

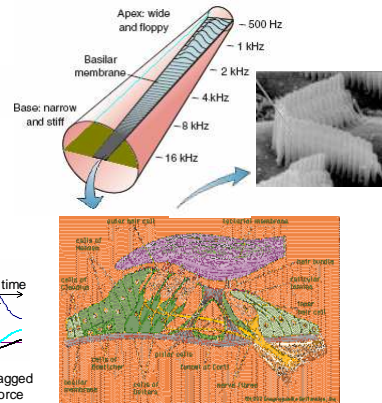
# Is there a Measure of Knowledge?

Tony Gardner-Medwin

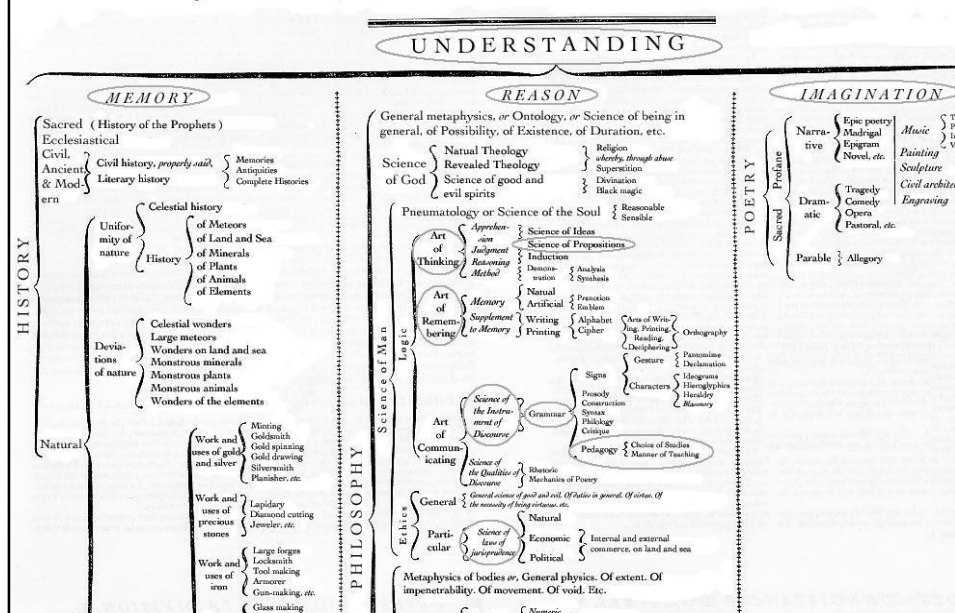
[a.gardner-medwin@ucl.ac.uk](mailto:a.gardner-medwin@ucl.ac.uk)

Emer. Prof. of Physiology, UCL

Nov 21, 2006: Lunch Hour Lecture



# MAP of the SYSTEM of HUMAN KNOWLEDGE



### ***Things I'm not going to do***

1. Compare knowledge in one topic with another  
- though I shall compare ignorance
2. Give a recipe for deciding if something is true or probable - the 'science of evidence' [Phil Dawid's Leverhulme Project]
3. Discuss relativism - frameworks of belief - whether knowledge comes from science, religion, experiment, argument, etc.

*I assume we start with a defined body of accepted consistent truths (a '**TOPIC**'), and we are interested in whether we can identify and measure how much knowledge or ignorance someone has about these*

### ***How do we assess knowledge?***

#### **Approaches**

Ask for an exposition of K .....

Ask for inferences made from K ....

Ask Qs and count correct answers ...

#### **Problems**

Was it rote-learned?  
Would that matter?  
Is understanding a part of K?

Are they based on guesses?  
How hard are the Qs?

... all of the above

**Are there real standards? Do assessors agree?**

... "65% is a 2:1" ... 65% of what? ... "If you think it's a 2:1, put 65%"

... is assessment of knowledge just a matter of opinion?

**Is there any gold standard - a definition of Knowledge or Ignorance that gives the same answer independent of details of protocol?**

## Measuring knowledge or ignorance within a topic

1. Identify the topic - a set of true propositions
2.
  - choose a proposition P whose truth or falsity is implied by the topic
  - ask the subject to assign a probability B (degree of Belief) that P is true
  - $C=B$  if P is true;  $C=1-B$  if P is false. (C is probability for a Correct inference)
  - add  $\log(1/C)$  to a cumulative score of ignorance, I
  - tell the subject whether the proposition P was T or F (*feedback*)
3. Repeat this till you have covered propositions *equivalent* to the topic (i.e. whose truth or falsity entails and is entailed by those in the topic)

The result is **IGNORANCE of the topic**, independent of :-

- a) *the choice of questions,*
- b) *whether they were part of the initial definition of the topic,*
- c) *how similar or repetitive they may be, or their order,*
- d) *how they relate to the way in which the subject learned the topic*

... **PROVIDED** that the subject:

1. holds consistent probabilities, and reports these correctly
2. is able to form inferences from propositions within the topic
3. uses feedback rationally to inform future responses

Any deviation from 1,2,3 will *increase* the ignorance measure (on average)  
No other measure is independent of a,b,c,d

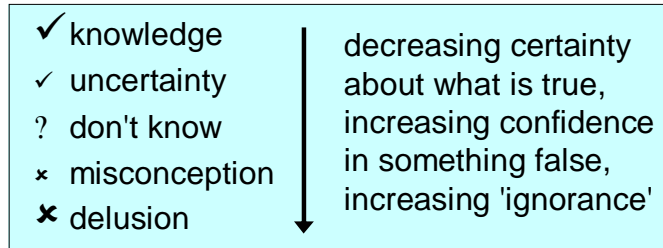
## Reservations?

- We can't realistically do this for substantial topics - it would take as long as learning the topics. We must rely on sampling.
- Real people will not be fully consistent or rational, so their ignorance score will be raised on that account, by an amount depending on the protocol

## Conclusions from theory?

- Ignorance has a unique measure, if the subject is consistent and rational
- Knowledge is inseparable from understanding. Without understanding, ignorance is ill-defined and potentially unbounded
- Parrots are ignorant, however well they can recite a topic
- Students likewise, if they cannot make inferences within the topic
- Ignorance for a statement is  $-\log(\text{probability assigned to its truth value})$
- Ignorance cannot be derived from just correct responses & errors
- Exposition of a topic is an unreliable way to test knowledge, because it may overlook (a) lack of understanding and (b) uncertainty
- Learning technology is well suited to obtaining the required information

### **Ordinary ways we think about Knowledge**



*Knowledge is a function of confidence (certainty, or degree of belief)  
The only anchor point is the top: total knowledge or zero ignorance  
There are states a lot worse than acknowledged ignorance*

"It ain't what you don't know that gets you into trouble.  
It's what you know for sure that just ain't so."

- attrib. to Mark Twain  
[Al Gore - *An Inconvenient Truth*]

### **'An Inconvenient Truth'**



***If you haven't seen it,  
book now at the Odeon, Panton St.!***

## Epistemology?

Knowledge is justified true belief

**Truth** - OK, we are assessing relative to given truths

**Belief** - We certainly have degrees of belief - a person may be certain, uncertain, or simply guessing - realistically, a continuum of probabilities

**Justification** - The trickiest and most important concept, critical to both assessment and learning

[Gettier? - discussion]

*NB modern pedagogic focus is on "Assessment for Learning" (e.g. Black, Sadler, Knight). Assessment should be part of the learning process, encouraging students' thinking about justification of answers, tying them in with other knowledge, implications and whether these enhance or go against belief in a conclusion. Assessment must encourage and reward such thinking. Reliable self-assessment is both constructive for learning and a desired outcome of education.*

**Your degree of belief is determined by your attempts to justify a proposition. Educational assessment should force you to make honest judgments about reliability -- Certainty-Based Marking (CBM)**

## The UCL (LAPT) Certainty-Based Marking scheme

Certainty Level:	C=1 (low)	C=2 (mid)	C=3 (high)	No Reply
Mark if Correct:	1	2	3	0
Penalty if wrong:	0	-2	-6	0
P(correct)	<67%	> P <	>80%	-

### London Agreed Protocol for Teaching (Of Physiology)



All about LAPT :

[www.ucl.ac.uk/lapt](http://www.ucl.ac.uk/lapt)

**Exercises, Publications, Tools, etc.**

**How should one choose the best CBM level?**

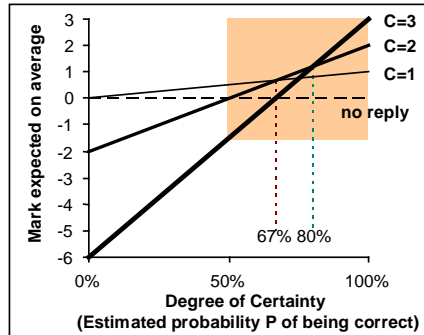
**Which line is highest?**

**C=3 when P(correct) >80%**

**C=1 when P(correct) <67%**

**The student is always motivated to be honest about low or high certainty, to get the best score.**

*[a proper scoring rule - P.Dawid]*



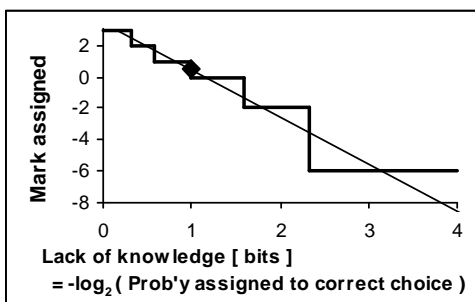
**NB the student gains:**

EITHER by finding justification for high confidence

OR by seeing reasons for reservation about an answer

If two students give the same answers, the one who distinguishes reliable from uncertain answers will do better.

**CBM quite closely follows the ideal ignorance measure**



**The student loses about 3 marks per 'bit' of ignorance**

**- up to a maximum of 3 bits**

## Example Qs

### Knowledge:

UK speed limits for a car (no trailer) on the following roads (not otherwise signposted) are? ....

A= 30 mph    B= 40 mph    C=50 mph    D= 60 mph    E= 70 mph

1.→ Dual carriage-way

1 ⌊ A ⌋ ⌊ B ⌋ ⌊ C ⌋ ⌊ D ⌋ ⌊ E ⌋

Confidence

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

2.→ Motorway

2 ⌊ A ⌋ ⌊ B ⌋ ⌊ C ⌋ ⌊ D ⌋ ⌊ E ⌋

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

3.→ Ordinary country road (2-way)

3 ⌊ A ⌋ ⌊ B ⌋ ⌊ C ⌋ ⌊ D ⌋ ⌊ E ⌋

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

4.→ Single lane country road with passing places

4 ⌊ A ⌋ ⌊ B ⌋ ⌊ C ⌋ ⌊ D ⌋ ⌊ E ⌋

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

### Numeracy:

5. Your antique clock rose by 80% in value in 10 years then fell by 40% in the next 10.

By what % did it rise in value overall, in the 20 years?

Confidence

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

### Logic:

6. Every branchiopod is a crustacean and every crustacean is an arthropod. No insect is a crustacean.

Given this, which of the following must be true? (Mark these T and the others F)

A.→Every branchiopod is an arthropod

A ⌊ T ⌋ ⌊ F ⌋

Confidence

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

B.→No insect is an arthropod

B ⌊ T ⌋ ⌊ F ⌋

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

C.→No branchiopod is an insect

C ⌊ T ⌋ ⌊ F ⌋

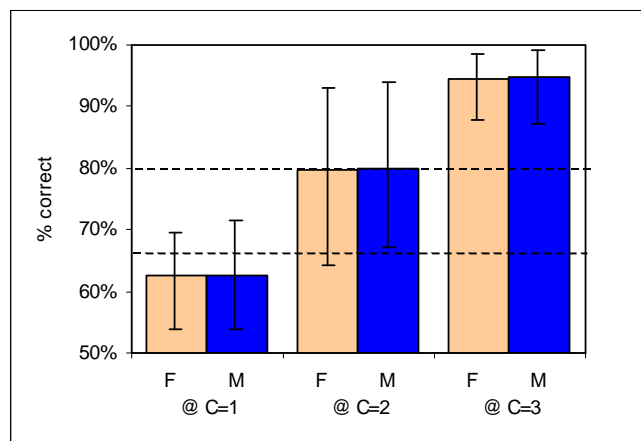
⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

D.→Some crustaceans are insects

D ⌊ T ⌋ ⌊ F ⌋

⌊ 1 ⌋ ⌊ 2 ⌋ ⌊ 3 ⌋

## How well do the students discriminate reliability?



Bars show range including 90% of students

[331 students, 500 T/F exam questions, 2001]

## Students adopt certainty-based marking very easily

### Principles that students seem readily to understand :-

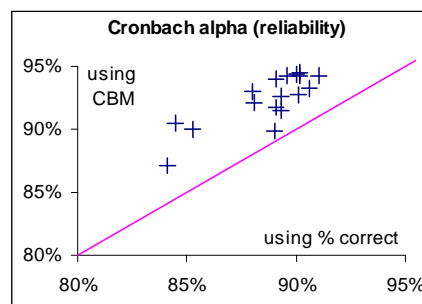
- You must know the reliability of each bit of your knowledge to use it
- confident errors are serious, requiring careful attention to explanations
- expressing uncertainty when you are uncertain is a **good thing**
- confidence is not just personality, it is being able to justify what you say
- if you are over- or under-confident, you should recalibrate with practice
- reflection integrates knowledge and understanding, and should be routine

In evaluation surveys, a majority of students have always said they like CBM, finding it useful and fair, and recently they voted 52%: 30% that it should be retained in Year 1&2 medical exams.

## ***CBM in Exams with True/False Questions***

17 medical exams, 250-300 questions in each, ca. 300-350 students

The standard measure of exam reliability, as a measure of student ability rather than chance factors, is called Cronbach Alpha.



Reliability increases if you use more exam questions.

To achieve these increases using only % correct would have required on average 58% more questions.



## Extended Matching Questions (EMQs)

- a new fashion in medical assessment

### CBM Trial (Feb '06) with EMQs

- Weak students over-estimated the reliability of their EMQ answers.
- Adjustment can be made to compensate, but ...
- this occurs much less with T/F Qs, and therefore CBM may be especially valuable with EMQs.
- *The Medical School has puzzlingly discontinued CBM trials with EMQs.*

Alcohol intoxication	Meningococcal septicaemia
Complete heart block	X Pulmonary embolus
Epilepsy	Supraventricular tachycardia
Gram negative septicaemia	Vaso vagal faint
<u>Left ventricular failure</u>	
Mechanical fall	

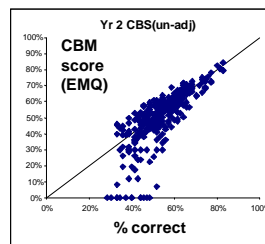
X Mark = -2 (conf=2) Proceed to the Next Question

For each of the scenarios below select the most likely diagnosis from the list above:

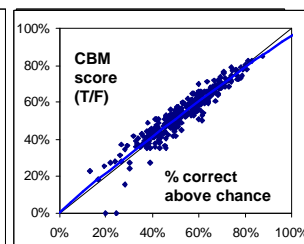
**Qu. 1:** An 80 year old man presents with collapse associated with severe shortness of breath. He has crackles throughout both lung fields.

**Model Answer:** Left ventricular failure

EMQ formative trial



T/F exam for comparison



## CBM - main current usage

Learning online or downloaded: self-assessment (> million/year)

- UCL, Imperial + students at >30 universities
- Follow-up exercises for lectures, practicals
- Revision, with past exams
- Student-written exercises (UCL & Imperial)

Formal online tests (using link to WebCT or VLE)

- Maths & key skills tests at UCL
- Formative module tests at Imperial
- Student teacher audit at Winchester College

Paper (OMR) tests

- Formative module tests at UCL & Imperial
- Yr 1,2 exams at UCL

Lecture/ seminar context

- Junior Doctor sessions at Imperial [Dr. Sara Marshall]  
"Are you prepared to act on your answer?"

### Lessons from experience with CBM

**Practice is needed before use in exams**

- not really a problem, since its objective is to encourage better thinking

**Exams should re-use questions from an open database only very sparingly**

- CBM places a premium on answers that the student has good reason to believe are correct - having seen the Q&A before is a good reason!

**Students can lose out through excessive self-confidence or risk-aversion**

- these are unhealthy traits that are moderated by practice with CBM
- adjustment can be made in exams to compensate for poor calibration\*\*

**With Best-Option Qs, students often think answers more reliable than they are**

- further research required
- students may benefit from confronting reality by using CBM

**Standard setting**

- the CBM mark range is unfamiliar, but scaling aligns it with the familiar\*\*

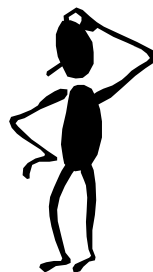
**\*\* see Publications**

**Positive Features :**

- students like CBM, consider it helps them study and is more fair
- they have voted at UCL to retain it in Yr 1,2 exams
- it is more reliable and valid than conventional marking in exams
- it is more closely related to what we mean by 'knowledge'

***We fail if we mark a lucky guess as if it were knowledge.***

***We fail if we mark delusion as no worse than ignorance.***



**[www.ucl.ac.uk/lapt](http://www.ucl.ac.uk/lapt)**