGE: 60 years old
ETHNICITY: Caucasian
HEIGHT: 1.79m
WEIGHT: 77kg (BMI 24, measured in 2002)
SMOKING STATUS: Gave up smoking 7 years ago
LAST BP: 132/86, taken 2 years ago
PAST MEDICAL HISTORY: Osteoarthritis of the knee 2002
MEDICATION: He is on no medication
LAST CONSULTATION: for knee pain 4 months ago.
APPEARANCE: As he enters the room you notice that he does not look ill.

PATIENT: "I've been getting this pain in my chest recently and my wife got a bit concerned. It's right here over the breastbone. It's like a dull ache, when I press right here in the middle of my chest. It's been 7 days now - I was helping my daughter move house, and as I was lifting the washing machine I felt it come on, like a dull sort of aching sensation. I thought I'd pulled a muscle in my chest or something."

2. Brief description of the scenario cues available for elicitation.

Chest pain every day, sometimes more than once.
Chest pain comes on with gardening.
Chest pain gets worse with gardening.
Gardening is the only exercise that he ever does, due to his knee osteoarthritis.
Chest pain is relieved at rest.
Chest pain is mostly in the middle of the chest but radiates down his right arm since yesterday.
Chest pain is not influenced by specific movements.
Chest pain is not related to inspiration.
Worsening duration and severity of chest pain.
Chest pain is not related to a particular time of day.
Chest pain does not feel like heartburn.
No nausea.
No vomiting.
No sweating.
No cough.
No sputum.
No haemoptysis.
No palpitations.
No breathlessness.
No feelings of anxiety.
He works as an insurance salesman.
He usually has 2 or 3 pints on a Friday night.
He lives with his wife. Their children left home years ago.
He has smoked about 10 cigarettes a day for 30 years. He gave up smoking 7 years ago.
His father had myocardial infarction aged 65 and now has heart failure.
His pulse is 80 beats per minute, regular, with a normal character.
His heart sounds are normal.
His respiratory rate is 16 breaths per minute.
His temperature is 36.8 °C.
His blood pressure is 136/74.
Total Cholesterol measured 1 year ago: 5.4 HDL 1.1 TG 2.1.
Some tenderness over the intercostal muscles on the right side of the chest.
His chest is resonant to percussion in all areas and sounds clear on auscultation.
His resting ECG is normal.
His chest X-Ray is normal.