FALL 1996
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
Ph.D. QUALIFYING EXAMINATION

Artificial Intelligence

MONDAY, SEPTEMBER 16
3:00 – 7:00 PM
ROOM 115 PSYCHOLOGY

GENERAL INSTRUCTIONS:

1. Answer each question in a separate book.
2. Indicate on the cover of *each* book the area of the exam, your code number, and the question answered in that book. On *one* of your books, list the numbers of *all* the questions answered. *Do not write your name on any answer book.*
3. Return all answer books in the folder provided. Additional answer books are available if needed.

SPECIFIC INSTRUCTIONS:

Answer:

- either one (1) of the questions G540-1 or G540-2, *and*
- both (2) questions in the section labeled B760, B761, or B766 corresponding to your chosen focus area, *and*
- any two (2) additional question in the sections B760, B761, and B766, where these two questions need *not* come from the same section, *and*
- both (2) questions in the section labeled A76x that corresponds to your focus area.

Hence, you are to answer a total of *seven* (7) questions.

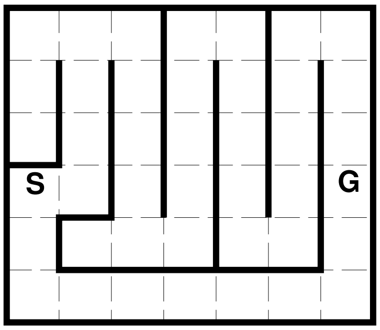
POLICY ON MISPRINTS AND AMBIGUITIES:

The Exam Committee tries to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced that a problem has been stated incorrectly, mention this to the proctor. If necessary, the proctor can contact a representative of the area to resolve problems during the *first hour* of the exam. In any case, you should indicate your interpretation of the problem in your written answer. Your interpretation should be such that the problem is nontrivial.

Answer either one (1) of the questions G540-1 or G540-2.

G540 – Introduction to AI Questions

G540-1. Consider the problem of trying to find a path in the following maze from the start location, S, at (1,3) to the goal location, G, at (7,3), where the lower-left grid square is at coordinates (1,1) and the upper-right grid square is at (7,6). The possible operators are **goN**, **goE**, **goS**, and **goW**, where each is valid only if there is a dashed line in that compass direction. Consider the heuristic to be the straight line (i.e., Euclidean) distance between the current location and the goal location.



(a) Show what happens, and why, when *hill-climbing* search is used.

(b) Name a search strategy that guarantees finding an optimal solution.

(c) Give the sequence of the first 7 squares expanded when using the A^* search strategy. Assume each move has unit cost.

(d) In general, if *h1* and *h2* are two *admissible heuristics*, which composite heuristic function, *min(h1, h2)* or *max(h1, h2)*, would be preferred? Why?

G540-2. Consider the following:

Hypotheses:

John is younger than Mary. Mary is younger than one of Fred's children.

Conclusion:

John is younger than Fred.

(a) Convert the hypotheses and conclusion to predicate logic, using predicates *younger(x,y)* and *child(x,y)*.

(b) Add to the hypotheses some REASONABLE real-world assumptions that make the conclusion really follow logically from the hypotheses.

(c) Then, convert the hypotheses and the negation of the conclusion to clauses and derive the empty clause by resolution.

Answer both (2) of the questions in the section labeled B76x that corresponds to your chosen focus area. Also answer any two (2) additional questions in any of the other sections (these two questions need NOT occur in the same section).

B760 – Machine Learning Basic Questions

B760-1. Consider the claim that all *inductive biases* in machine learning are a consequence of Occam's Razor. First define the phrases "inductive bias" and "Occam's Razor." Then describe three (3) inductive biases; for each, also name an existing algorithm that uses it and discuss the bias' relationship to Occam's Razor. Finally, debate the validity of the above claim.

B760-2. Imagine you plan to use *reinforcement learning* to construct a controller of a physical system. Consider the following two approaches:

1. Watkins' Q-learning algorithm is directly applied to the physical system.
   That is, no world model is learned, and during training the chosen actions are executed on the actual physical system.
2. A world model is first learned by interacting with the physical system,
   and then Q-learning is performed using this model (i.e., "mental simulation").

First, briefly describe and motivate the (one-step) Q-learning algorithm. Second, explain the strengths and weaknesses of each of the two approaches above. (The phrase "world model" refers to both the next-state and reward functions.)

B761 – Deductive Problem Solving & Logic Programming Basic Questions

B761-1. Derive the empty clause from the following, using binary resolution plus paramodulation.

1. $x = x$
2. $x \cdot (y \cdot z) = (x \cdot y) \cdot z$
3. $x \cdot x = 1$
4. $1 \cdot x = x$
5. $x \cdot 1 = x$
6. $x \cdot y \neq x \cdot z \ \vee \ y = z$
7. $a \cdot b \neq b \cdot a$

Here, x, y, z are variables and $a, b, 1$ are constants. A "human" proof would be: $ab\,ab = 1 = abba$ (applying (3) three times), so, cancelling the ab (by (6)), we get $ab = ba$.

B761-2. Consider the **Prolog** program:

```
member(X, [X | T]).
member(X, [Y | T]) :- member(X,T).
wow :- member(X,X).
```

The query `?- member(f(a,Y), [c, f(X,b), d])` will result in `X` $=$ `a`, `Y` $=$ `b`, and the query `?- wow` will result in the answer `yes` in most Prologs (including Quintus). Explain how Prolog derives these answers. Your explanation should address how Prolog searches for a solution, the unifications involved, and how Prolog represents lists as terms.

B766 – Computer Vision Basic Questions

B766-1.

(a) Define (i) the *correspondence problem*, (ii) *disparity*, and (iii) the *epipolar plane* as used in the problem of recovering depth by *binocular stereo*.

(b) Specify an intensity-based (not edge-based) method for solving the correspondence problem.

(c) If a single continuous surface is completely visible in the two images, what constraint can be used to help solve the correspondence problem?

B766-2.

(a) Describe the *tangent versus arc-length representation* of a curve. Include as part of your answer an example of this representation for a curve containing line segments and arcs of circles of various sizes.

(b) Describe a position- and orientation-invariant method for determining if two curves have similar shape using this representation.

Answer both (2) of the questions in the section labeled A76x that corresponds to your chosen focus area.

A760 – Machine Learning Advanced Questions

A760-1.

(a) Briefly describe Towell and Shavlik's KBANN algorithm that performs *theory refinement*.

(b) Imagine someone has written you the following letter. What would be your response?

> The final network KBANN produces is dependent on the training examples. Clearly, if you have a "good" training set, the imperfections in the domain theory will be corrected, and an accurate trained network will result. But why would not a standard neural network (i.e., one not given the domain theory), when given these same good training examples, not do as well? If a neural network is given good training data it will accurately approximate the function, otherwise it will not. It is understandable that giving a neural network some hints will result in faster training, but why should one expect that generalization would be better?

(c) Do any of your arguments in Part (b) change if instead Pazzani and Kibler's FOCL algorithm was being compared to Quinlan's FOIL? Explain.

A760-2. One problem with using *genetic algorithms* in a learning-from-examples setting is that they can overfit their fitness functions.

(a) Define *overfitting* and explain how a fitness function might be overfit.

(b) One solution for the above problem is to have two learners. One of them learns the fitness function, while the other optimizes the current fitness function. First, explain how this solution addresses the overfitting problem, then provide a concrete design for such a dual-learner system for the specific case of learning from a large database of labeled examples.