A Cognitive Tutoring System for Visual Diagnostic Classification

Reasoning, Representation and Reification

Rebecca Crowley
University of Pittsburgh School of Medicine

April 7, 2004
Intelligent Tutoring System

Empirical Research

- Expert Model
- Student Model
- Developmental Model
- Pedagogic Model

How do these skills develop naturally?

Real-time data obtained during the learning process

Test different approaches in a controlled environment

What strategies can be used to learn these skills and what determines their effectiveness?
Background

- Numerous previous theories about and studies of expertise in medical domains that are not primarily visual
  - Heuristic classification (Clancey, 1985)
  - Forward versus backwards reasoning (Patel, 1986)
  - Ill-structured problem space (Pople, 1982)
- Several previous studies of expertise in radiology and dermatology (Lesgold, 1988; Kundel, 1983; Norman 1989)
- No previous studies of expertise in microscopic diagnosis
- Microscopic diagnosis may be similar in some respects and different in other respects
Methods

- Think-aloud protocols
- Standardized case set in Breast Pathology
- Videotape collected from the microscope
- Determine accuracy – specific and categoric
- Code processes (operators) and content (knowledge states) using process coding scheme
- Identify errors using error coding scheme
- Measure times to important events
- Aggregate and compare codes and times for level, or case
- ANOVA, Scheffe tests to measure differences
Subjects

- 10 Novices – just finished year 2
- 10 Intermediates – 2\textsuperscript{nd} and 3\textsuperscript{rd} year residents
- 8 Experts – Attending pathologists with >10 years experience
Data Collection
<table>
<thead>
<tr>
<th>Case</th>
<th>Gold Standard Diagnosis</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Infiltrating Ductal Carcinoma</td>
<td>Focal lesion of poorly differentiated cancer adjacent to biopsy site and scar.</td>
</tr>
<tr>
<td>2</td>
<td>Ductal Carcinoma in Situ (DCIS)</td>
<td>Widespread solid and cribriform in-situ carcinoma present throughout majority of sample.</td>
</tr>
<tr>
<td>3</td>
<td>Infiltrating Lobular Carcinoma</td>
<td>Widespread classical type infiltrating lobular carcinoma. Scant adjacent normal tissue.</td>
</tr>
<tr>
<td>4</td>
<td>Lobular Carcinoma in Situ (LCIS)</td>
<td>Small focus of LCIS with retrograde extension in otherwise normal breast.</td>
</tr>
<tr>
<td>5</td>
<td>Fibroadenomas, Sclerosing Adenosis and Intraductal Papilloma</td>
<td>Multiple focal lesions, including sclerosing adenosis - a benign lesion that shares some visual features with cancer.</td>
</tr>
<tr>
<td>6</td>
<td>Paget’s Disease</td>
<td>Nipple with focal area of intra-epidermal Paget’s disease. No underlying carcinoma.</td>
</tr>
<tr>
<td>7</td>
<td>Adenomyoepithelioma</td>
<td>Small circumscribed lesion with uniform features.</td>
</tr>
<tr>
<td>8</td>
<td>Atypical Papilloma</td>
<td>Large lesion with numerous atypical features.</td>
</tr>
</tbody>
</table>
An expert diagnosing DCIS:

**Expert E7:**

3. Okay let’s look at low power
4. I think the tissue is breast → **identify-anatomic-location**
5. I recognize some normal
6. Here is in situ carcinoma → **statement-of-hypothesis**
7. I have to find out if there is any invasion → **set-goal-identify**
An intermediate diagnosing DCIS

Intermediate I5:

20 and some of the ducts that are expanded with small cells with focal, possibly central, area of necrosis.
4 So just scan this slide around and try to determine some focal areas that I want to concentrate and focus on.
5 Now I'm looking at some of the ducts that are expanded.
25 And some of these ducts, they also have holes, and these are sort of punched-out holes,
26 very uniform, which...
29 So at this magnification, I think it is a DCIS.
Overall accuracy

![Accuracy Chart]

- Novice
- Intermediate
- Expert

**Accuracy**

- Specific
- Category
# Process Differences

<table>
<thead>
<tr>
<th>Measure</th>
<th>Novice Mean</th>
<th>Novice SD</th>
<th>Intermediate Mean</th>
<th>Intermediate SD</th>
<th>Expert Mean</th>
<th>Expert SD</th>
<th>F Value</th>
<th>P Value</th>
<th>Pairwise Comparison</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td>14.1</td>
<td>5.4</td>
<td>20.5</td>
<td>8.2</td>
<td>8.0</td>
<td>2.8</td>
<td>9.4</td>
<td>.001</td>
<td>N, I</td>
<td>.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, I</td>
<td>.132</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.077</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, I</td>
<td>.511</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.741</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.191</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, I</td>
<td>.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.794</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.027</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, I</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.084</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.153</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, I</td>
<td>.119</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.408</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, I</td>
<td>.094</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.981</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, I</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.210</td>
</tr>
<tr>
<td>Data exploration and explanation</td>
<td>2.6</td>
<td>1.6</td>
<td>4.5</td>
<td>2.4</td>
<td>2.0</td>
<td>1.6</td>
<td>4.4</td>
<td>.023</td>
<td>N, I</td>
<td>.000</td>
</tr>
<tr>
<td>Data interpretation</td>
<td>4.9</td>
<td>1.6</td>
<td>11.4</td>
<td>4.1</td>
<td>8.4</td>
<td>3.6</td>
<td>9.8</td>
<td>.001</td>
<td>N, I</td>
<td>.000</td>
</tr>
<tr>
<td>Operational processes</td>
<td>2.6</td>
<td>2.1</td>
<td>4.4</td>
<td>1.8</td>
<td>1.5</td>
<td>1.6</td>
<td>5.4</td>
<td>.011</td>
<td>N, I</td>
<td>.119</td>
</tr>
<tr>
<td>Goal-setting</td>
<td>0.4</td>
<td>0.4</td>
<td>1.7</td>
<td>1.3</td>
<td>1.8</td>
<td>2.0</td>
<td>3.3</td>
<td>.052</td>
<td>N, I</td>
<td>.094</td>
</tr>
<tr>
<td>Unique hypotheses</td>
<td>2.2</td>
<td>0.6</td>
<td>4.0</td>
<td>0.9</td>
<td>3.3</td>
<td>1.2</td>
<td>11.3</td>
<td>&lt;.001</td>
<td>N, I</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.210</td>
</tr>
</tbody>
</table>

Crowley et al. JAMIA, 2003
<table>
<thead>
<tr>
<th>Process</th>
<th>Identification</th>
<th>Novice Mean</th>
<th>Novice SD</th>
<th>Intermediate Mean</th>
<th>Intermediate SD</th>
<th>Expert Mean</th>
<th>Expert SD</th>
<th>ANOVA F Value</th>
<th>P Value</th>
<th>Cohen's Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Examination</td>
<td>Identify tissue configuration</td>
<td>0.1</td>
<td>0</td>
<td>0.4</td>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>&lt;0.001</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Identify structure</td>
<td>5.6</td>
<td>4</td>
<td>3.1</td>
<td>2.5</td>
<td>0.6</td>
<td>1</td>
<td>28.4</td>
<td>&lt;0.001</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Identify anatomic location</td>
<td>1.9</td>
<td>2</td>
<td>1.7</td>
<td>1.4</td>
<td>0.9</td>
<td>0</td>
<td>5.3</td>
<td>0.006</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Locate lesion</td>
<td>0.2</td>
<td>0</td>
<td>0.3</td>
<td>0.6</td>
<td>0.5</td>
<td>1</td>
<td>3.7</td>
<td>0.028</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Identify procedure</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
<td>0.2</td>
<td>0.06</td>
<td>0</td>
<td>1.4</td>
<td>0.262</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Identify descriptive cue</td>
<td>1.7</td>
<td>2</td>
<td>0.5</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>21.1</td>
<td>&lt;0.001</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Identify histopathologic cue</td>
<td>4.6</td>
<td>4</td>
<td>13</td>
<td>7.3</td>
<td>6</td>
<td>6</td>
<td>22.3</td>
<td>&lt;0.001</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Quick Eval</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>1.2</td>
<td>0.2</td>
<td>1</td>
<td>8</td>
<td>0.001</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Identify margins</td>
<td>0.2</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>1</td>
<td>5.1</td>
<td>0.008</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Rule out finding</td>
<td>0.1</td>
<td>0</td>
<td>0.5</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>8.9</td>
<td>&lt;0.001</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Identify artifact</td>
<td>0.1</td>
<td>0</td>
<td>0.29</td>
<td>0.8</td>
<td>0.2</td>
<td>1</td>
<td>1.1</td>
<td>0.346</td>
<td>0.83</td>
</tr>
<tr>
<td>Compare</td>
<td>Compare to normal</td>
<td>1</td>
<td>1</td>
<td>1.4</td>
<td>1.5</td>
<td>0.3</td>
<td>1</td>
<td>6.5</td>
<td>0.002</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Compare to expected</td>
<td>0.08</td>
<td>0</td>
<td>0.05</td>
<td>0.2</td>
<td>0.06</td>
<td>0</td>
<td>0.1</td>
<td>0.924</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Compare at case level</td>
<td>0.6</td>
<td>1</td>
<td>0.5</td>
<td>1.2</td>
<td>0</td>
<td>0</td>
<td>4.3</td>
<td>0.016</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Compare findings from multiple locations</td>
<td>0.6</td>
<td>1</td>
<td>1.2</td>
<td>2</td>
<td>0.4</td>
<td>1</td>
<td>3.4</td>
<td>0.037</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Determine degree of abnormality</td>
<td>0.05</td>
<td>0</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>1</td>
<td>3.4</td>
<td>0.038</td>
<td>1</td>
</tr>
<tr>
<td>History</td>
<td>Read</td>
<td>0.6</td>
<td>1</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>1</td>
<td>1.8</td>
<td>0.178</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Identify historical cue</td>
<td>0.6</td>
<td>1</td>
<td>0.6</td>
<td>0.7</td>
<td>0.3</td>
<td>0</td>
<td>3.2</td>
<td>0.047</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Relate diagnosis to known/possible history</td>
<td>0.2</td>
<td>1</td>
<td>0.5</td>
<td>0.8</td>
<td>0.4</td>
<td>1</td>
<td>2.3</td>
<td>0.104</td>
<td>0.9</td>
</tr>
</tbody>
</table>
## Task Errors

<table>
<thead>
<tr>
<th>Errors Case Level</th>
<th>Description</th>
<th>Novice Errors coded as present or absent in each case</th>
<th>Intermediate Errors coded as present or absent in each case</th>
<th>Expert Errors coded as present or absent in each case</th>
<th>Statistics</th>
<th>Pairwise Comparison</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Number of cases / Total (%)</td>
<td>Number of cases / Total (%)</td>
<td>Number of cases / Total (%)</td>
<td>Chi-Square</td>
<td>P Value</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Lesion never brought under objective</td>
<td>8/30 (26.7 %)</td>
<td>1/28 (3.6 %)</td>
<td>0/23 (0 %)</td>
<td>13.25</td>
<td>.0021</td>
<td>N, I</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
</tr>
<tr>
<td>2</td>
<td>Lesion traversed without recognition</td>
<td>7/30 (23.3 %)</td>
<td>0/28 (0 %)</td>
<td>0/23 (0 %)</td>
<td>8.11</td>
<td>.0209</td>
<td>N, I</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
</tr>
<tr>
<td>3</td>
<td>Error in identifying anatomic location</td>
<td>14/40 (35.0 %)</td>
<td>2/38 (5.3 %)</td>
<td>0/32 (0 %)</td>
<td>22.76</td>
<td>&lt;.0001</td>
<td>N, I</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
</tr>
</tbody>
</table>

### Segment Level

<table>
<thead>
<tr>
<th>Segment Level</th>
<th>Errors counted for each case</th>
<th>Mean/SD</th>
<th>Mean/SD</th>
<th>Mean/SD</th>
<th>F Value</th>
<th>P Value</th>
<th>Pairwise Comparison (TukeyHSD)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Incorrectly names normal structure</td>
<td>0.35 0.74</td>
<td>0.11 0.31</td>
<td>0.0</td>
<td>5.28</td>
<td>.012</td>
<td>N, I</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.656</td>
</tr>
<tr>
<td>5</td>
<td>Incorrectly names histopathologic cue</td>
<td>0.93 1.4</td>
<td>0.76 1.08</td>
<td>.003 0.18</td>
<td>7.05</td>
<td>.004</td>
<td>N, I</td>
<td>.669</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.004</td>
</tr>
<tr>
<td>6</td>
<td>Error in assigning significance,</td>
<td>0.48 0.82</td>
<td>0.32 0.57</td>
<td>.003 0.18</td>
<td>3.17</td>
<td>.059</td>
<td>N, I</td>
<td>.552</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N, E</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E, I</td>
<td>.298</td>
</tr>
</tbody>
</table>

Crowley et al. JAMIA, 2003
Intermediates detect but cannot accurately classify.
Task Analysis Latencies

Crowley et al. JAMIA, 2003
Developmental Model

**Novice**
- AL not attempted
- Failure of search and detection
- Few features
- No refinement
- Few hypotheses
- No testing
- Diagnostic Errors

**Intermediate**
- AL correct
- Successful search and detection
- Many features
- Many errors
- Uncertainty
- Many comparisons
- Many Hypotheses (wide)
- Errors in Testing
- Work backwards & Forwards
- Diagnostic Errors

**Expert**
- AL correct
- Successful search and detection
- Few features
- Many Hypotheses (narrow)
- Efficient Testing
- Work backwards & Forwards
- Diagnostic Accuracy

**Abstraction & Refinement**
Intelligent Tutoring System

Expert Model

Student Model

Developmental Model

Pedagogic Model

Forwards and backwards reasoning
Abstraction and refinement
Granular enough to support change

Model different types of skills

Visual and symbolic feedback
Encourage feature identification
Encourage hypothesis testing

effect of empirical model
Scaffolding skill acquisition

- **Searching skills**
  - Prevent students from interpreting non-diagnostic areas
  - Require that students see the entire slide
  - Monitor and provide feedback on specific search skills

- **Feature identification**
  - Encourage identification and full specification of evidence
  - Correct errors and give explanations
  - Combine visual and symbolic aspects of training

- **Hypothesis triggering and testing**
  - Provide feedback for forwards and backwards reasoning
  - Support efforts to distinguish between hypotheses

- **But allow enough flexibility so that system will support students who are in many stages of skill acquisition**
## Visual Features

- **SUPERFICIAL PERIVASCULAR DERMATITIS**
  - Lymphocytes predominate
  - Extravasated erythrocytes and/or siderophages
  - Eosinophils in variable number
  - Neutrophils prominent
  - Eosinophils prominent
  - Neutrophils and eosinophils prominent
  - Mast cells
  - Lymphocytes predominate
  - Ballooning and individual necrotic keratinocytes
  - Normal cornified layer
  - Corroded layer normal
  - Discrete food in papillary dermis

## Diagnoses

- **1/11 patterns**
  - Tinea versicolor
  - Dermatophytosis
  - Erythrasma
  - Pitted keratolysis
  - Vitiligo
  - Schamberg's disease
  - Viral exanthems of some kinds
  - Drug eruption (one type)
  - Erythema figuratum, superficial
  - Schamberg's disease
  - Pruritic unferial papules and plaques of pregnancy
  - Dermatitis herpetiformis
  - Linear (g.) dermatitis
  - Erythema annulare centrifugum
  - Discoid lupus erythematosus
  - Cellular granulomatous/abscess
  - Bullous pemphigoid/Herpes gestations, urticarial
  - Verruca vulgaris, urticarial
  - Anthropic assault, superficial
  - Pruritic unferial papules and plaques of pregnancy
  - Macula

- **Postinflammatory pigmentedary alteration**
  - Mucous amyloidosis
  - Stasis changes
  - Monocytone-induced eruption
  - Urticaria pigmentosa
  - Erythema multiforme
  - Mucha-Habermann disease
  - Graft vs. host reaction
  - Effects of nitrogen mustard on mycosis fungoides, patch/plaque
  - Drug eruption (one type)
  - Discoid lupus erythematosus
  - Dermatomyositis
  - Lichen sclerosus et atrophicus (morpha)
  - Postinflammatory pigmentary alteration
  - Fixed drug eruption, superficial
  - Lichen planus
  - Lichen planus-like drug eruption
  - Lichenoid panniculitis
  - Mucha-Habermann disease
  - Lichen striatus
  - Graft vs. host reaction
  - Lichenoid purpura of Gougerot and Blum
  - Lichen planus-like keratitis
  - Disseminated superficial actinic porokeratosis
  - Mycosis fungoides, plaque
  - Lichen nitidus
  - Sarcoïdosis
  - Langerhans' cell histiocytosis
  - Herpesvirus infection
  - Measles
  - Erythema multiforme
  - Graft vs. host reaction
  - Fixed drug eruption, superficial
  - Toxic shock syndrome
  - Cut-off
  - Miller's nodule

(continued)
Domain Model Class Structure

Case Data

After Motta et al, 2001
Representation and Reification

- Domain Knowledge
- Algorithm
- Case Knowledge

Abstract PSM

- Pedagogic Knowledge
- Problem-Space

Interface

- Tutoring System
- Tutoring System
- Tutoring System

Representation

Reification
Identified evidence cluster supports same hypotheses. More evidence cluster now differentiates A hypothesis… reasonable but wrong.

Cluster does not support Acute Burn.

“Backwards” reasoning.

A hypothesis… reasonable but wrong.
### Differences between conditions

<table>
<thead>
<tr>
<th>Case-focused training</th>
<th>Knowledge-focused training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutoring of knowledge within context of a single case</td>
<td>Tutoring of each case in context of the entire knowledge base</td>
</tr>
<tr>
<td>Student can focus on the relationships within a case</td>
<td>Students can ask and answer what-if questions outside of a single case</td>
</tr>
<tr>
<td>Allows students to explore problem-space in open ended fashion</td>
<td>Teaches the most efficient way to move through the problem-space</td>
</tr>
<tr>
<td>Cognitive load low for single case but high for integration</td>
<td>Cognitive load high for single case but low for integration</td>
</tr>
<tr>
<td>Students must build their own mental model</td>
<td>Students can use the tutor’s model</td>
</tr>
</tbody>
</table>

---

25/31
Effect of Reification

- **Multiple Choice Test 1A**
  - 8 total whole-case test cases:
    - 4 test-pattern cases: A1, B1, C1, D1
    - 4 control-pattern cases: N1, O1, P1, Q1
  - 20 total tutoring cases:
    - 4 test-pattern cases (2 instances each): A2, A3, B2, B3, C2, C3, D2, D3
    - 3 non-test pattern cases (2 instances each): E1, E2, F1, F2, G1, G2
    - 6 non-test pattern cases (1 instance each): H1, I1, J1, K1, L1, M1

- **Group 1: N=10**
  - Pre-Test
  - Interface 1 Training:
    - Work with Interface 1
  - Post-Test

- **Group 2: N=10**
  - Pre-Test
  - Interface 2 Training:
    - Work with Interface 2
  - Post-Test

  - 1 week

- **Retention Test**
  - Multiple Choice Test 1B
  - 8 total whole case test cases:
    - 4 test-pattern cases: A1, B1, C1, D1
    - 4 control-pattern cases: N1, O1, P1, Q1

  - 8 total whole-case test cases:
    - 4 new instances tutored patterns: A4, B4, C4, D4
  - 4 tutor cases: A2, B2, C2, D2

  - 1 week
Capturing Interaction Data

WEB SERVER
- Login Servlet
- Project DB
- Tutor Servlet
- Protocol Collection Filter
- Java Webstart Download Manager

Student Modeling System
- Student Files
- Probabilistic Student Model
- Pedagogic Model
- Pedagogic Production Rules
- Pedagogic Ontologies

Expert Module
- Dynamic Solution Graph
- Production Rules
- Domain Ontology
- Slide
- Pedagogic Ontology
- Jess
- Protégé-2000

 IMAGE DELIVERY SYSTEM
- Viewer Servlet
- Image Pump Application
- Whole-Slide Images
- Whole-Slide Images
- Case DB
- 2 Student Interfaces & 1 Authoring Interface

Client GUI Download with Java WebStart

WEB SERVER
- Login Servlet
- Project DB
- Tutor Servlet
- Protocol Collection Filter
- Java Webstart Download Manager

Student Modeling System
- Student Files
- Probabilistic Student Model
- Pedagogic Model
- Pedagogic Production Rules
- Pedagogic Ontologies

Expert Module
- Dynamic Solution Graph
- Production Rules
- Domain Ontology
- Slide
- Pedagogic Ontology
- Jess
- Protégé-2000

 IMAGE DELIVERY SYSTEM
- Viewer Servlet
- Image Pump Application
- Whole-Slide Images
- Whole-Slide Images
- Case DB
- 2 Student Interfaces & 1 Authoring Interface

Client GUI Download with Java WebStart
Hypotheses

• “Knowledge-focused” training results in:
  – *Better overall learning gains*, especially in:
    • Learning and applying relationships outside of cases tutored
    • Decreasing errors related to overgeneralization
  – *More efficient learning*
    • *Fewer hints needed, less time on task*

• “Case-focused” training results in:
  – *Better overall retention* of skills, especially in:
    • Learning and applying relationships learned within a case
    • Learning and applying visual recognition skills
Future Work

Study #1 - What is the relationship of reification to learning gains?
Study #2 - What student modeling formalisms are most predictive?
Study #3 - How does the timing of intervention effect learning?
Study #4 - How does the ITS compare to current practice?
Conclusions

- Studies of human performance can provide data for models needed in medical tutoring systems.
- Intelligent tutoring systems can provide useful laboratories in which to study basic questions about how expertise develops.
Acknowledgements

- **Expertise Studies**
  - Chuck Friedman
  - Gregory Naus
  - Jimmie Stewart
  - Elizabeth Legowski
  - National Library of Medicine Training Grant 5-T15-LM07059

- **Intelligent Tutoring System**
  - Olga Medvedeva
  - Eugene Tseytlin
  - Girish Chavan
  - Ellen Roh, Drazen Jukic, Scott Binder, George Xu, Sean Cowper
  - Elizabeth Legowski
  - National Library of Medicine 1R01 LM007891-01
Clinical Info

Biopsy from a 24 year old female with new onset vesicles, macules and papules, over trunk, buttocks and posterior neck. Also with Gluten-sensitive enteropathy.

Findings and Hypotheses

- Neutrophilic inflammatory infiltrate
  - LOCATION: dermal
  - QUANTITY: moderate

- Blisters
  - LOCATION: subepidermal

Diagnoses

Message Area
Dynamic Solution Graph

- Directed acyclic graph that generates valid paths through the problem space based on case data, domain model, task model.
- No predefined solution – each cycle generates the current state plus all valid next steps
- Expresses negated nodes, negated relationships, models forwards and backwards reasoning
- Cluster nodes – relate ALL Evidence to Hypotheses, and limits the need for disambiguation
- Student can abstract or refine evidence, hypotheses, and diagnoses or delete them – the graph changes appropriately
- Graph can model more complex events such as finding the features that best distinguish between multiple hypotheses
- Visualization of large, complex problem-space accomplished using JGraph
A General Architecture

(Crowley and Medvedeva, AMIA Proceedings 2003 modeled after Fensel et al, WICSA1, 1999)
Pedagogic Model Ontology

Mark-up language for hint and bug templates
Domain Model Ontology
<table>
<thead>
<tr>
<th>Process</th>
<th>Error Remediated by Tutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Identification</td>
<td>Evidence identified by student is not present</td>
</tr>
<tr>
<td></td>
<td>Evidence identified by student is explicitly absent</td>
</tr>
<tr>
<td></td>
<td>Evidence identified by student is not present, but can be present for some other hypothesis under consideration</td>
</tr>
<tr>
<td></td>
<td>Power used by student is too low to identify evidence</td>
</tr>
<tr>
<td></td>
<td>Evidence identified by student exists at another location</td>
</tr>
<tr>
<td></td>
<td>Evidence identified by student exists elsewhere, but evidence present in this location has been missed</td>
</tr>
<tr>
<td></td>
<td>Evidence not fully specified (too general)</td>
</tr>
<tr>
<td></td>
<td>Location of Absent Evidence identified by student is incorrect</td>
</tr>
<tr>
<td></td>
<td>Absent Evidence identified by student is not an important feature</td>
</tr>
<tr>
<td></td>
<td>Evidence identified by student as absent is present in this location</td>
</tr>
<tr>
<td></td>
<td>Evidence identified by student as absent is present in another location</td>
</tr>
<tr>
<td></td>
<td>Power used by student is too low to identify absent evidence</td>
</tr>
<tr>
<td>Feature Specification</td>
<td>Wrong quality for evidence</td>
</tr>
<tr>
<td></td>
<td>Evidence doesn’t require further specification of qualities</td>
</tr>
<tr>
<td></td>
<td>Quality can have that value for hypotheses under consideration but not in this case</td>
</tr>
<tr>
<td></td>
<td>Correct quality, but incorrect value for quality</td>
</tr>
<tr>
<td>Hypothesis triggering</td>
<td>No evidence for this hypothesis (early intermediate)</td>
</tr>
<tr>
<td></td>
<td>Hypothesis does not match all evidence (late intermediate)</td>
</tr>
<tr>
<td>Hypothesis Testing:</td>
<td>Evidence does not support/refute that hypothesis</td>
</tr>
<tr>
<td></td>
<td>Evidence modified no longer supports/refutes hypothesis</td>
</tr>
<tr>
<td></td>
<td>Insufficient argument made for diagnosis</td>
</tr>
</tbody>
</table>