BMC Medical Education

BioMed Central

Research article

The role of strategy and redundancy in diagnostic reasoning
Ralph F Bloch*, Daniel Hofer, Sabine Feller and Maria Hodel

Address: Institute for Medical Education (IAWF), Faculty of Medicine, University of Bern, Inselspital 37a, 3010 Bern, Switzerland

Email: Ralph F Bloch* - bloch@iae.unibe.ch; Daniel Hofer - Dan.Hofer@bluewin.ch; Sabine Feller - Sabine.Feller@iae.unibe.ch; Maria Hodel - Maria.Hodel@bag.admin.ch

* Corresponding author

Abstract

Background: Diagnostic reasoning is a key competence of physicians. We explored the effects of knowledge, practice and additional clinical information on strategy, redundancy and accuracy of diagnosing a peripheral neurological defect in the hand based on sensory examination.

Method: Using an interactive computer simulation that includes 21 unique cases with seven sensory loss patterns and either concordant, neutral or discordant textual information, 21 3rd year medical students, 21 6th year and 21 senior neurology residents each examined 15 cases over the course of one session. An additional 23 psychology students examined 24 cases over two sessions, 12 cases per session. Subjects also took a seven-item MCQ exam of seven classical patterns presented visually.

Results: Knowledge of sensory patterns and diagnostic accuracy are highly correlated within groups (R^2 = 0.64). The total amount of information gathered for incorrect diagnoses is no lower than that for correct diagnoses. Residents require significantly fewer tests than either psychology or 6th year students, who in turn require fewer than the 3rd year students ($p < 0.001$). The diagnostic accuracy of subjects is affected both by level of training ($p < 0.001$) and concordance of clinical information ($p < 0.001$). For discordant cases, refutation testing occurs significantly in 6th year students ($p < 0.001$) and residents ($p < 0.01$), but not in psychology or 3rd year students. Conversely, there is a stable 55% excess of confirmatory testing, independent of training or concordance.

Conclusions: Knowledge and practice are both important for diagnostic success. For complex diagnostic situations reasoning components employing redundancy seem more essential than those using strategy.

Background

A major part of the undergraduate medical curriculum is dedicated to teaching the art and science of diagnosing illness and disease. Furthermore, when assessing the clinical competence of medical students, examiners must infer knowledge and reasoning skills from the behavior and the responses of the candidates.

It stands to reason then that medical teachers should possess a thorough understanding of diagnostic reasoning as a "basic science" of medical education. In reality, however, our comprehension of the diagnostic reasoning process is hazy at best.

Table 1: Factors in empirical research on diagnostic reasoning.

Subjects studied	Experts [5], outstanding physicians [2], physicians-at-risk [6], "typical physicians" [7], learners [10] with differing levels of expertise.
Clinical information provided	Sequential text fragments [8], x-rays [9,5], pictures [10] (e.g. dermatology [11], pathology), standardized patients [2], other simulations [12].
Products of reasoning analyzed	Diagnostic success [13], items collected [4,14], recall [15,16], reflection, "thinking aloud", introspection [5]

The present study attempts to explore diagnostic reasoning by analyzing detailed recorded data-gathering behavior of experimental subjects with different levels of expertise in a computer simulation of patients with neurological lesions of the peripheral nervous supply to the hand.

Serious reasoning research started in psychology [1] during the 1950s. It has taken another 20 years for diagnostic reasoning to become an area of empirical research in medicine [2,3]. At a time when pragmatic medical educators believed in the existence of generic problem-solving skills, diagnostic reasoning research reestablished the primacy of content specific knowledge [4].

Initially research evolved along two intertwined threads, which alternatively supported and confused each other: the reasoning by (medical) experts and the reasoning by computers. By now, these two fields of research have largely gone their separate ways.

Three factors (Table 1.) have determined the type of experimental studies of diagnostic reasoning: Firstly the subjects studied, secondly the clinical information provided to subjects both by content and method, and thirdly the products of reasoning subjected to analysis.

This type of research is very labor intensive and, consequently, expensive. Thus it is difficult to collect sufficient data to reach adequate statistical power based on diagnostic success and process items alone. As a consequence, diagnostic reasoning research leans heavily on recall, introspection and reflection data [17]. It comes, therefore, as no surprise that the theories derived from this research tend towards models of semantic, analytical reasoning [18,19]. The literature is replete with a panoply of cognitive structures [20] – mainly semantic in nature – that are supposed to underlie diagnostic reasoning. The situation may be obscured further by the effect of social desirability bias, which may restrain experimental subjects from admitting to employing less than superlative reasoning strategies.

There is ample evidence [21,22] that analytical, semantic models alone do not fully explain diagnostic reasoning.

Research based primarily on semantic recall, introspection and reflection contains blind spots, when it comes to unconscious and implicit reasoning processes that are not based on semantic information. Methods focusing on such processes are thus required to look beyond semantic networks.

For further discussion, we define inference or inferential reasoning as: logical, algorithmic, mainly semantic, sequential, propositional, forward and/or backward directed, purposeful, open to reflection and introspection. In contrast, pattern recognition is: holographic, heuristic, mainly perceptual, parallel, redundant, unconscious, probabilistic and intuitive. Inferential reasoning is characterized by strategy, pattern recognition by redundancy.

By "strategy" we mean a purposeful sequence of tests, where the specifics of the next test are selected on the basis of previous tests such as to return maximum new information. "Redundancy" on the other hand expresses the number of tests that fail to provide any new information for inference.

A suitable experimental model should, therefore, involve a sufficient number of perceptual cues to allow for good statistical power. One such candidate is eye-movement scanning in the interpretation of histological slides or x-rays. Unfortunately, the fact that the ocular axis is directed at a certain location on the image does not indicate, what is actually seen by the central visual field or that visual information is indeed being recorded and processed.

We have selected a simple deterministic computer simulation involving the (sensory) neurological examination of the peripheral nervous system in the hand. The collected sequence of responses and coordinates of each sensory stimulus allow statistical inference on the reasoning strategies, be they inferential or based on pattern recognition.

For this experiment we asked ourselves 5 questions:

1. How do subjects pick the specific locations on the hand to be tested (strategy)?
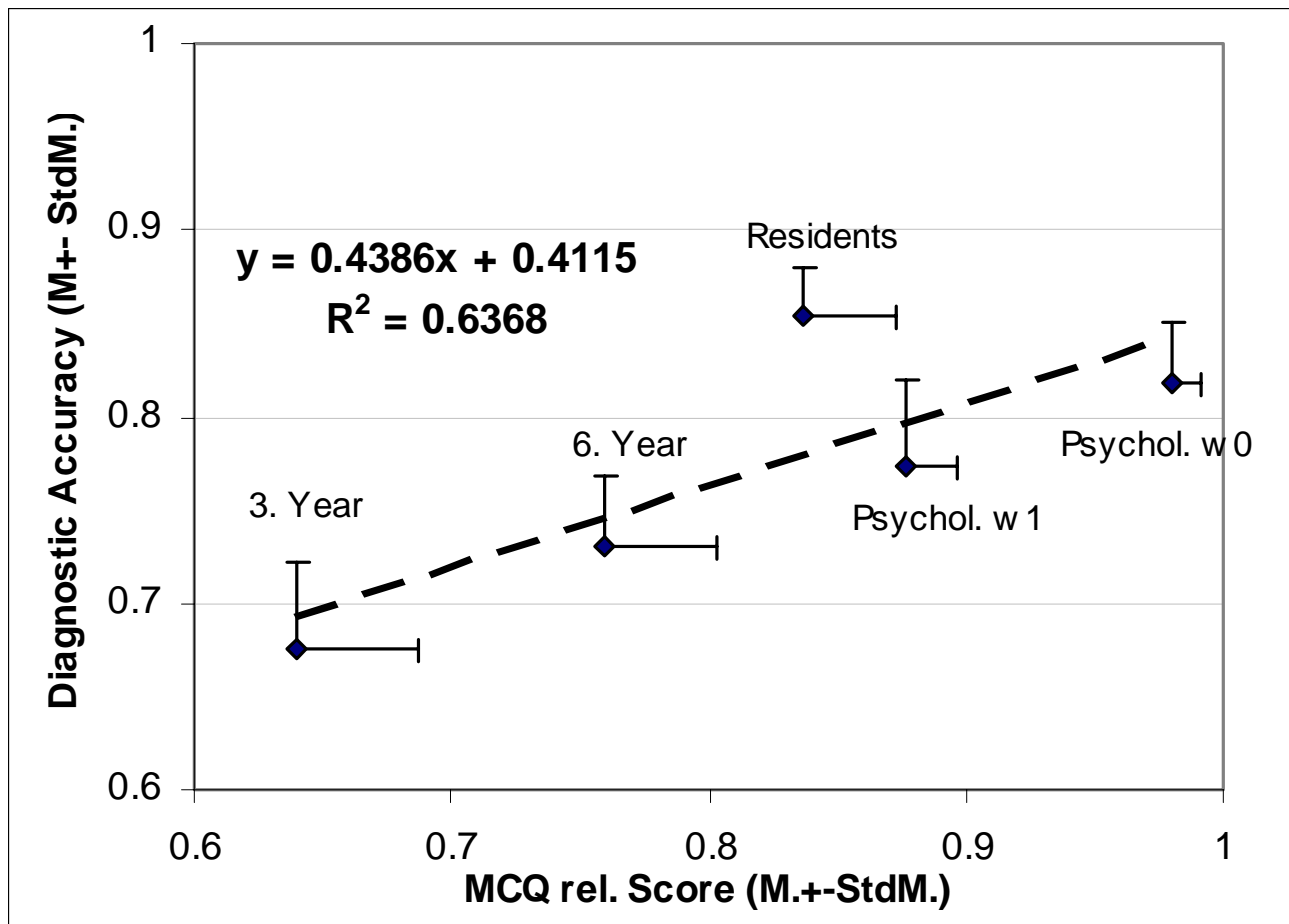
Figure 1
MCQ score and diagnostic accuracy fort the experimental groups. The values for psychology students in their initial session and in the second session, one week later, are plotted separately.

2. How many additional points in excess of what is required for strict inference, do they test before reaching a diagnosis (redundancy)?

3. How often is the selected diagnosis correct (accuracy)?

4. How are strategy, redundancy and accuracy related to knowledge and practice?

5. How are strategy, redundancy and accuracy affected, if subjects receive additional clinical information (symptoms and history) that is concordant, neutral or discordant with respect to the sensory pattern?

The key to answering these questions is the ability to quantify the information revealed by each successive sensory test. The accepted measure of information content is entropy, as introduced by Claude Shannon [23] in 1948 (Appendix A, see Additional file 1). Specifically, it indicates the potentially available information not yet revealed by the test sequence. An entropy value of 1.0 indicates that none of the available diagnostic information has yet been revealed and that all diagnostic possibilities are still equally likely. Conversely, an entropy value of 0.0 indicates that all relevant diagnostic information has been revealed and that only one diagnosis remains possible.

Table 2: Diagnostic Accuracy – Results of ANOVA: Tests of between sequence effects

Source	df	F	Sig.
CONCORD	2	17.392	.000
LEVEL	3	12.720	.000
CONCORD * LEVEL	6	2.877	.009
Error	1454		
Total	1466		
Corrected Total	1465		

Entropy does not attempt to estimate or model the current state of a typical diagnostician's knowledge regarding the case. It indicates simply, how much information has been revealed to an ideal inference engine. This allows us to demonstrate the gap that exists between the information content revealed and the information actually used by the diagnostician.

If an individual sensory test ("pin prick") does not change entropy, the test adds no new information – it is redundant. Thus redundancy is defined as the total number of sensory tests in a sequence that did not alter entropy.

The faster a subject accumulates sufficient information to arrive at the correct diagnosis, the more efficient is his diagnostic strategy. Quantitatively, this is indicated by a smaller area under the entropy/number-of-test curve (Figure 6).

A subject's strategy can be strictly inferential (i.e. no redundant tests), in which case it is automatically optimal, subjects could systematically attempt to refute the apparently likeliest hypothesis (Popperian strategy) or, as happens often in reality, they may try to confirm those abnormal findings that support their currently favorite hypothesis.

To determine, which information gathering strategy was used, three measures were calculated: (i) how quickly relevant information was collected as expressed by the area under Shannon's entropy as a function of the number of tests; (ii) the specific number of refutations of discordant cues (Refutation matrix, Appendix B, see Additional file 2); and (iii) the excess of confirmatory testing (Confirmation matrix, Appendix C, see Additional file 3).

Methods
Seven familiar neurological patterns of sensory loss in the hand were simulated: C6, C7 and C8 nerve root injury, radial, median and ulnar nerve lesion and poly-neuropathy. Photographs of the dorsal and volar aspects of either the left or right hand were displayed on the screen. With

mouse clicks subjects could "test" individual points on the hand. Depending on the location tested and the underlying predetermined diagnosis, one of three verbal responses was returned deterministically in a small pop-up window at the point tested:

• "it feels normal",

• "it feels different", or

• "I can hardly feel it".

The simulation ran as a Java applet within a regular Web page. Subjects were not provided with feedback regarding individual diagnoses during the actual experiment; they did, however, receive detailed feedback after they had completed all the cases.

Each pattern was presented in the context of additional clinical information (symptoms, history and a functional photograph of the hand) concordant, neutral or discordant relative to the sensory pattern. The additional clinical information was relatively bland, providing only subtle suggestions as to the actual diagnosis whether concordant or discordant, although the concordant information was more specific. The neutral items contained no clues. For example, the discordant cases of radial and median nerve deficits had a history vaguely suggestive of a mild cervical injury. Sensory patterns and additional clinical information were repeatedly checked by experienced neurologists for realism.

The experimental subjects consisted of a convenience sample of 23 psychology students, 21 3rd year medical students, 21 6th year medical students (Switzerland has a six year medical curriculum; during the first two years students concentrate on basic sciences) and 21 senior neurology residents. The junior medical students had studied neuroanatomy, but were unfamiliar with the detailed sensory patterns and clinical pictures. They had never practiced sensory examination. Senior medical students had studied sensory patterns, had limited knowledge of
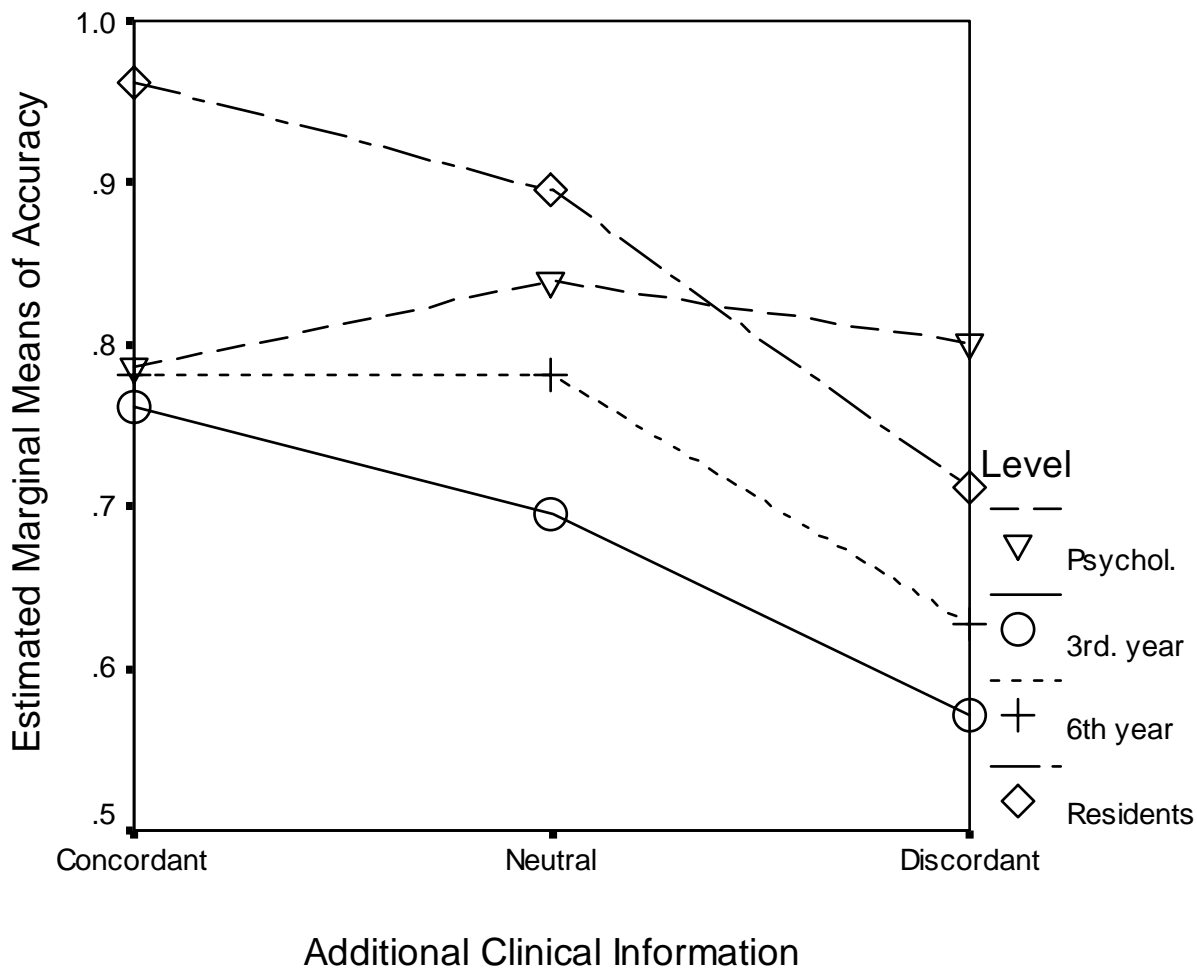
Figure 2
Estimated marginal means of diagnostic accuracy as function of the level of training and the degree of concordance.

clinical pictures and had been introduced to sensory testing. Neurology residents acted as substitute experts, since we were not able to recruit sufficient certified neurologists. The psychology students served as a control group with roughly matching intelligence but no medical education. Psychology students knew neither neuroanatomy nor clinical pictures. Neither had they been taught sensory examination. They were exposed to visually presented maps of the sensory patterns as part of the experimental protocol.

Psychology students participated in two sessions one week apart, the rest in one session each. In their first session psychology students were shown the seven patterns

as visual maps together with diagnostic labels during 15 minutes. Otherwise all sessions followed the same sequence: (i) an MCQ test of the seven patterns presented visually as sensory maps; (ii) a single practice case that was not recorded; (iii) a series of 12 cases each for the psychology students and 15 cases each for the rest in a balanced block design. As result of an oversight, the blocks were not perfectly balanced across the 21 possible combinations (6 × 7 / 2). There were only seven unique blocks each with three different sequences of cases. Altogether, all 21 cases occurred with equal frequency for each group. We do not, therefore, believe that this error introduced any significant bias.
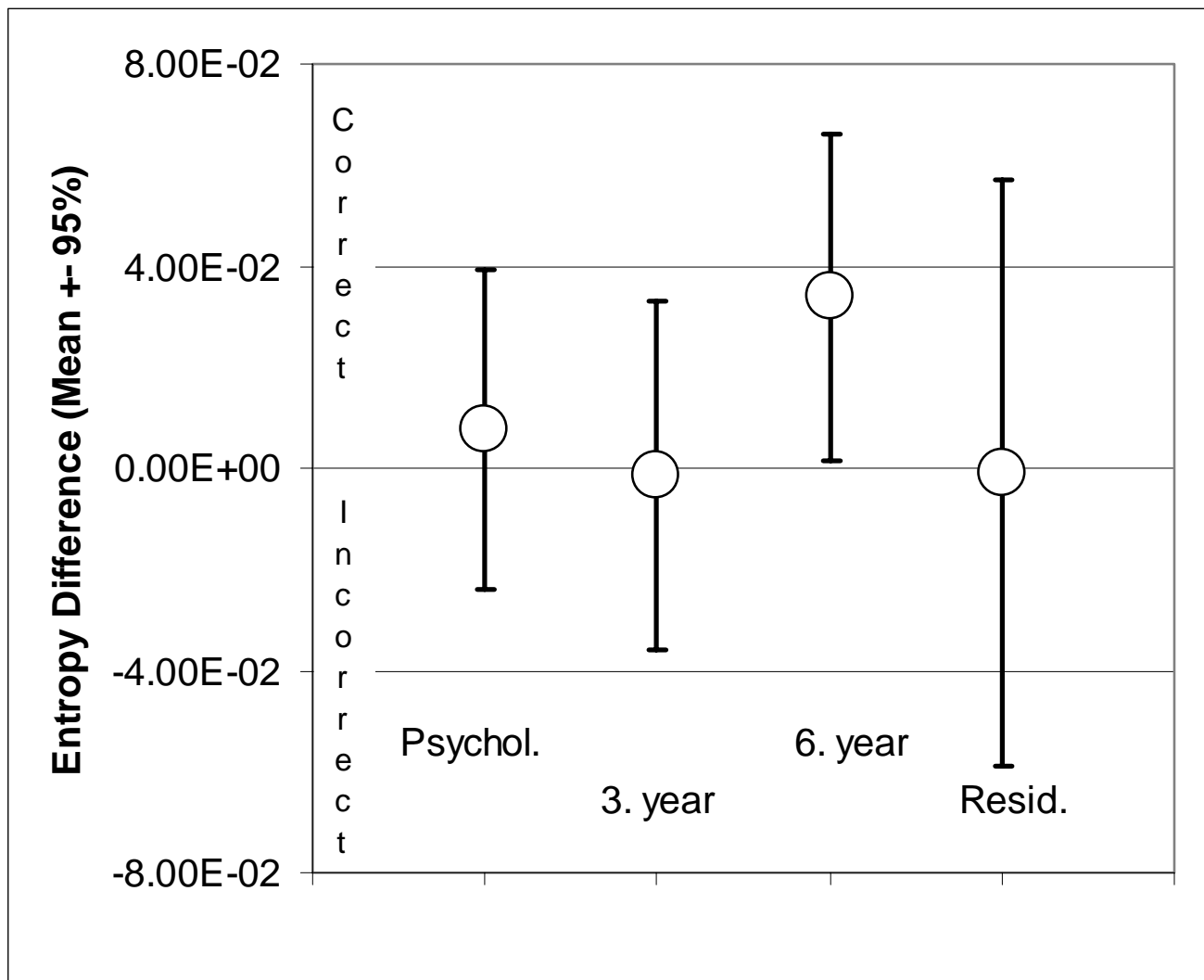
Figure 3
Results of unpaired T-tests of residual entropy for correct and incorrect diagnoses.

After each test subjects had the option of picking a diagnosis from a menu and proceeding to the next case. As part of a further study to be reported separately, the test sequence was interrupted automatically at 5, 10, 20 and 40 tests. Subjects were then asked to indicate their current best estimates for the likelihood of each diagnostic hypothesis.

Test coordinates and time since the previous test were stored test by test for the whole case sequence in the client side Java applet and sent to the Web server as part of the active server page request, upon the selection of a specific diagnosis. Data were automatically stored in a relational database (Microsoft Access) keyed to case and subject. After completion of the experimental phase of the study, data were preprocessed by means of a Microsoft Visual Basic program to determine the expected findings at each point tested for the actual diagnosis as well as for the alternative hypotheses. These results were again stored in a relational table. Based on these findings, plausibility, entropy, redundancy, refutation- and confirmation-counts were calculated with a second MS-VB program (Method described in Appendix A, B & C, see additional files 1, 2

Table 3: Results of MANOVA: Tests of Between-Sequence Effects

Source	Dependent Variable	df	F	Sig.
Model	Redundancy	12	207.878	.000
	Area	12	160.755	.000
LEVEL	Redundancy	3	46.174	.000
	Area	3	16.524	.000
CONCORD	Redundancy	2	15.078	.000
	Area	2	.539	.583
LEVEL * CONCORD	Redundancy	6	2.216	.039
	Area	6	.974	.441
Error	Redundancy	1416		
	Area	1416		
Total	Redundancy	1428		
	Area	1428		

and 3). SPSS was used for the statistical analysis of these derived dependent variables.

For each subject the knowledge of sensory patterns was calculated as the ratio of correctly identified patterns over seven, the total number of patterns in the multiple choice exam. Diagnostic accuracy was calculated for each subject as the ratio of correct diagnoses over total cases processed.

Psychology students participated in two sessions with 12 cases each thus diagnosing a total of 24 cases. Since only 21 unique cases existed, each of these subjects encountered three of 21 cases twice. Using a random number generator, either the first or second of these double cases was dropped from further analysis.

Results
A total of 1,428 sequences with 27,524 test points were analyzed. In 17 sequences subjects guessed the diagnosis without performing any tests. Residents guessed 12 concordant cases correctly. Students guessed three of the remaining five sequences incorrectly. The two correct guesses were in concordant cases, one by a 3rd and one by a 6th year student.

Group means of MCQ scores and diagnostic accuracy correlate well (Fig. 1). For psychology students the mean values for the initial and follow-up session one week later were calculated separately. Diagnostic accuracy of residents is higher than one would expect from the knowledge of patterns alone.

The diagnostic accuracy of the diagnosis is examined by ANOVA (Table 2). Diagnostic accuracy is affected significantly both by the level of training and the degree of concordance.

The diagnostic accuracy of psychology students is not affected by the degree of concordance, while the accuracy of the residents is significantly eroded by discordant information (Fig. 2).

Discordant cases form a homogeneous subset against neutral and concordant cases at $\alpha = 0.05$. The level of training does not separate into homogeneous subsets.

For all four groups of experimental subjects the residual entropy does not differ significantly for cases diagnosed correctly and incorrectly (Fig. 3). The 6th year students, in fact, show borderline increased entropy (lower certainty) for correct diagnoses. Diagnostic errors, therefore, do not appear to be due to insufficient information gathering but rather to flawed reasoning.

Area under the entropy curve and redundancy were examined by MANOVA (Table 3).

Both level of training and degree of concordance have a significant effect on redundancy. But the area under the entropy curve depends only on the level of training, not the degree of concordance. Post hoc analysis (Scheffé test) shows the area under the entropy curve to split up into two homogeneous subsets: 3rd year medical students versus the rest. Redundancy splits into three homogeneous subsets: 3rd year students, residents, and 6th year medical together with psychology students as a middle group.

In regards to degree of concordance, redundancy splits into two homogeneous subsets: concordant versus neutral and discordant.

The redundancy of psychology students is not affected by the additional clinical information – they do not recog-
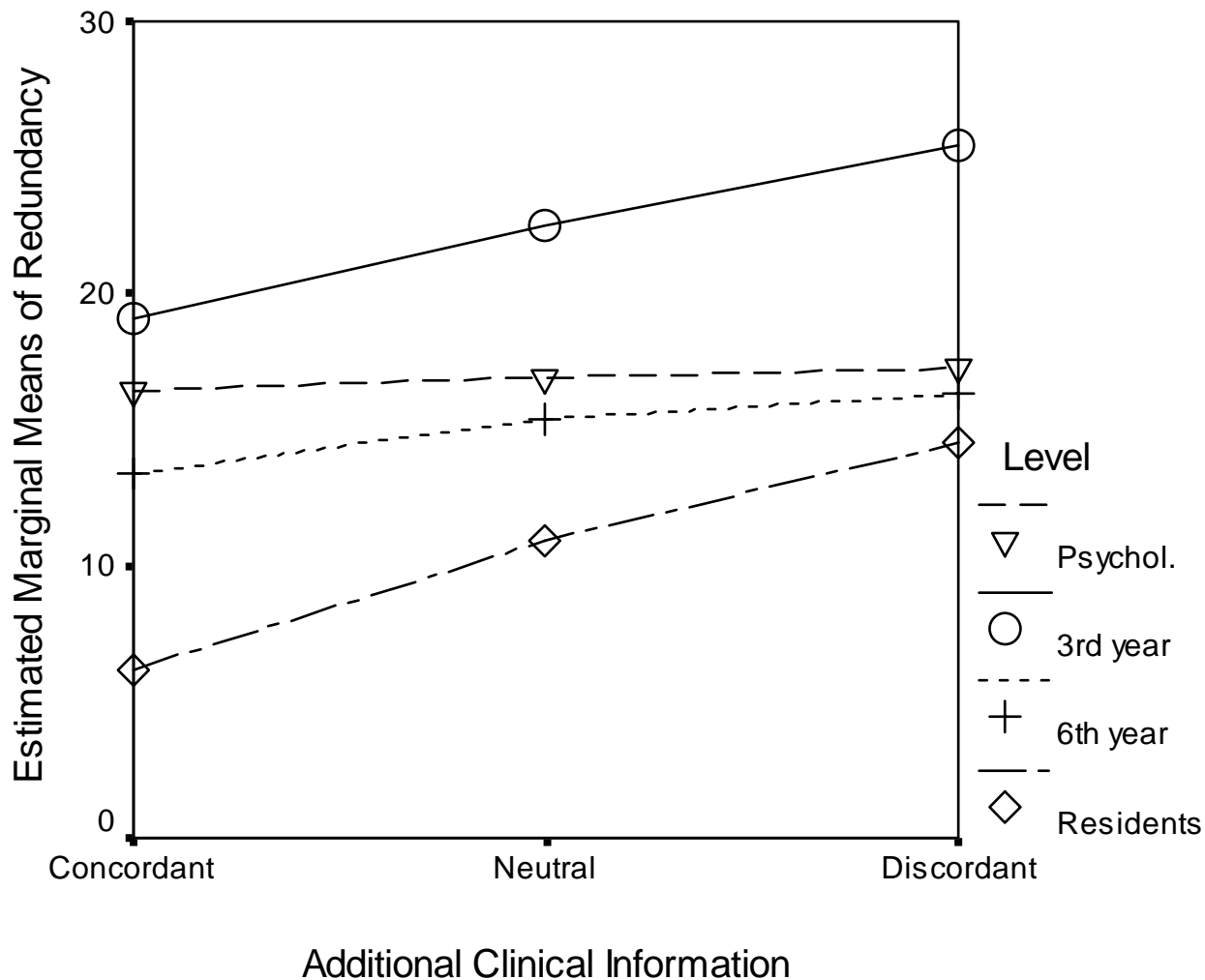
Figure 4
Estimated marginal means of redundancy as function of level of training and degree of concordance.

nize its implication (Fig. 4). For the residents, on the other hand, it affects redundancy by almost a factor of three.

For the area under the entropy curve (strategy), the effect of either the level of training or the degree of concordance is less clear cut (Fig. 5).

It is obvious, though, that 3rd year medical students show least evidence of strategy, independent of additional clinical information.

To test whether subjects specifically attempt to refute the diagnostic hypotheses suggested by the additional clinical information in the discordant cases, we employ Popperian analysis (Table 4) as described in Appendix B (see Additional file 2).

Psychology and 3rd year medical students are not affected by the additional discordant clinical data. They don't know enough about the clinical syndromes.
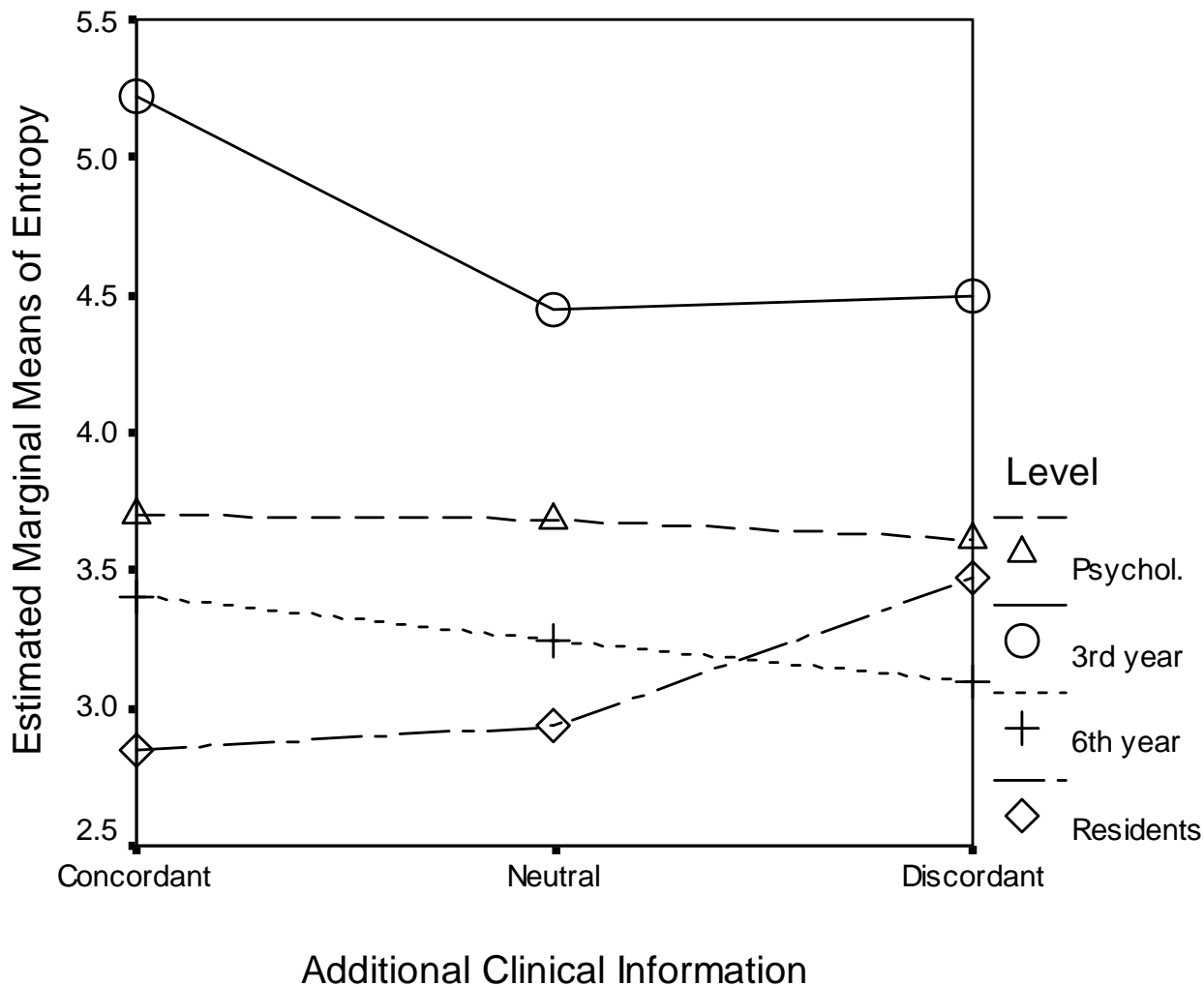
Figure 5
Estimated marginal means of area under the entropy curve as function of the level of training and the degree of concordance.

Table 4: Contingency table analysis of Popperian refutation counts.

	Psychol.	3. Year	6. Year	Residents
Chi-Square	1.225	0.263	20.282	7.385
p-Value	n.s.	n.s.	< 0.001	< 0.01
Common Odds Ratio	1.030	0.985	1.168	1.111
Redundancy Ratio	1.012	1.129	1.059	1.266

Table 5: Chi-square, significance and estimated ratio of actual over expected confirmatory tests.

	Ratio actual/expected	χ^2	p-Value
All	1.57	2084.3	<0.001
Psychol.	1.65	873.5	<0.001
3. year	1.52	549.0	<0.001
6. year	1.51	368.1	<0.001
Resid.	1.55	313.1	<0.001
Concord	1.61	690.0	<0.001
Neutral	1.54	648.0	<0.001
Discord	1.56	756.3	<0.001

Residents and 6th year students show a significant though small attempt at refuting the clinically suggested diagnoses. The excess of specific refutations is about 11% for residents and 17% for 6th year students respectively (Common Odds Ratio).

There is a significant difference in the increase of redundancy between residents and 6th year students: χ^2 = 17.24; p < 0.001; C.O.R. = 1.24. In other words, in the presence of discordant information 6th year students seem to use more strategy while residents rely more on redundancy. This could be explained by the "intermediacy effect".

Finally, we looked for evidence of confirmatory testing (Appendix C see Additional file 3). In fact, confirmatory testing seems to be a stable feature, independent of level of training or degree of concordance (Table 5).

The tendency to selectively confirm expected hypotheses rather than testing randomly or refute alternative hypotheses appears inherent in this diagnostic reasoning experiment.

Discussion and Conclusions

Diagnostic accuracy, strategy and redundancy depend primarily on the knowledge of sensory patterns and associated syndromes. The effect of knowledge on accuracy and redundancy appears to be stronger than on strategy. In fact, effective data-gathering strategies seem to play a minor role. Even where appropriate, little refutation of alternative hypotheses occurs. Just the opposite: confirmatory testing seems to be dominant.

In addition, both accuracy and redundancy, but not strategy appear to depend on practice independently from knowledge.

These results appear somewhat counterintuitive. Experts should have vastly better problem-solving strategies than novices. True, in the real world, experts also have an edge on knowledge. The knowledge spread in our experiment was insufficient to demonstrate that aspect.

There might be another explanation, however. In our experiment, to reach a diagnosis by inference requires not only the seven diagnostic hypotheses to be present in short-term memory, but also the roughly seven tests in strategically placed locations and their combinations must be available at all times. In other words, for purely inferential diagnostic reasoning one needs to operate on approximately 49 items or 5.6 bits of information. As George A. Miller [28] has shown the capacity of short term memory is only about seven items or 2.8 bits of information. The scope of short-term memory, therefore, would appear insufficient to support pure inference. Short of using memory substitutes, such as paper and pencil, the only alternative is to resort to what Miller refers to as "recoding" – an implicit reasoning strategy. This is a hypothesis that requires further testing.

It remains surprising, however, that the psychology students were able to set up an efficient recoding scheme after only 15 minutes' training that allows easy shifting from overt to latent pattern recognition.

The reported findings may also have implications for teaching and assessment. If the rate limiting factor for inference is the number of items that have to be kept in short term memory, teachers can assist learners by constructing diagnostic trees that involve only two or three branches at each decision point, rather than long lists of differential diagnoses. Such cognitive structures correspond to Bordage's [29] key features or Mandin's schemes [30].

In the assessment of diagnostic reasoning, redundancy of requested information appears as a second independent, sensitive measure of competence besides the accuracy of the diagnosis.
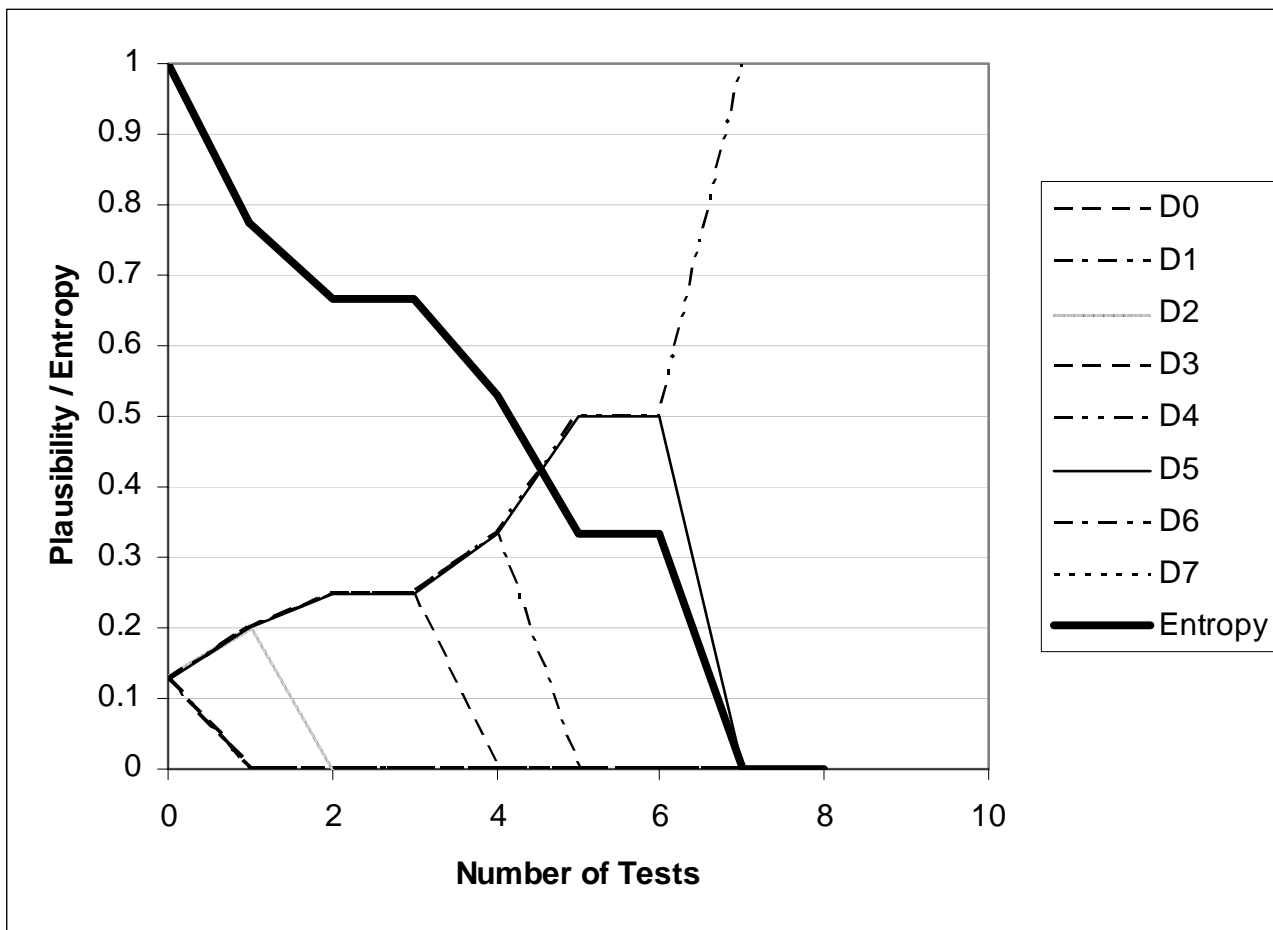
Figure 6
Example of one illustrative sequence. For all diagnoses except D4 the plausibilities successively disappear. Correspondingly, the entropy falls from 1.0 to 0.0. After seven tests total certainty exists.

Table 6: Popperian refutation matrix for the seven discordant cases. The '+' indicates the cells favored by the discordant information, whereas the '-' designates non-favored cells.

		Diagnostic Hypotheses						
		C6	C7	C8	Ra	Me	Ul	Pn
Actual Diagnoses	C6	0	-	-	+	+	+	+
	C7	-	0	-	+	+	+	-
	C8	-	-	0	-	-	+	-
	Ra	+	+	+	0	-	-	-
	Me	+	+	+	-	0	-	-
	Ul	-	-	-	-	+	0	-
	Pn	-	-	+	-	-	-	0

Competing interests

None declared.

Authors' contributions

RB conceived the experiment, wrote the initial Java applet, analyzed the results and wrote the paper. DH and SF refined the software, designed and supervised the experimental details. MH and SF prepared the cases, recruited subjects and collected the data.

Additional material

Additional File 1

Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6920-3-1-S1.doc]

Additional File 2

Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6920-3-1-S2.doc]

Additional File 3

Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6920-3-1-S3.doc]

Acknowledgments

References

1. Bourne LE and Dominows RL **Thinking.** *Annual Rev Psychol* 1972, **23:**105-130
2. Elstein AS, Kagan N, Shulman LS, Jason H and Loupe MJ **Methods and theory in the study of medical inquiry.** *J Med Educ* 1972, **47:**85-92
3. Barrows HS and Bennett K **The diagnostic (problem solving) skill of the neurologist: experimental studies and their implications for neurological training.** *Arch Neurol* 1972, **26:**273-277
4. Elstein AS, Shulman LS and Sprafka SA **Medical Problem Solving. An Analysis of Clinical Reasoning.** *Harvard University Press, Cambridge Mass* 1978,
5. Raufaste E, Verderi-Raufaste D and Eyrolle H **[Radiological expertise and diagnosis. II. Empirical study].** *J Radiol* 1998, **79(3):**235-40
6. Turnbull J, Carbotte R, Hanna E, Norman G, Cunnington J, Ferguson B and Kaigas T **Cognitive difficulty in physicians.** *Acad Med* 2000, **75(2):**177-81
7. Barrows HS, Norman GR, Neufeld VR and Feightner JW **The clinical reasoning of randomly selected physicians in general medical practice.** *Clin Invest Med* 1982, **5(1):**49-55
8. Bordage G, Grant J and Marsden P **Quantitative assessment of diagnostic ability.** *Med Educ* 1990, **24(5):**413-425
9. Babcook CJ, Norman GR and Coblentz CL **Effect of clinical history on the interpretation of chest radiographs in childhood bronchiolitis.** *Invest Radiol* 1993, **28(3):**214-217
10. Brooks LR, LeBlanc VR and Norman GR **On the difficulty of noticing obvious features in patient appearance.** *Psychol Sci* 2000, **11(2):**112-7
11. Kulatunga-Moruzi C, Brooks LR and Norman GR **Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis.** *Teach Learn Med* 2001, **13(2):**110-6
12. Myers JH and Dorsey JK **Using diagnostic reasoning (DxR) to teach and evaluate clinical reasoning skills.** *Acad Med* 1994, **69(5):**428-9
13. Regehr G, Cline J, Norman GR and Brooks L **Effect of processing strategy on diagnostic skill in dermatology.** *Acad Med* 1994, **69(10 Suppl):**S34-S36
14. Schwartz W **Documentation of students' clinical reasoning using a computer simulation.** *Am J Dis Child* 1989, **143(5):**575-579
15. Norman GR, Brooks LR and Allen SW **Recall by experts and novices as a record of processing attention.** *J Exp Psychol Learn Mem Cogn* 1989, **5:**1166-74
16. Patel VL, Groen GJ and Frederiksen CH **Differences between medical students and doctors in memory for clinical cases.** *Med Educ* 1986, **20(1):**3-9
17. Schmidt HG, Norman GR and Boshuizen HP **A cognitive perspective on medical expertise: theory and implication.** *Acad Med* 1990, **65(10):**611-21
18. Bordage G and Lemieux M **Semantic structures and diagnostic thinking of experts and novices.** *Acad Med* 1991, **66(9 Suppl):**S70-S72
19. Schmidt HG, Norman GR and Boshuizen HP **A cognitive perspective on medical expertise: theory and implication.** *Acad Med* 1990, **Vol 65(10):**611-621
20. Custers JFM, Regehr G and Norman GR **Mental representations of medical diagnostic knowledge: a review.** *Acad Med* 1996, **71:**555-61
21. Norman GR and Brooks LR **The non-analytical basis of clinical reasoning.** *Advances in Health Sciences Education* 1997, **2:**173-184
22. Elstein AS **Heuristics and biases: selected errors in clinical reasoning.** *Acad Med* 1999, **74(7):**791-4
23. Shannon CE **A mathematical theory of communication.** *The Bell System Tech J* 1948, **27:**379-656
24. Jaynes ET and Bretthorst GL **Probability Theory: The Logic of Science. To be published in July 2003.** [http://omega.albany.edu:8008/JaynesBook]
25. Collins A and Michalski RS **The Logic of Plausible Reasoning: A Core Theory.** *Cognitive Science* 1989, **13:**1-49
26. Gelman A, Carlin J, Stern H and Rubin D **Bayesian Data Analysis.** *Boca Raton, FL: Chapman & Hall* 1995,
27. Popper K **The logic of scientific discovery.** *Routledge, London* 1934,
28. Miller GA **The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.** *Psychol Rev* 1956, **63:**81-97
29. Bordage G **Elaborated knowledge: a key to successful diagnostic thinking.** *Acad Med* 1994, **69(11):**883-5
30. Mandin H, Jones A, Woloschuk W and Harasym P **Helping students learn to think like experts when solving clinical problems.** *Acad Med* 1997, **72(3):**173-9

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6920/3/1/prepub