6.034 Quiz 2
25 October 2013

Name	Daenerys Targaryen
email	KHALEESI OF THE GREAT GRASS SEA, BREAKER OF CHAINS, MOTHER OF DRAGONS, MHYSA, THE QUEEN ACROSS THE SEA, STORMBORN

Circle your TA **(for 1 extra credit point)**, so that we can more easily enter your score in our records and return your quiz to you promptly.

Michael Fleder	**Giuliano Giacaglia**	**Dylan Holmes**
Casey McNamara	**Robert McIntyre**	**Duncan Townsend**
Mark Seifter	**Sam Sinai**	**Prashan Wanigasekara**

Problem number	Maximum	Score	Grader
1	50		
2	50		
Total	100		

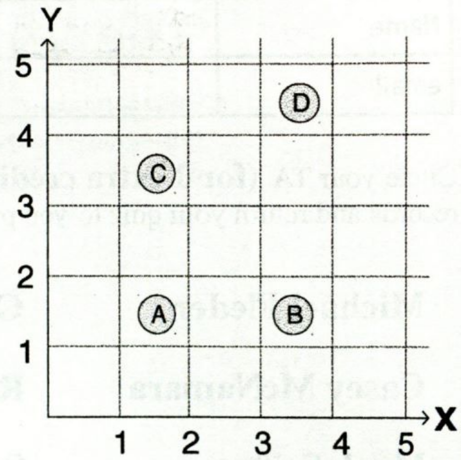
3 (SRN)	5		

There are **12** pages in this quiz, including this one, but not including blank pages and tear-off sheets. Tear-off sheets with duplicate drawings and data are located after the final page of the quiz. As always, open book, open notes, open just about everything, including a calculator, but no computers.

Problem 1: ID Trees (50 points)

Part A: Data points from different classes (12 points)

While studying for a 6.034 exam, two of your friends ask you to resolve a disagreement about ID trees. They show you a problem from lecture in which you're using an ID tree to classify four points **of different classes**.



Xavier, who took the class last year, heard Prof. Winston say that the **X > 3** test has the lowest disorder, since it divides the training set into two subgroups with less disorder.

Yvonne, who's taking the class this year, heard Prof. Winston say that the **Y > 4** test has the lowest disorder, since it can completely separate one of the classes (D) from the rest.

A1.(8 points) Compute the average disorder of each test. Express your answer in terms of sums, products, integers, ratios of integers, and logarithms.

Test	Disorder
X > 3	$\frac{1}{2}\left(-\frac{1}{2}\log\frac{1}{2} + -\frac{1}{2}\log\frac{1}{2}\right) + \frac{1}{2}\left(-\frac{1}{2}\log\frac{1}{2} + -\frac{1}{2}\log\frac{1}{2}\right) = \log 2 = 1$
Y > 4	$\frac{1}{4}\left(1\log 1\right) + \frac{3}{4}\left(-\frac{1}{3}\log\frac{1}{3} + -\frac{1}{3}\log\frac{1}{3} + -\frac{1}{3}\log\frac{1}{3}\right) = \frac{3}{4}\log 3$ ($1\log 1 \to 0$)

A2.(4 points) Which test has lower average disorder? (Circle one)

[Xavier's test: X > 3] Yvonne's test: Y > 4 Neither one; their disorders are equal

Here's a table of logarithms if you need one. This table is repeated on a tear-off sheet.

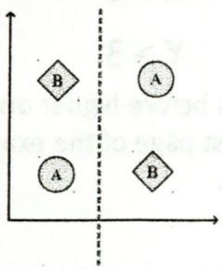
n	1	2	3	4	5	6	7
$\log_2(n)$	0	1	≈1.584	2	≈2.321	≈2.584	≈2.807

$$\log(a \cdot b) = \log a + \log b \qquad \log\frac{a}{b} = \log a - \log b$$

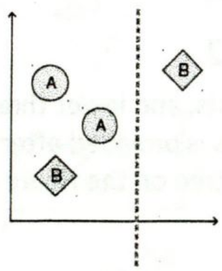
Part B: Ordering chaos (12 points)

Your Solomonic wisdom in Part A impresses your friends, who are eager to learn more. They show you a number of different training sets with possible tests and ask you to tell them which is best.

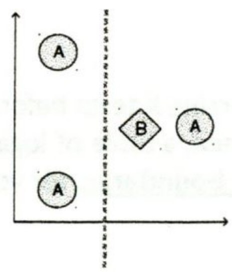
Sort the following tests in order of increasing disorder. In case of a tie, put the alphabetically earlier scenario first. (You may solve by inspection; you don't need to show your calculations.)



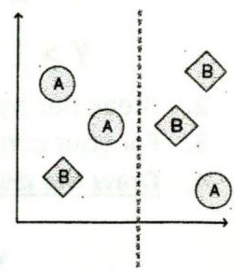
Scenario A



Scenario B



Scenario C



Scenario D

The correct sorted order is:

Scenario C (least disorder)   Scenario B   Scenario D   Scenario A (highest disorder)

You may use this space to show your work:

Part C: Dichotomous trees (26 points)

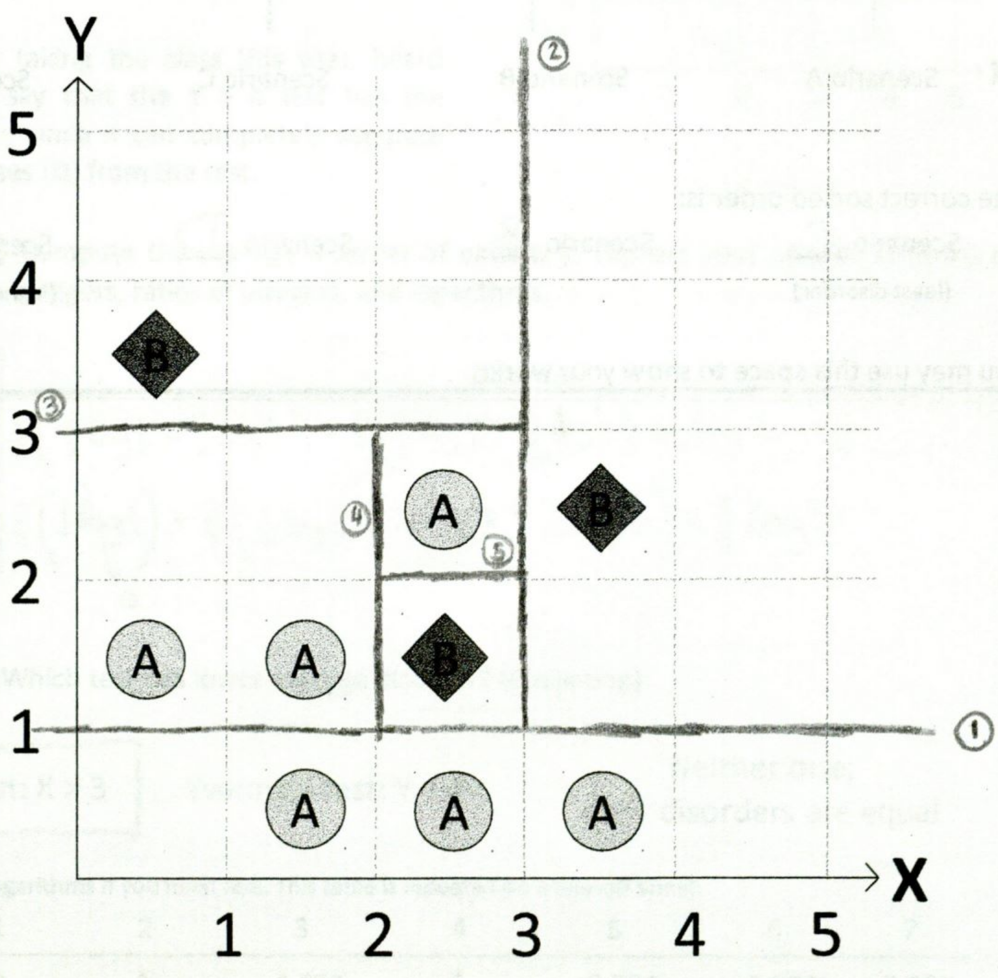
Your friends, now firmly convinced of your limitless ability to classify data, ask you to construct an ID tree for a large dataset.

C1.(13 points) Construct the complete greedy disorder-minimizing ID tree for dataset depicted below.

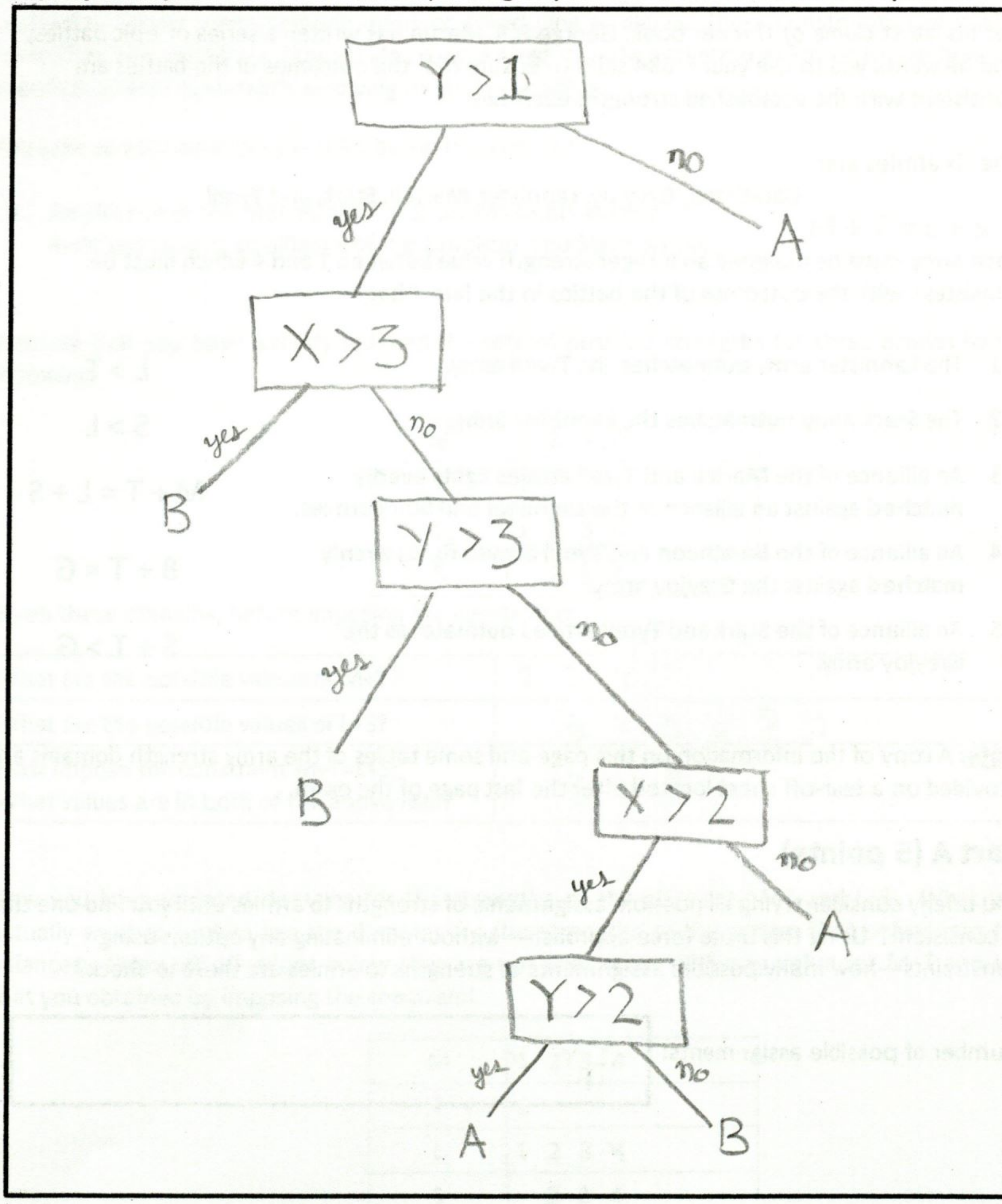
1. Choose tests from the following list:

X > 1	X > 2	X > 3
Y > 1	Y > 2	Y > 3

2. Break ties by preferring X tests before Y tests, and lower thresholds before higher ones.
3. For your convenience, a table of logarithms is provided after the last page of the exam.
4. **Draw the decision boundaries** for your ID tree on the figure below.



C2.(13 points) Draw the ID tree corresponding to your decision boundaries from part C1.



Problem 2: Authorial execution (50 points)

For his latest *Game of Thrones* book, George R.R. Martin has written a series of epic battles, and he wants you to use your 6.034 skills to ensure that the outcomes of the battles are consistent with the established strengths each side.

The six armies are:

Baratheon, **G**reyjoy, **L**annister, **M**artell, **S**tark, and **T**yrell

Each army must be assigned an integer strength value between 1 and 4 which must be consistent with the outcomes of the battles in this latest book:

1. The **L**annister army **outmatches** the **T**yrell army. $L > T$
2. The **S**tark army **outmatches** the **L**annister army. $S > L$
3. An alliance of the **M**artell and **T**yrell armies fights **evenly matched** against an alliance of the **L**annister and **S**tark armies. $M + T = L + S$
4. An alliance of the **B**aratheon and **T**yrell armies fights **evenly matched** against the **G**reyjoy army. $B + T = G$
5. An alliance of the **S**tark and **T**yrell armies **outmatches** the **G**reyjoy army. $S + T > G$

Note: A copy of the information on this page and some tables of the army strength domains are provided on a tear-off sheet located after the last page of the exam.

Part A (5 points)

You briefly consider trying all possible assignments of strengths to armies until you find one that is consistent. Using this brute force approach—without eliminating any options using constraints—how many possible assignments of strengths to armies are there to check?

Number of possible assignments: 4^6

$4 \cdot 4 \cdot 4 \cdot 4 \cdot 4 \cdot 4$

B G L M S T

Part B – Constraints containing many variables (10 points)

The constraints imposed by alliances winning battles are a bit different from the constraints you remember seeing when coloring maps or scheduling airplanes: these constraints can involve more than two variables. You decide to work through a hypothetical example to see how you should deal with constraints involving multiple variables.

You take as your example the third battle from the list:

3. An alliance of the **M**artell and **T**yrell armies fights **evenly matched** against an alliance of the **L**annister and **S**tark armies.

$$M + T = L + S$$

Suppose that you have already reduced the sets of possible strengths for these armies to the following:

M	1 2 3 4
T	1
L	1 2 3 4
S	2 3 4

Given these domains, before imposing the constraint:

What are the possible values of M+T?	2 3 4 5
What are the possible values of L+S?	3 4 5 6 7 8
Now impose the constraint M+T=L+S: What values are in both of the above lists?	3 4 5

Now you have updated domains for the strengths of the alliances M+T and L+S. What you actually want to update are the domains for the strengths of the armies themselves, not the alliances. **Cross off** all values below that are **not consistent with any value** for M+T and L+S that you obtained by imposing the constraint.

M	~~1~~ 2 3 4
T	1
L	1 2 3 ~~4~~
S	2 3 4

These would be your updated domains for the strengths of the armies after enforcing the constraint that M+T=L+S.

Part C – Last one standing (35 points)

C1.(25 points) Unfortunately, your employer is getting impatient, so you decide to forego brute force search. Instead, use **depth-first search with forward checking and propagation through domains reduced to size 1** to search for a solution.

1. Consider the armies in alphabetical order.
2. Try assigning lower strength values before higher ones.
3. **To receive credit, draw your search tree on the next page.**
4. Recall that your starting domains are as follows: (Duplicate copies of the constraints and domains are provided on a tear-off sheet after the last page of the exam)

B	1 2 3 4
G	1 2 3 4
L	1 2 3 4
M	1 2 3 4
S	1 2 3 4
T	1 2 3 4

L > T

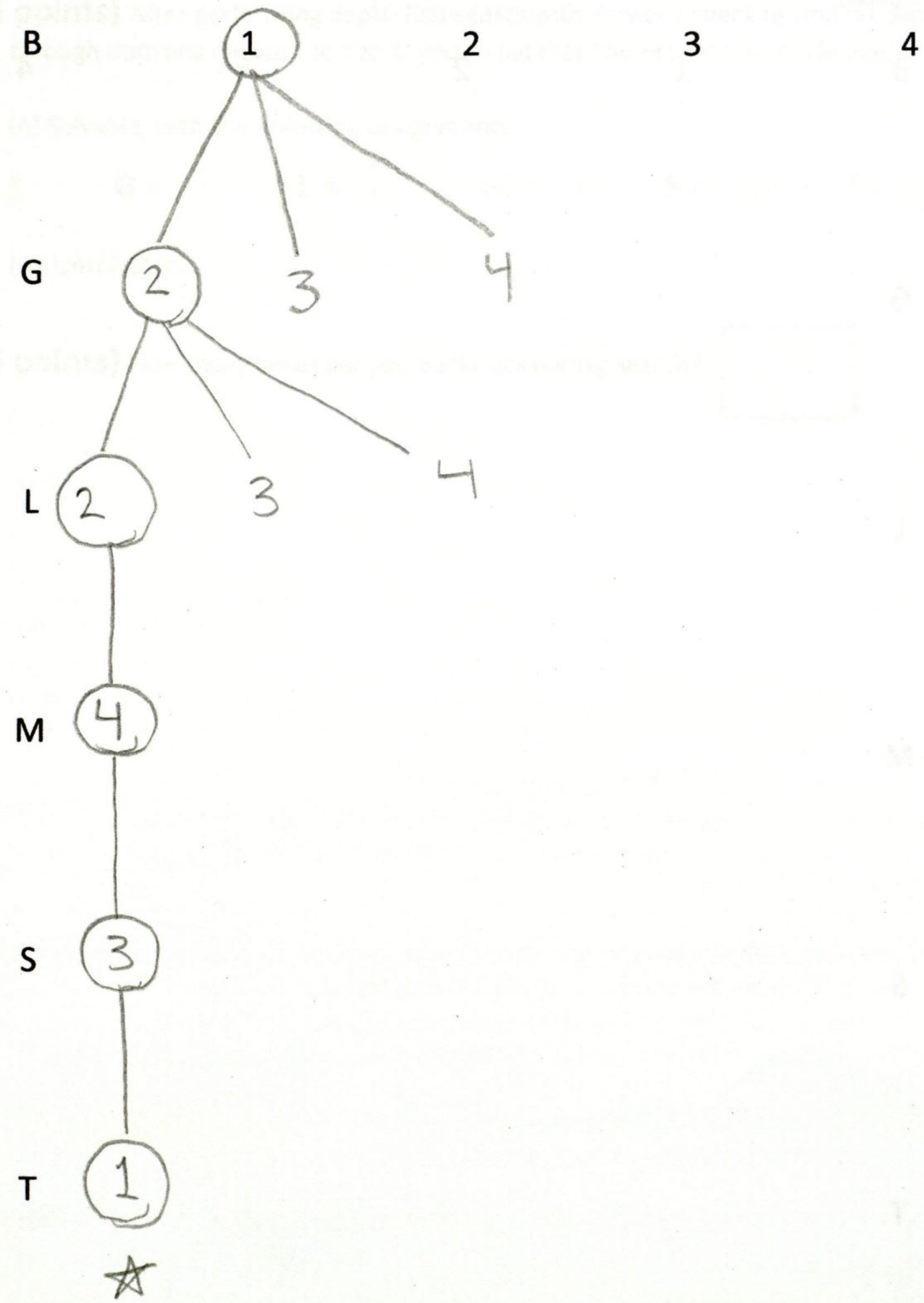
S > L

M + T = L + S

B + T = G

S + T > G

(Draw your tree on this page. For your convenience, an extra copy of this page is provided on the next page.)



CONSTRAINT PROPAGATION

① Assign B = 1

$$\overset{1}{B} + \overset{1,2,3,\cancel{4}}{T} = \overset{\cancel{1},2,3,4}{G}$$

② Assign G = 2

$$\overset{1}{B} + \overset{1,\cancel{2},\cancel{3}}{T} = \overset{2}{G}$$

$$\overset{\cancel{1}234}{S} + \overset{1}{T} > \overset{2}{G}$$

③ Propagate T

$$\overset{\cancel{1}234}{L} > \overset{1}{T}$$

$$\overset{\cancel{1}\cancel{2}34}{M} + \overset{1}{T} = \overset{23\cancel{4}}{L} + \overset{23\cancel{4}}{S}$$

$$\overset{1}{B} + \overset{1}{T} = \overset{2}{G} \;\checkmark$$

$$\overset{234}{S} + \overset{1}{T} > \overset{2}{G} \;\checkmark$$

④ Assign L = 2

$$\overset{2}{L} > \overset{1}{T} \;\checkmark$$

$$\overset{\cancel{2}3}{S} > \overset{2}{L}$$

$$\overset{\cancel{3}4}{M} + \overset{1}{T} = \overset{2}{L} + \overset{3}{S}$$

⑤ Propagate S

$$\overset{3}{S} > \overset{2}{L} \;\checkmark$$

$$\overset{4}{M} + \overset{1}{T} = \overset{2}{L} + \overset{3}{S} \;\checkmark$$

$$\overset{3}{S} + \overset{1}{T} > \overset{2}{G} \;\checkmark$$

⑥ Propagate M

$$\overset{4}{M} + \overset{1}{T} = \overset{2}{L} + \overset{3}{S} \;\checkmark$$

⑦ Assign M = 4

$$\overset{4}{M} + \overset{1}{T} = \overset{2}{L} + \overset{3}{S} \;\checkmark$$

⑧ Assign S = 3

$$\overset{3}{S} > \overset{2}{L} \;\checkmark$$

$$\overset{4}{M} + \overset{1}{T} = \overset{2}{L} + \overset{3}{S} \;\checkmark$$

$$\overset{3}{S} + \overset{1}{T} > \overset{2}{G} \;\checkmark$$

⑨ Assign T = 1

$$\overset{2}{L} > \overset{1}{T} \;\checkmark$$

$$\overset{4}{M} + \overset{1}{T} = \overset{2}{L} + \overset{3}{S} \;\checkmark$$

$$\overset{1}{B} + \overset{1}{T} = \overset{2}{G} \;\checkmark$$

$$\overset{3}{S} + \overset{1}{T} > \overset{2}{G} \;\checkmark$$

☆

C2.(5 points) After performing depth first search with forward checking and propagation through domains reduced to size 1, you found that this problem is (circle one):

(A) Solvable, with the following assignments

B = 1 **G =** 2 **L =** 2 **M =** 4 **S =** 3 **T=** 1

(B) Unsolvable.

C3.(5 points) How many times did you backtrack during search? 0

Problem 3: Spiritual and Right Now

Circle the **one best** answer for each of the following question. There is **no penalty for wrong answers**, so it pays to guess in the absence of knowledge.

1. Sussman's explanation of propagators featured an example involving:

2
1. Spread of epidemics.
2. Electronic circuit analysis.
3. Calculating the distance to a galaxy.
4. Calculating a route from Boston to Providence.
5. Evolution of birds following ice ages.

2. Wilson indicated that damage to the human hippocampus leads to inability to:

2
1. Remember faces of family and friends.
2. Form new memories.
3. Recall childhood experiences.
4. Do crossword puzzles.
5. Pay attention in lectures.

3. Wilson indicated work on rats demonstrates that rat memory:

4
1. Declines with age.
2. Declines with lack of sleep and poor nutrition.
3. Includes if-then rules used while making choices while running mazes.
4. Includes ability to play place sequences backwards after reaching a goal.
5. Improves with practice.

4. Ullman said that fMRI response in the human brain to high-information recognition features:

3
1. Is greater than for low-information features throughout the brain.
2. Is the same as for low-information features throughout the brain.
3. Is greater than for low-information features in brain parts specialized to object recognition.
4. Is unusually high for hands and faces.
5. Increases rapidly throughout the first year of life.

5. Ullman theorizes that infants learn what hands look like because:

2
1. They move.
2. They move other things.
3. Muscle sensors tell the infant where his hands are.
4. Infants spend a lot of time sucking on their thumbs.
5. Infants have innate abilities that include hand and face recognition.