RPM Final Project

Allen Worthley
mworthley3@gatech.edu

***Abstract*—** This paper discusses an agent that utilizes various heuristics and comparison methodologies to solve Raven's Progressive matrices problems. Ultimately, the agent correctly solves 61 of the 96 problems.

1 DESIGN

The design of this agent leans on two core methodologies: using logical and affine transformations with Tversky tranformations, and using a dark pixel ratio ("DPR") and intersection pixel ratio ("IPR") similar to Joyner's et all (2015) agent three. This agent solves Raven's Prgressive matricies problems of sizes 2 x 2, and 3 x 3. Methodogies differ with the size of the presented matrix dimentions of the given problems. Each methodology uses a voting system to track how each employed heuristic ranks for each answer image. The proceding sections ellaborate on the nuances of applying these methodologies to each size.

1.1 Design of 2x2

The agent attempts to infer transformations in the form of several basic operations across three heuristics: at a row, column, and diagonal level. The operations follow Kunda's, Megreggor's, and Goel's (2013) design which included: Identity, rotation of 90 degrees, rotation of 180 degrees, rotation of 270 degrees, reflection from left to right, and reflection from top to bottom transformations. (Kunda, & Goel, 2013) To deduce from the problem set which operations is applicable, the agent produces an image and compares the answer set to that image. The image with a Tversky similarity that meets a given criteria receives a vote. (Tversky 1977) This type of comparison yields the best results as no pixel-wise comparison after transformations are almost never exactly matching pixel for pixel.

In addition, as Joyner et all (2015) specifies in agent three, the agent in this paper uses a DPR and IPR methodology to solve the matrices problems. Joyner et all describes DPR as:

> "the difference in percentage of the number of dark-colored pixels with respect to the number of total number of pixels in the continuous pixel set of two matrixes." (Joyner et al., 2015)

Further, the author describes IPR as:

> "the difference in percentage of the number of dark-colored pixels present at the same coordinates with respect to the total number of dark-colored pixels in both cells for a given set of contiguous pixels." (Joyner et al., 2015)

These ratios act as a benchmark for comparing row, column, and diagonal relationships within the given problem matrices. The agent compares the following relationships: A to B, A to C, and B to C. The ratios calculated between these problem images represent the "base" DPR and IPR. These base ratios are compared to the relationships calculated for each answer set. The comparison starts from either cells C (row), B (column), or A (diagonal) to the answer set image. If the relationship from given comparison matches the observed relationships and is within tolerance, then it receives a vote.

1.1.1 Parameters and Thresholds in 2x2

Several parameters act as hard thresholds for the Tversky similarity comparisons and ratios (DPR and IPR). The value 0.01 stages the cutoff value for the Tversky comparisons. The Tversky similarity metric is at most 1 when there is perfect similarity or, when every pixel matches between two images. The cutoff statistic is one minus the calculated Tversky metric of the transformed image (based on the operation employed) and a given answer image. If the cutoff statistic is less than 0.01, then the agent votes for that given answer.

Similarly, a threshold of 0.05 acts a cutoff value for the DPR and IPR heuristic tests. If a DPR or IPR calculated statistic varies from the observed DPR IPR relationships in the problem image sets less than 0.05, then that answer image receives a vote.

1.1.2 Voting Mechanism for 2x2

This agent manages the various heuristic results in a voting mechanism to find the best answer for the problem. If a given heuristic meets the threshold criteria for Tversky or DPR/IPR, then that answer receives a vote. For 2x2 matrices, the vote is scaled by the confidence in the heuristic. For instance, if a given heuristic metric is within the 0.05 DPR / IPR threshold at 0.01, then the vote value is calculated as 1 divided by 0.01 or 100. This model of scaling favors answers that are closest to zero. Or, in other words, are the most similar to the observed relationships within the given problem matrices.

1.2 Design of 3x3

The design of the 3x3 agent is very similar to the 2x2 design. However, there are no affine transformations and the DPR/IPR methodology includes some additional logic. Also, logical comparisons of images were added.

Logical functions check for relationships of logical AND, OR, or XOR of two images. These logical relationships are compared at the row, column, left diagonal, and right diagonal level. Then the resulting image is compared to all answer images. The answer image with the highest Tversky similarity receives a vote.

For the 3x3 matrix DPR/IPR methodology, further abstraction is necessary. The agent uses the 2x2 approach as previously described for the basis of the analysis. Except it uses composite cells which are the logical AND of two images. For instance, the agent incorporates the cells D & E, and B & E to calculate the DPR and IPR for (D&E) to F, and (B&E) to H as shown in Figure 2. These composite images incorporate more information than if only E were used for the calculations.

Additionally, the agent employs an aggregate change in DPR and IPR for row, column, left diagonal, and right diagonal levels. This encapsulates more information from the larger matrix. Some matrices may have information embedded in the series of rows or columns. Therefore, the agent attempts to capture this information by row and column. For example, in Figure 2, the difference in the Change in Row % and the Observed Change in Row % from a given answer image should be relatively similar. In practice, the agent operates on this principle with a tolerance of 0.15 in either DPR or IPR.

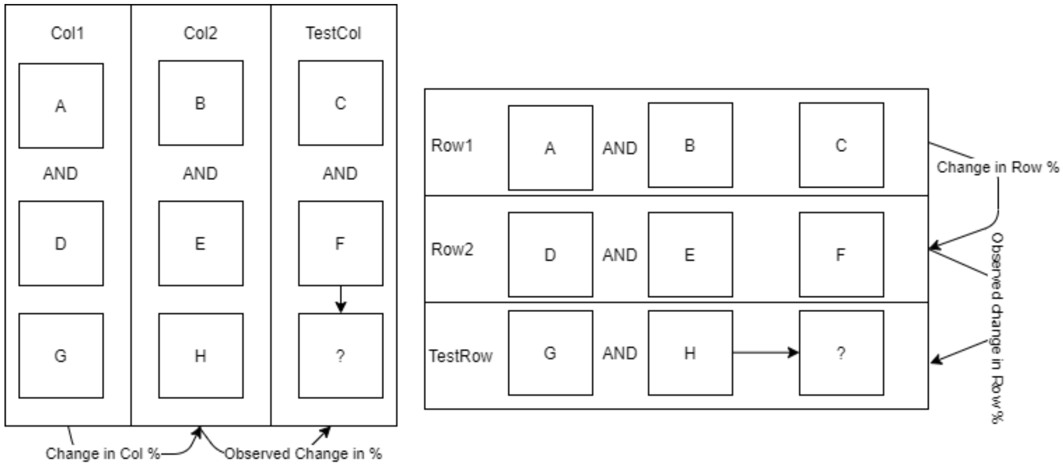

Figure 1 — Visual of column and row aggregate change.

1.2.1 Parameters and Testing

Like the 2x2, parameters and thresholds define cutoffs for the testing. These parameters maximize the performance of the given problem set. In the 3x3 case, the logical Tversky similarity tests have a cutoff of 0.02. In addition, two other parameters control the cutoff values for DPR/IPR. A cutoff of 0.01 controls the tolerance for the single row, col, etc. test. And, a threshold of 0.12 controls the aggregated changes in rows, cols, etc.

1.2.2 Voting in the 3x3 case

If the calculated statistic is within tolerance, then that answer set image receives one vote. In the 3x3 case, the vote is not scaled and merely represents a 1 for meets the threshold and 0 for does not meet. Votes are aggregated and the answer set with the most votes is the returned answer.

2 PERFORMANCE

The agent successfully answers 61 of the 96 problems in Gradescope with a runtime of 9.228 seconds. In the local test environment, it correctly solves problems in the following problem sets: 8 of set B, 9 of C, 9 of D, 5 of E, 1 of challenge B, 5 of challenge C, 4 of challenge D, and 3 of challenge E.

2.1 Agent's Success and Intended Function

The agent succeeds at employing a vast number of heuristics and testing these heuristics on the given problem set for example three problems show the robustness of the agent. In each example, a table exists to show the reader what answer image each heuristic votes for.

For example, in Basic Problem C-10, the agent tests all Tversky and ratio (DPR/IPR) heuristics. These tests combine to correctly answer a commutative trend. In problem C-10 shown in Figure 2, the top left cell progresses in both the row and column direction. The results of the testing and voting are in Table 1.

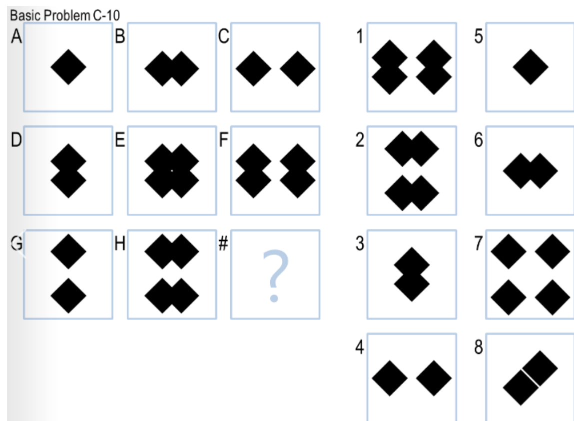

Figure 2— Representation of Basic Problem C -10.

Interpreting Table 1, the first row represents that the Heuristic for the change in DPR of A & B to C and D &E to F is within tolerance to the change in DPR from D&E to F and G & H to the answer image. However, this is only for answers 1, 2 and 7. The combination of all other heuristics shown in Table 1 in aggregate choose the correct answer, answer 7.

Table 1 — Vote summary of AI Agent for Basic Problem C – 10.

Heuristic Used	**Answer**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Ratio-Agg_Row_Change-DPR	1	1	0	0	0	0	1	0
Ratio-Agg_Col_Change-DPR	1	1	0	0	0	0	1	0
Ratio-Agg_Col_Change-IPR	1	0	0	0	0	0	1	0
Ratio-Agg_Left_Change-DPR	1	1	1	1	0	1	1	1

Heuristic Used	Answer							
	1	2	3	4	5	6	7	8
Ratio-Agg_Left_Change-IPR	0	0	1	1	0	1	0	1
Ratio-Row1_Row2-DPR	0	0	0	0	0	0	1	0
Ratio-Agg_Row_Change-IPR	0	0	0	0	0	0	1	0
Ratio-Col1_Col2-DPR	0	0	0	0	0	0	1	0
Total	**4**	**3**	**2**	**2**	**0**	**2**	**7**	**2**

Further, in Basic Problem D – 07, the agent can correctly classify complex relationships through an amalgamation of different heuristics. As shown in Figure 3, the problem has a left diagonal trend between shapes. The agent correctly identifies 3 heuristics as displayed in Table 2 for the left diagonal relationships.

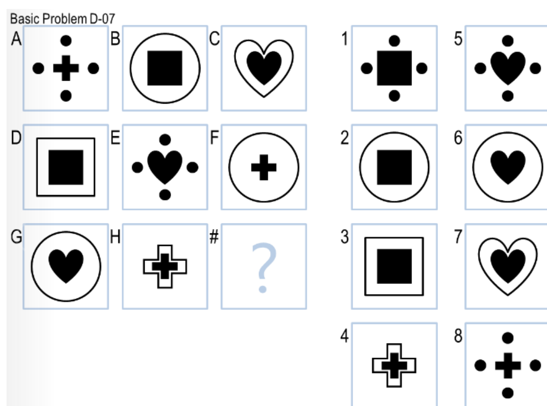

***Figure* 3** - Representation of Basic Problem D- 07.

However, it is interesting to note that the agent believes this problem has a trend at the row level as represented by the *Ratio-Agg_Row_Change-IPR* heuristic. Further, the strongest case for the left diagonal comes from the comparison between the single diagonals test as indicated by the *Ratio-Left1_Left2-IPR* heuristic. This signifies that similar progressions of shapes existed in a left diagonal direction.

Table 2 - Vote summary of AI Agent for Basic Problem D– 07.

Heuristic Used	**Answer**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Ratio-Agg_Row_Change-IPR	1	1	1	0	1	1	1	1
Ratio-Left1_Left2-IPR	1	0	0	0	0	0	0	0
Ratio-Agg_Left_Change-DPR	1	1	1	1	1	1	1	1
Ratio-Agg_Left_Change-IPR	1	1	1	0	1	1	1	0
Total	**4**	**3**	**3**	**1**	**3**	**3**	**3**	**2**

In addition, the agent also uses the combination of heuristics to correctly answer the reflection problem in Basic problem C-07 as shown in Figure 4. Here the agent identifies that there is an aggregate reflection from C to G as a special case. The agent then applies the reflection to A and tests which answer Tversky matches the reflected image. This special case receives an additional vote to stand out from the other votes.

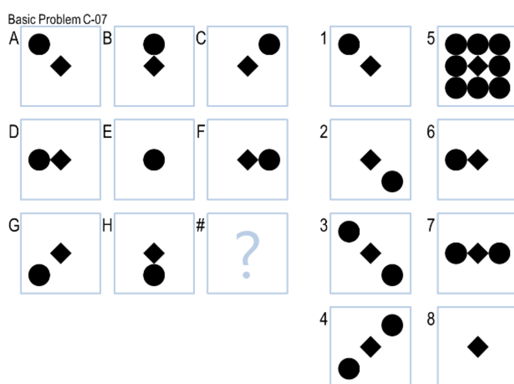

Figure 4 - Representation of Basic Problem C- 07.

2.2 Agent's Shortcomings and Unintended Noise

One of the downsides of the approach in this paper is the unintended noise that comes with the aggregation of so many heuristics. In some cases, the heuristics' votes send the incorrect signal.

For instance, in Basic Problem E-05, the agent does not have a heuristic to answer this type of problem. Yet, the agent finds "within" tolerance DPR and IPR relationships on a row, column, and right diagonal perspective. The agent returns the answer 1 as this has the most votes albeit incorrect votes.

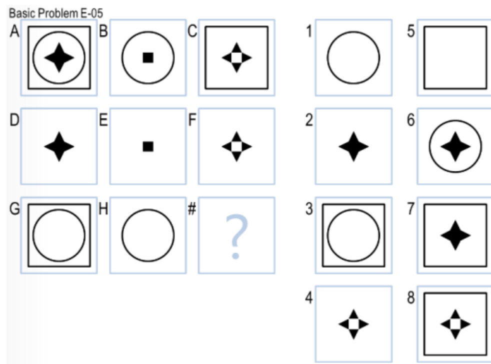

Figure 5 - Representation of Basic Problem E- 05.

Similarly in Challenge Problem E-07, the agent does not have a heuristic to address the addition of vertices. However, the agent tries to find a relationship and chooses the answer with the largest number of votes.

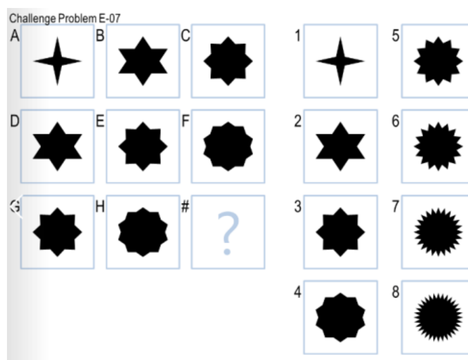

Figure 6 - Representation of Basic Problem E- 07.

3 COMBINED APPROACH

The approach taken for this agent varied. At first, 2x2 problems could easily be solved by using a combination of transformations and comparing the transformed image to the answer set with Tversky similarity. However, as complexity grew with 3 x3 problems, the combination of transformations became cumbersome in processing. A quicker and simpler solution needed to be found. Thus, I started researching and discovered Joyner's et all (2015) paper with the success of agent 3. Agent 3 used DPR and IPR with a voting system.

The DPR / IPR approach granted the agent efficiency in deducing pixel relationships with quite a small amount of code. Yet, the first iterations of this approach relied on very simple heuristics and correspondingly, had poor performance. Therefore, more heuristics on the row, column, and diagonals level were added. Though, the performance of the 2x2 matrix suffered. Ultimately, the two approaches were combined.

4 COMPARISON TO HUMAN

The intended design of this agent attempts to mimic a human's understanding of a problem; start with a host of heuristics and choose the ones that have the most support. The Tversky similarity heuristics of affine transformations mimic this. The agent effectively asks what if this image is rotated and is it like the next image. This action is very similar to a human.

However, the DPR/IPR does not make much sense to a human. The number of intersecting pixels does not mean a whole lot. While the computer can easily calculate the intersection pixel ratio, a human would find it tedious and frivolous. Humans can deduce patterns from changes in shapes as entire objects and not changes in relative number of dark colored pixels between two images.

5 REFERENCES

1. Joyner, D., Bedwell, D., Graham, C., Lemmon, W., Martínez, Ó., & Goel, A.K. (2015). Using Human Computation to Acquire Novel Methods for Addressing Visual Analogy Problems on Intelligence Tests. ICCC.
2. Kunda, M., K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. Cognitive Systems Research, 22-23, 46-66.
3. Tversky, A. (1977). Features of similarity. Psychological Review, 84(4), 327-352.