Investigating domain-knowledge and reasoning as predictors of diagnostic performance with PATSy.

Richard Cox
School of Cognitive & Computing Sciences
University of Sussex, Falmer, BN1 9QH, UK
richc@cogs.susx.ac.uk

Carmel Lum
Department of Experimental Psychology
University of Cambridge, Downing Site, Cambridge, CB2 3EB, UK
cl278@cam.ac.uk

Jonathan Kilgour
Human Communication Research Centre
University of Edinburgh, Edinburgh, EH8 9LW, UK
jonathan@cogsci.ed.ac.uk

Abstract.

This study investigated the clinical reasoning of speech and language pathology students who used PATSy (a web-based, multimedia database of clinical cases for education and research) in the context of a cognitive neuropsychology course in aphasia. Theoretically, the course was oriented around an information processing account of normal and pathological language production. PATSy contains numerous 'virtual patient' cases consisting of multimedia data on individuals who have language production disorders. The aims were a) to assess PATSy as a research tool in the study of clinical reasoning, b) to investigate students' clinical reasoning and that of experts, and c) to learn more about the relative contributions of domain-specific knowledge and general reasoning ability to diagnostic performance on PATSy.

Experts and students used PATSy under exam conditions to diagnose a previously unseen patient case. Subjects' notes, hypotheses and conclusions were logged in the course of their reasoning, together with a trace of their assessment test selections. Students were pre-tested on domain-general and domain-specific knowledge. The results showed that end-of-course diagnostic reasoning performance was predicted by pre-course domain knowledge, the number of PATSy cases accessed by students during the course, and deductive reasoning performance. The implications for a) teaching clinical reasoning skills and b) adding intelligent support to PATSy are discussed.

Keywords: web-based multimedia, case-based teaching, human reasoning, diagnostic reasoning, clinical reasoning, expert-novice differences, Wason selection task, analytical reasoning, model-tracing.

1 Introduction

This paper presents findings from an investigation of clinical reasoning by speech and language pathology students who used PATSy (a web-based, multimedia database of clinical cases for education and research) in the context of a cognitive neuropsychology course on aphasia. In the preliminary study reported here, experts and students used PATSy to diagnose a previously unseen patient case. Subjects' written notes, hypotheses and conclusions were logged, together with details of the neuropsychological tests that they selected during the course of their diagnostic reasoning. These protocols were analysed and preliminary models of expert and novice diagnostic reasoning and hypothesis generation skills were developed. Participants were also pre-tested on a range of domain-independent (generic reasoning) and domain-knowledge specific reasoning tests.

1.1 Cognitive demands on the learner

The diagnosis of neuropsychological disorders is a knowledge-rich and relatively ill-structured domain. There are no 'prototypical' cases and a considerable amount of prior knowledge is required. However, testing a patient also requires 'weak' methods [11] such as knowledge of search heuristics (*e.g.* binary-split), and other general strategies. As a mixture of strong and weak methods, then, diagnosis imposes a very heavy cognitive load on the learner - it requires several distinct kinds of knowledge: knowledge of acquired language pathologies, knowledge of the information processing model, specific knowledge of neuropsychological tests and what they measure, more general psychometric knowledge, scientific enquiry skills and heuristic search skills. The two latter types of knowledge are domain-independent, whereas the first three are domain-specific. A third type (general psychometric knowledge) falls in between.

Thus, when reasoning about a particular patient case, students are required to generate and test hypotheses about how a model of normal speech and language processing might be degraded in a manner which would account for the language disorder observed in the patient.

The goal of hypothesis testing is to 'make accurate predictions about the portion of the world you're dealing with' [9]. There are two kinds of hypothesis testing - inductive and deductive. In the former, several events are observed and then explanatory hypotheses are devised to explain the observations and to make generalisations from a few cases to many or all cases. In deductive reasoning, the hypothesis is proposed first and systematic relevant observations are then made in order to see if it is supported or whether it must be abandoned (rejected). The kind of hypothesis testing that speech and language students are required to engage in when analysing a patient is mainly deductive, although both modes are used.

Previous research in medical diagnostic reasoning [6,10,18], suggests that experts tend to generate fewer and more general hypotheses than novices. It has been suggested that this tendency of experts reflects their use of case-based reasoning (calling to mind past patients *i.e.* inductive reasoning). We suggest that this general strategy on the part of experts also serves to rapidly reduce problem space size early in diagnosis. Expert knowledge invariably includes experience of a wide variety of cases.

Novices lack an extensive case-base to influence their reasoning, and so are required by necessity to rely predominantly upon deductive reasoning. This mode of reasoning is also desirable within scientific disciplines and is necessary as a counterweight to the inevitable and difficult-to-suppress practise of (inductive) case-based reasoning that tends to dominate the reasoning of the clinically experienced practitioner.

On PATSy, data for a particular patient case consists of results from a battery of neuropsychological tests. The tests are interpretable at various levels of granularity (item by item, overall test scores, error patterns). The absolute scores and the error patterns are both of theoretical significance. A clinician's interpretation may also be moderated by more global factors such as the 'severity' of the patient's condition and/or the patient's social and educational background. Confronted with a large menu of tests, students must plan and adopt an orderly approach to their enquiries. They must be capable of isolating and controlling particular factors and recognise that some factors may be confounded with others.

Successful performance in diagnosis depends on knowledge of the domain (i.e. knowledge of tests and a theory of language information processing) and the ability to search and reason validly in generic terms. It was therefore predicted that the ability to formulate clinical arguments in diagnosis would correlate positively with good performance on measures of domain-specific and general reasoning skills.

1.2 PATSy

PATSy is a web-based multimedia database of clinical cases currently in use at 11 UK Universities (representing 80% of UK institutions that teach speech and language science). Detailed descriptions can be found in [12,13] and on the project's web site[1] PATSy is used as resource in teaching students how to diagnose language impairment in brain-injured patients and serves as a repository of patient cases for researchers and clinicians. PATSy provides a learning environment in which students may practice deductive reasoning in diagnosis. It provides students with opportunities for practise on a wider range of cases than they would Typically see in clinical placements, including access to rare cases. PATSy allows students to obtain pre-clinical experience and to rehearse diagnosis several times on the same (virtual) patient. The system also provides 'anytime, anywhere' access to training in an important area of clinical education.

The 'speech and language sciences' version of PATSy currently contains 22 language disorder cases. The range of disorders includes agrammaticism, aphasia and dyslexia. PATSy also contains 46 interactive neuropsychological tests/sub-tests such as the Test for Reception of Grammar (TROG) [3]. Each patient case on the system has full data from around 27 tests, on average. This represents very rich, 'research grade' data for each case on the system.

Each PATSy 'case' consists of digitised samples of speech, images, video and psychometric data collected by clinicians and researchers from actual patients in hospital and rehabilitation settings. A typical session might proceed as follows: using a web browser, a student logs on to PATSy and selects a case (patient) using menu driven selection criteria. A menu of case information is then presented. At this point, most subjects choose to read the medical history and observe video clips of the patient speaking with a therapist. Next, a set of tutorial questions related to the video clips become available to the student. Following this, the learner is required to formulate a clinical hypothesis based on her initial observations. The

[1] `http://www.patsy.ac.uk/ or http://patsy.cogsci.ed.ac.uk` (n.b. the system is currently being expanded to include 3 new domains - childhood dyslexia, medical rehabilitation and cognitive neuropsychology in collaboration with York, Edinburgh and Cambridge Universities.)

student then has the option of ‘administering’[2] a test to the patient by selecting the desired test from a structured menu of test titles or viewing a summary table of the patient’s total test scores for various tests. As mentioned above, a wide range of test data is available for each patient, consisting of medical history data plus assessment data based on the presentation of visual and auditory stimuli to the patient.

Users are able to simulate the administration of a test to a patient by stepping through any test at item-by-item level[3] (ie. finest possible testing granularity). An example - the ‘TROG’ test - is provided in Figure 1.

Figure 1: Item from TROG - Test of Reception of Grammar (Bishop, 1989). The user can click the ‘Listen to stimulus’ button and hear the spoken stimulus presented to patient (‘The horse is chased by the man’), and then see the patient’s pointing response (simulated by an arrow revealed beside one of the stimulus pictures when ‘Show patient response’ button is clicked).

Throughout a PATSy training session students are encouraged to state their model-derived hypotheses in advance, to reflect on their conclusions and to explicitly state what they plan to do next based on those conclusions (figure 2).

In this study, our aims were a) to assess PATSy as a research tool in the study of clinical reasoning, b) to investigate students’ clinical reasoning and that of experts, c) to learn more about the relative contributions of domain-specific knowledge and general reasoning ability

[2]In a virtual, simulated sense

[3]Researchers and students in advanced training are able to by-pass the item-by-item analysis format of PATSy and proceed straight to a summary of the patient’s total scores for each test.

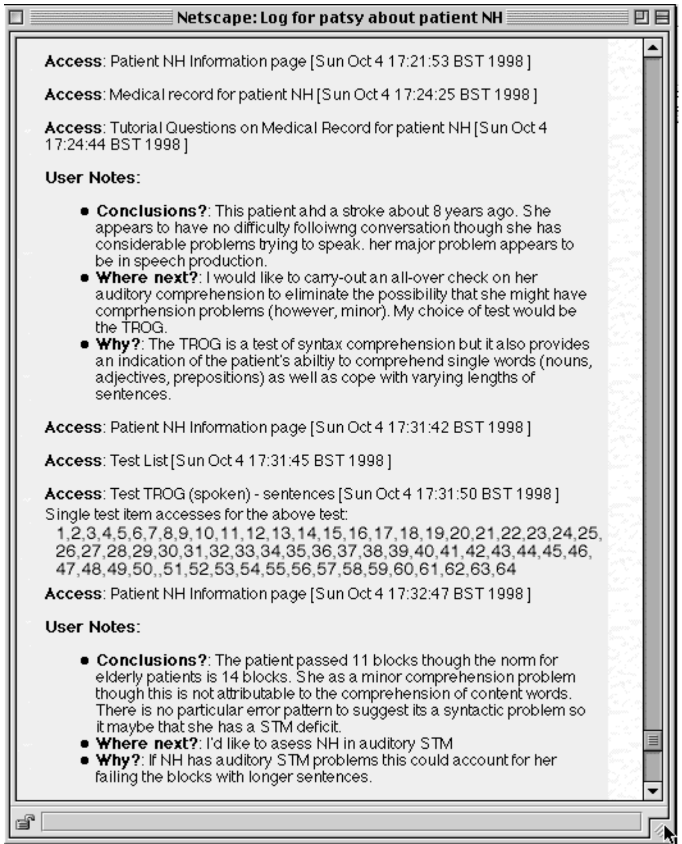

Figure 2: Example of PATSy log - note system generated data with student's own notes interleaved.

to diagnostic performance on PATSy and d) to investigate which of these component skills are amenable to scaffolding via the addition of intelligent support to PATSy.

2 The study

The study was conducted on a class of 38 3rd year speech therapy students at Queen Margaret University College, Edinburgh. The lecturer closely integrated PATSy into her course on cognitive neuropsychological approaches to acquired language disorders. Students were referred to numerous PATSy cases during the semester in lectures, and were expected to use PATSy as a course requirement. Participants were informed that their participation in the study would be used to assist the further development of the PATSy system.

Three expert participants were also included - they were speech and language science clinician/educators who used PATSy to diagnose the previously unseen examination case (DBL).

2.1 Pre-course tests: General reasoning

All student participants were administered 3 tests of general reasoning.

River pollution task Search heuristics are an important part of test-based diagnostic reasoning. Yet search skills are not directly taught to students and they may therefore be expected

The ADA minimal pairs test tests:	In the ADA minimal pairs tests, the person:
a) semantics	a) sees two written stimuli
b) phonological discrimination	b) hears two rhyming words
c) hearing	c) hears two words or non-words
d) word recognition	d) hears two phonemes

Table 1: Sample items from the DK test.

to differ widely in ability. As a measure of search skill, participants attempted a computer-based 'River pollution task' [2]. In this task the goal is diagnose which of 16 factories is responsible for polluting a river. There are 24 different chemicals in total, with each factory discharging between 3 to 10 chemicals. A single water analysis tests for one chemical. Chemical testing has an associated cost. Efficient solution requires a a binary-split strategy. The computer recorded, for each participant, the time taken to 'diagnose' the guilty factory, number of chemical 'tests' required to isolate factory and the number of factory guesses.

Deductive reasoning - Wason 4-card selection task The Wason 4-card task is a much-used reasoning task [19]. The original abstract version was used. Subjects are shown 4 cards, each depicting letters or numbers ('A', 'K', '4', '7'). They are informed of the rule 'If a card has a vowel on one side, it has an even number on the other' ($P \rightarrow Q$). They are then asked to indicate all and only those cards they would turn over in order to verify the rule. The cards that *must* be turned are A and 7 ('P' and 'not-Q'). A second 'realistic' law version was also used. In this version, cards depicting 'drinking beer', drinking coke', 16 years of age, 22 years of age) were presented along with the rule 'If a person is drinking beer, then the person must be over 18 years old'. The scoring system produced a score range of -2 to +2. For example, a score of +2 was awarded for both correct cards, +1 for 1 of the 2 correct cards, -1 for one incorrect card turned, etc.

Analytical and verbal reasoning tests The analytical reasoning test was based on an instrument ('GRE') used for graduate school selection in the US. This test has more realistically themed and 'puzzle-like' items than tests such as the Wason 4-card selection task. Examples of GRE-type test items can be found in [5]. The *analytical* sub-scale poses constraint satisfaction puzzle problems for which diagrams are often useful aids to solution. The *verbal* sub-scale poses verbal argument analysis exercises (eg. assessing the credibility of evidence). The 'GRE' test used in this study contained 5 verbal items and 17 analytical items.

2.2 Pre-course tests: Domain-specific knowledge (DK)

Test of test knowledge and theory This consisted of a test with 2 sections. The first is a multiple-choice test of students' knowledge of neuropsychological tests. The 33 item test[4] assesses students' knowledge of what the test measures (test purpose) and the stimulus and response modalities (auditory, visual, motor) that the test employs (test format). Typical items are shown in Table 1.

[4] Developed by Dr. Julie Morris

The domain knowledge test also assessed the students' pre-course understanding of the cognitive neuropsychological model of normal language production [16]. Students were asked to draw a 'box and arrrow' diagram of the model. This was scored by awarding points for correct components (labelled boxes) and directional links (arrows). The total DK score consisted of the test knowledge score plus the theory model diagram score.

2.3 During-course measures

As mentioned above, PATSy was not used in a 'stand alone' manner, but was integrated into a course of lectures. For each student, PATSy log archives allowed several parameters to be derived - number of cases 'seen' by each student during the course, time spent per case, test selections by student for each case, typed comments in response to 'Conclusions?', 'Why?' and 'Where next?' prompts by PATSy, etc. The average time spent per PATSy case was 78 minutes (s.d = 47.29, range 21 to 201 minutes). The average number of PATSy cases accessed by students during the course was 4 (s.d. = 1.3, range 2 to 6 cases). Hence each student typically logged around 5.3 hours of PATSy interaction over the course semester prior to the unseen exam case 'DBL'.

2.4 Post-course assessment

Diagnosis of a previously unseen case The case (DBL) used for the examination has lexical access ('word finding') difficulties in speech production. Three experts and the 38 student subjects used PATSy[5] in examination mode. PATSy logged which tests were administered, with time stamps. No time limit was imposed. All students had used PATSy during their 3rd year aphasiology training prior to the exam. Students were informed that the PATSy logs would be marked and would 'count' towards their assessment. Detailed PATSy protocols (similar to the one shown in Figure 2) were gathered from students and experts as they diagnosed a previously unseen patient's language disorder. Experts (clinician educators) also performed the task.

Exam case assignment marks Students submitted a 2000-word written assignment based on the PATSy exam case. The students were instructed to "Analyze and interpret DBL's data to give an account of his speech production problems in accordance with a cognitive neuropsychological perspective." A major component of the mark awarded was based on the lecturer's assessment of how well the student argued from a cognitive neuropsychological perspective. The emphasis was upon argument and comprehensive consideration of all relevant cognitive subsystems. Marks were also awarded for student responses to the log prompts. The coursework assessment mark formed the dependent variable for the regression analyses reported below.

[5] One of the clinician educators provided a hand-written protocol.

Subject	E1	E2	E3	S6	S10	S30
Sequence						
1	PS(Pt)	PS(Pt)	AMPSSy(Pt)	AMPSSy(Pt)	AMPSSy(Pt)	A(In)
2	VS(In)	VS(Pt)	A(In)	A(In)	A(Sp)	AMPSSy(Pt)
3	AS(In)	SP(Sp)	A(In)	A(In)	A(In)	A(In)
4	SP(Sp)	A(Sp)	PS(Pt)	PS(Pt)	V(In)	A(In)
5	A(Sp)		SP(Sp)	AS(In)	V(Sp)	A(In)
6			A(Sp)	A(Sp)	V(Sp)	AS(In)
7				SP(Sp)	VPS(Pt)	PS(Pt)
8				V(In)	PS(Pt)	V(Sp)
9					SP(Sp)	A(Sp)
10					VM(In)	V(Sp)
11					VM(Sp)	V(Sp)
12					AM(In)	V(Sp)
13						V(In)

Table 2: Examples of TDL representations of subjects' testing sequence for case DBL in the course of diagnosis. E1, E2, E3 = clinician educators (experts), S6, S10 & S30 are 3rd year students (*n.b.* shorter than average student test sequence examples chosen to save space - mean test sequence length for students = 20.6, s.d. = 9.2). Key to coding scheme: capital letters indicate stimulus modality(ies) of test; letter in brackets indicates mode in which patient responds. Test stimulus modality codes: A=auditory, V=visual (eg written word), P=picture, S=semantic, Sy=syntax. Patient response modalities (in brackets): In=indicate (eg raise finger), Sp=spoken, Pt=point.

3 Results

3.1 Results: PATSy log data & the test description language (TDL)

In order to analyse test selection sequences recorded in the PATSy logs, it was necessary to represent the numerous tests and what they measure at a more abstract level. A 'test description language' (TDL) was therefore devised. This was for two reasons. Different tests may assess similar cognitive processes and may be interchangeable for some purposes. Secondly, TDL representation makes comparison of individual reasoning traces easier.

The TDL represents a test in terms of the cognitive modalities it assesses and the mode of response required from the patient. An example of the scheme in use is shown in in Table 2. The TROG (figure 1) would be coded 'APSSy(Pt)'. Log data processing also yielded quantitative scores consisting of the number of times text was entered by each student in response to each of the 'Conclusion?', 'Why?' and 'Where next?' prompts.

The experts' protocols, and the test administration sequence data (Table 2) showed that their performance on the diagnosis task is characterised by efficient search in which broader tests are used to cut a broad swathe through the search space early in the investigation. The expert's testing approach is informed by the video of the patient in conversation, the medical history, expert knowledge of the information processing model and fluency with a wide range of tests. Typically, broader tests are followed by finer, more specific tests. Experts efficiently explore areas of the information processing model, judiciously deploying coarse or finer grained tests as they proceed. As the expert protocols show, initially they search the information processing model in terms of modality of processing and whether the source of the difficulty is input, central or output. They tend to finish by focussing on the 'grain size' ie word level, sentence level of information that patient can accommodate.

Taking the experts' traces together, an expert sequence was derived consisting of a broad test like the TROG first [AMPSSy(Pt)] followed by either an auditory lexical decision test or an auditory discrimination test [both coded as A(In)] followed by a picture naming test [SP(Sp)] and 2 further subsequent tests. This yields an 'optimal' 5-test hybrid expert trace of: AMPSSy, A(In), SP(Sp), PS(Pt) or VS(Pt), A(Sp) or A(Sp) or V(Sp).

3.2 Log analyses

The PATSY system log protocols and TDL sequences revealed that the students differed strikingly in their diagnostic strategies. Students varied in the amount of tests that they 'administered' to the exam patient. The TDL sequences also revealed great variability in terms of test sequencing and the order in which components of the language production model were tested. Compared to the experts' average of 5 tests to reach diagnosis, the student average (mean) test sequence length was 20.6 (s.d. 9.2, range 6 - 41 tests).

A comparison of the students' TDL sequences with those of the experts provided a basis for estimating the number of tests conducted by each student, together with scores for test ordering accuracy. 'Spurious' tests were defined as tests that the experts did not use at all in their reasoning sequence. A median split on the 'spurious' test variable yielded 2 groups of students - non-prolific testers and prolific testers (n=20 and 18 respectively). However, these two groups did not differ significantly in terms of their coursework assignment marks for the examination case (13.45 versus 14.56 marks). The groups differ more in terms of their reasoning strategies and style than in their performance on the exam assignment.

3.3 Test data analyses

Quantitative scores from the domain-independent and domain-knowledge pre-test consisted of: Wason abstract version (WASONA), Wason realistic version (WASONR), analytical reasoning (GREA), verbal reasoning (GREV), GRE speed (number of gre items attempted in time allowed - GREATT) and river pollution task (RTIME, RTESTS, RFACTORY (number of factory guesses).

Space prohibits comprehensive reporting of results on all the pre-test measures. Regarding the Wason test results, however, these were well aligned with many previous findings (e.g. [4]) with fewer than 10% of the students making correct card selections on the abstract version of the task. Also in line with much previous research, 'affirming the consequent' (selecting 'A' and '4') was by far the most common response pattern (68%). On the natural law or realistic (drinking age rule) version, 66% of respondents selected correct cards (drinking beer, 16 years of age), showing the usual dramatic effect of making the rule context more 'concrete' [4].

A bivariate correlation matrix was examined and three variables were found to correlate positively and significantly with assignment mark - pre-course domain knowledge (DK) scores (.41*[6]), the number of cases viewed on PATSy during the course (.39*), and the abstract version of the Wason 4-card selection task score (.34*). These variables had been predicted *a priori* to correlate positively with clinical reasoning performance. The 3 variables were input as predictor variables into an exploratory regression analysis.

[6] * = significant at $p < .05$ (2 tailed)

A multiple linear regression (MLR) model was derived using the SPSS MLR procedure with stepwise forward entering of predictor variables. The dependent variable was the written coursework assignment mark. The resultant regression model was significant at $p<.02$, with a multiple R = .42. The model accounted for 15% (adjusted R squared) of the variance in assignment marks. The stepwise model contained one significant predictor variable - number of cases accessed on PATSy during the course ($p<.02$). The remaining variables, which were not included in the final equation by the stepwise regression procedure but which were very close to meeting the $p<.05$ criterion for inclusion, were domain knowledge test score (p=.054) and performance on the Wason 4-card selection task (p=.06). When these variables were included in the regression model, the multiple R increased to .61 ($p<.004$), accounting for 30% (adjusted R squared) of assignment mark variance.

4 Discussion

Of the three reasoning pre-tests (analytical reasoning, Wason selection and river pollution) only the Wason task predicted performance on the dependent measure.

Experience with the PATSy system in terms of the number of patient cases experienced during the course had a positive and generalised effect on learning in terms of the written end-of-course assessment assignment. Marks were awarded on the assignment for arguments based on hypothesis testing, evidence from theory and reasoning from a cognitive neuropsychological perspective.

The results suggest that domain-specific knowledge is a major factor in learning to reason in a hypothesis-driven manner. This is not surprising. Knowledge of tests and of theory are essential as a basis for hypothesis-driven deductive reasoning. It seems surprising that pre-course knowledge should emerge as a significant component of the predictive model. However, the students were in their 3rd year and had some experience of tests and assessment, though not necessarily from a cognitive neuropsychological perspective.

The fact that scores on the Wason task correlated positively with the assignment marks was interesting. Most students' responses on the Wason task were of a 'rule affirmation' rather than of a 'rule disconfirmation' type. This accords with conditional reasoning practised clinically where confirmation for theory-driven hypotheses is usually sought in terms of the 'if-then' rule [17].

Although other tests of general reasoning ability (analytical reasoning, river pollution task) were predicted to correlate positively with the assignment mark, they did not. Analytical reasoning problems are generally of the constraint satisfaction puzzle variety. On reflection, this type of reasoning is little-exploited clinically.

The results also showed that prolific and non-prolific testers showed similar average marks on the assignment. This was probably because the students were credited for arguments even when these were 'flanked' by spurious tests. Future work aims to explore the relationships between the domain-specific and general reasoning tests and reasoning strategies represented by subjects' TDL traces. Of interest is the content and nature of reflective commenting in the logs - is a particular type of reflective comment (self-explanatory?) associated with more efficient reasoning, for example.

In our study, the experts used broad spectrum testing early in diagnosis to rapidly reduce the size of the search space. This results in the generation of more general (and fewer) hypotheses than novices might produce - a pattern consistent with previous research in medical diagnostic reasoning [6,10,18], though in our study it was not established that case-based reasoning was used by the experts.

As stated earlier, novices lack an extensive case-base to influence their reasoning, and so are required by necessity to rely predominantly upon deductive reasoning. PATSy's design encourages deductive reasoning and the aim of adding intelligent support to PATSy would be to better facilitate the development of this mode of clinical reasoning by scaffolding deficits in domain knowledge (theory, tests, etc) and general reasoning. Learner support could be provided for some of the diagnostic sub-skills using a modular design approach. On line (look-up) reference material on information processing accounts of normal and pathological language and on tests could be made available. Online simulation-based exercises such as the River Pollution task [2] could be provided for practising basic search strategies.

Support for diagnosis could take the form of feedback based on model-tracing [1]. The TDL representation of expert (or more advanced student) testing sequences could be used as a basis for evaluating the user's selections and providing interventions such as hints or guidance on test selection and sequencing. Such an intelligent tutoring approach would also require domain knowledge representations of the IP model (with the patient's pathology over-layed) and an extended-TDL representation of the tests. PATSy could detect inappropriate test selection in terms of a) the diagnostic utility of a test's input and response modalities, b) the cognitive demands on the patient that they pose and, c) the semantic level of test items and the granularity of test items. PATSy's test description language (TDL) scheme needs to be extended in order for it to also encode the type of test stimulus material (*i.e.* phoneme, word (or non-word), sentence). This will be one focus of future work.

General reasoning skills can be effectively taught (e.g. [8,15]). A *post-hoc* exercise in teaching various reasoning skills, such 'recognising arguments', 'identifying conclusions' and 'identifying reasons' using practise material from the MENO Thinking Skills Inventory [14], showed that PATSy students could master these argumentation skills quickly. It remains to be seen, however, whether these benefits generalise to diagnostic reasoning contexts on PATSy. If so, then suitable exercises could be made available on-line as part of the PATSy system.

In conclusion, the results of this study indicate that PATSy is a useful research tool for the study of clinical reasoning. Used in conjunction with selected domain knowledge and reasoning tests, PATSy provides a structured and controlled context for the study of clinical reasoning.

5 Acknowledgements

The PATSy project is funded by the Nuffield Foundation whose support is gratefully acknowledged. Many thanks to Dr. Julie Morris for development of the test knowledge test and for multimedia data preparation. Thanks too to the 3rd year speech and language science class of 1998/1999 at Queen Margaret College, Edinburgh. Valuable assistance and patient data contributions to PATSy by the Department of Speech, University of Newcastle and the Departments of Psychology, Universities of York and Cambridge are also gratefully acknowledged. Finally, thanks to the anonymous reviewers for useful comments and suggestions.

6 References

[1] Anderson, J. R., Conrad, F.G. & Corbett, A.T. (1993) The LISP tutor and skill acquisition. In J.

Anderson (Ed.) *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[2] Broadbent, D.E. & Berry, D.C. (1987) Explanation and verbalization in a computer-assisted search task. *The Quarterly Journal of Experimental Psychology*, **39A**, 585–609.

[3] Bishop, D. (1989) TROG: Test of reception of grammar. Department of Psychology, University of Manchester, UK.

[4] Cheng, P.W. & Holyoak, K.J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology*, **17**, 391–416.

[5] Cox, R. (1999) Representation construction, externalised cognition and individual differences. *Learning and Instruction*, **9**, 343–363.

[6] Cuthbert, L., du Boulay, B., Teather, D., Teather, B., Sharples, M., & du Boulay, G. (1999) Expert/novice differences in diagnosic medical cognition - A review of the literature. *Cognitive Science Research Papers CSRP 508*, School of Cognitive & Computing Sciences, University of Sussex.

[7] Ellis, A.W. & Young, A.W. (1996) *Human Cognitive Neuropsychology*. Hove, UK: Psychology Press.

[8] Halpern, D.F. (1996) *Thought and Knowledge: An introduction to critical thinking.* Mahweh, NJ: Lawrence Erlbaum.

[9] Holland, J.H., Holyoak, K.J., Nisbettt, R.E. & Thagard, P.R. (1986) *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

[10] Jonassen, D.H. (1996) Scaffolding diagnostic reasoning in case-based learning environments. *Journal of Computing in Higher Education*, **8(1)**, 48–68.

[11] Luger, G.F. & Stubblefield, W.A. (1989) *Artificial Intelligence and the design of expert systems*. Redwood City, CA: Benjamin/Cummings.

[12] Lum, C. & Cox, R. (1998) PATSy - A distributed multimedia assessment training system. *International Journal of Communication Disorders*, **33**, 170–175.

[13] Lum, C., Cox, R., Kilgour, J., Morris, J. & Tobin, R. (1999) PATSy: A multimedia distributed web-based resource for aphasiologists in research and education. *Aphasiology*, **13(7)**, 573–579.

[14] MENO test (1993) Critical thinking component. Cambridge, UK: University of Cambridge Local Examinations Syndicate.

[15] Nickerson, R.S. (1994) The teaching of thinking and problem solving. In R.J. Sternberg (Ed.) *Thinking and Problem Solving*. San Diego, CA: Academic Press.

[16] Patterson, K. & Shewell, C. (1987) Speak and spell: Dissociations and word-class effects. In M. Coltheart, G. Sartori & R. Job (Eds.), *The Cognitive Neuropsychology of Language*, Hillsdale, NJ: Lawence Erlbaum Associates.

[17] Robertson, S.I. (1999) *Types of thinking*. London: Routledge.

[18] Sisson, J., Donnelly, M., Hess, G., & Woolliscroft, J. (1991) The characteristics of early diagnostic hypotheses generated by physicians (experts) and students (novices) at one medical school. *Academic Medicine*, **67**, 130–132.

[19] Wason, P.C. (1966) Reasoning. In B. Foss (Ed.), *New Horizons in Psychology*. Harmondsworth: Penguin, 135–151.