FOR INTERNAL SCRUTINY (date of this version: 13/3/2009)

UNIVERSITY OF EDINBURGH

COLLEGE OF SCIENCE AND ENGINEERING

SCHOOL OF INFORMATICS

COGNITIVE MODELLING (LEVEL 10)

COGNITIVE MODELLING (LEVEL 11)

Friday 1 April 2005

00:00 to 00:00

Year 4 Courses

Convener: G Plotkin
External Examiners: J Gurd, M Wooldridge

MSc Courses

Convener: A Smaill
External Examiners: J Carroll, M Hepple, E Hull, I Marshall

INSTRUCTIONS TO CANDIDATES

Answer QUESTION 1 and ONE other question.

Question 1 is COMPULSORY.

All questions carry equal weight.

1. **You MUST answer this question.**

   (a) This question asks you to consider the relative merits of the rational analysis and cognitive architecture approaches for modelling human memory.

   i. Describe three different psychological phenomena (behavioural effects) that relate to human memory. [*3 marks*]

   ii. Discuss the advantages and disadvantages of each modelling approach for modelling human memory, using the effects you mentioned as examples. [*5 marks*]

   (b) In a criminal court case, an expert witness provides testimony regarding a test for detecting certain markers in a person's blood. The test was performed on blood collected at the crime scene and on the defendant's blood. The expert states the following: 1) the test showed a match between the two blood samples, 2) the test always shows a match when there is one, 3) the test correctly shows no match in 99% of cases where there is no match, and 4) the probability of a true match between two randomly selected people's blood is only .0001.

   i. Assuming the expert's information is correct and no other information is known, write down how you would determine the probability that the markers in the blood at the crime scene actually match those of the defendant, and give the approximate probability. [*3 marks*]

   ii. Why might the expert's testimony be problematic when presented this way? [*2 marks*]

   (c) A coin is flipped five times, with the following results: H T H H T. Let $P(h)$ be the probability that the coin comes up heads on the next flip.

   i. What is the maximum-likelihood estimate of $P(h)$ based on the observed data? [*2 marks*]

   ii. If Bayesian inference were used, would the estimate of $P(h)$ be higher or lower than the maximum-likelihood estimate? Mention any assumptions you make in answering this question. [*2 marks*]

   iii. Explain why Bayesian inference may be a better basis for probabilistic cognitive models than maximum-likelihood estimation. [*4 marks*]

   (d) Suppose that you are implementing a production rule model, and that multiple production rules are applicable on the same cycle. Which rule(s) will fire on that cycle

   i. in Cogent (assuming the default behaviour)? [*2 marks*]

   ii. in ACT-R? [*2 marks*]

2. **You should either answer this question or question 3.**

   Developmental psychologists have devised several methods for testing the capacity of pre-verbal infants to evaluate numerical quantities. In a typical experiment, two cups are placed next to each other in front of the infant subject. First, a small number of identical treats (e.g., biscuits) is placed, one by one, into the left-hand cup. Once they are in the cup, the infant is unable to see the treats. Next, some treats are placed one by one into the right-hand cup (again disappearing from view). Finally, the infant is allowed to reach for one of the cups. Infants reliably choose the cup with more treats when the comparison is between small values such as 1 vs. 2 or 2 vs. 3, but not when the comparison is between larger values such as 4 vs. 6.

   (a) Based on the information above, propose a possible explanation for the difference in behaviour between the comparisons mentioned. [*4 marks*]

   (b) Based on your answer to part (a), design a Cogent model for the experiment described above. Give the box-and-arrow diagram of the model, and explain the function of each of the buffers and processes you assume. (You do not have to give the production rules for the processes.) Include in your description any predicates used to represent the information being processed or stored by the model. [*10 marks*]

   (c) Explain how your model accounts for the behaviour described. Which properties of rules or buffers are important for the model to produce this behaviour? What additional testable predictions does your model make? [*8 marks*]

   (d) As well as failing to discriminate pairs such as 4 vs. 6, infants also fail to discriminate 2 vs. 4 and 1 vs. 4 (in fact, they choose a cup at random in these cases). Do either of these cases present a problem for your model? Explain. [*3 marks*]

3. **You should either answer this question or question 2.**

Tenenbaum (2000) performed the following experiment to test people's generalisation behaviour in a mathematical domain. Subjects interacted with a computer program and were told that the program would choose a subset of the numbers between 1 and 100 based on simple rules such as "all even numbers", "all multiples of 10", "all numbers between 43 and 55", etc. Their task was to guess which numbers belonged to the set. First, the computer would present some examples of numbers chosen randomly from the set (e.g., $\{2, 32\}$ or $\{45, 47, 50, 51\}$). Then the subject had to predict, for various other numbers, how likely each of these numbers was to belong to the set as well. Results of the experiment will be discussed below.

Tenenbaum modelled subjects' behaviour using a Bayesian model. Let $X = \{x_1 \dots x_n\}$ be the set of examples subjects were presented with, and R be the rule chosen by the computer to define the full set of numbers (e.g., "all even numbers"). Subjects are asked to predict $P(y \in R|X)$, for some new example number y. The hypothesis space H consists of all possible definition rules. For this question we will assume that all numbers lie between 1 and 20 rather than 1 and 100.

(a) Write down an expression for computing the likelihood under this model. *[2 marks]*

(b) Write down an expression for computing $P(y \in R|X)$ by summing over hypotheses. *[2 marks]*

(c) Suppose that the only hypotheses under consideration are

- h_1: powers of two
- h_2: even numbers
- h_3: $\{8\}$
- h_4: $\{7,8,9\}$

and that each has equal prior probability. Also suppose $X = \{8\}$.

i. Which is the maximum-likelihood hypothesis? *[1 mark]*

ii. If R is taken to be the maximum-likelihood hypothesis, what is $P(y \in R|X)$ for $y = 4$, $y = 6$, $y = 8$, and $y = 9$? What if R is taken to be the maximum a posteriori (MAP) hypothesis? *[2 marks]*

iii. Tenenbaum found that when only one example was provided, subjects gave most test numbers similar probabilities of acceptance, with slightly higher probabilities for numbers that were intuitively more similar to the example. For example, when 16 was the example number, 8 was rated higher than 6 or 9, and 17 was rated higher than 87. Discuss these results in relation to the maximum-likelihood and MAP predictions you gave above. *[2 marks]*

iv. What is the posterior probability of each of the hypotheses? You do not need to reduce fractions, i.e. answers of the form $\frac{1/2}{4/5+1/6}$ are acceptable. *[4 marks]*

v. Consider $P(y \in R|X)$ for $y = 4$, $y = 6$, and $y = 9$ using Bayesian inference (i.e. using the expression from part b.). Which value of y has the highest predicted probability of being in R, and why? How might this explain the behavioural results in part iii? *[8 marks]*

(d) Tenenbaum also found that when several examples were provided, subjects tended to accept only numbers that followed the most specific rule consistent with the examples. So, for examples $\{4, 32, 2, 64\}$, subjects would accept 8 and 16 with probability near one, and all other numbers with probability near zero. According to the model, why do people's predictions become more rule-like as the number of examples goes up? *[4 marks]*