A Production System to (almost) solve 3x3 Ravens Progressive Matrices

Manikandan Ravikiran
mravikiran@gatech.edu

1 INTRODUCTION, SOLUTION PROPOSAL & ALGORITHM

In this report, a production system (agent) to solve 3x3 Raven's Progressive Matrices (RPM) (Raven, 1962) is presented (Figure 1b) . The developed agent solves **12/12 on Basic, 10/12 on Challenge sets and 6/12 across Test/Raven's** respectively. The report explains details on the incremental design of the solution, error analysis, efficiency, generality, cognitive connection and concludes with some implications on future work.

Knowledge Representation, Agent Reasoning and Design: Inline to Project 1[1] in this work, a production system was developed **using pixel based visual representations** for images (Figure 1a). The production system consists of series of rules, which was designed and developed, to solve one or more RPM at a time. These rules consists of two parts namely ***if case*** which identifies the type of problem and ***then*** case that identifies the solution, provided ***if case*** is valid. Both of these exploit various relation ships (Kunda, McGreggor, and Goel, 2013) that exists across rows/column/diagonals within the RPM images (Figure 1a). Further these rules use **Root Mean Square (RMS)** and **Euclidean distance (ED)** metrics for image similarity estimation.

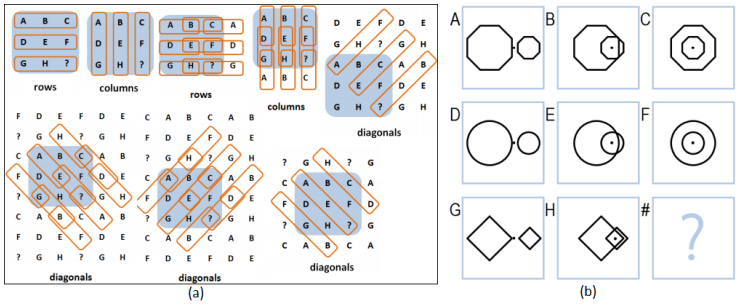


Figure 1—Affine symbolic Reasoning (a) and Sample 3x3 RPM (b).

The final developed production system consists of **17** production rules a.k.a ***if else*** cases incrementally developed by testing on auto-grader, which encompasses

1 Detailed description of algorithm used in project 1 is in section **??**

various relationships in RPM's. During processing, each of the rules is executed in sequence and whenever the input RPM violates a given rule, the agent moves onto the next rule otherwise computes the result and outputs the corresponding answer choice. These ***17*** rules are explained across the submissions section 2.1 - 2.10 and consolidated rule list is in appendix section **??**.

Performance Evaluation Metrics: Performance of the agent is accessed using accuracy, efficiency & generality metrics. Also, errors are categorized as **Wrong Principle (WP)** and **Incomplete Correlate (IC)** (Kunda et al., 2016) highlighted in blue and orange (Tables 1-9) & in-depth descriptions in sections **??** & **??**.

2 EXPERIMENTAL RUNS AND DISCUSSION

This section presents and analyzes, submissions tested on the auto grader. Each submission's description begins with selecting one or more RPM's, followed by its analysis and solution description development, ending with cognitive connection, errors and improvement proposals. **Due to page limitations, the solution to problems are explained in brief and in-depth explanation with code snippets of rules are in appendix section ??, for understanding purposes.** Also, basic problems and challenge problems are represented by **BP** and **CP** respectively.

2.1 Submission-1: Solving BP-1 to BP-7 (2019-09-24 15:44:48 UTC)

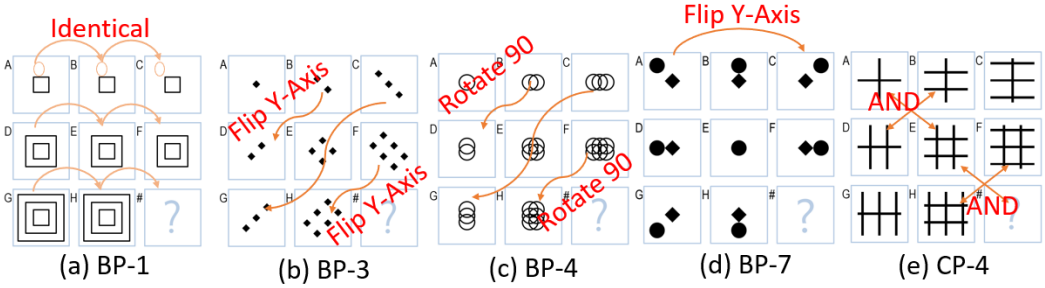


Figure 2—Examples of relationship across BP's and CP's.

Intuition: To begin with, **manual analysis of BP-1 to BP-7 was carried out**. As seen in Figure 2, these problems satisfy ***identity, reflection and rotation*** relationships. Hence, **rules 1, 2, 3.1 & 4** (Figure 3) was added to the agent, with an option to skip when none of the rules solve the problem. This submission uses only ***RMS for similarity estimation***. The results so obtained are in Table 1, with total execution time of 6.8 secs. Code snippets are added in section **??**.

1	Identity	If A=B and B=C then find image same as H from answer choices.
2	DiagonalFlip	If **Flip**(C)=G and **Flip**(F)=H and **Flip**(B)=D then find image like E from the answer choice
3.1	Rotate270	If **Rotate270**(C) = G and **Rotate270** (F) = H and **Rotate270** (B) = D, then find image like E from the answer choice
3.2	LogicalRotate270	If **Rotate270**(C) = G and **Rotate270** (F) = H and **Rotate270** (B) = D and A **&&** B = E **&&** A, then find answer S such that E **&&** S = H **&&** F
4	RowFlip (A, C)	If **Flip**(A)=C and **Flip**(D)=F, then find image like Flip(G)

Figure 3—Rules developed for submission-1.

Cognitive Connection: RPM's when subjected to human testing, reveal incremental nature in human solving process (Carpenter, Just, and Shell, 1990). In this submission, the agent simulates this, by analyzing various relationships such as rotation, reflection and identity to find the solution. At the same time, while human's exhibit meta-cognition while producing the solution, the agent doesn't gauge certainty of proposed solution and lacks skills to think on whether it should return an answer or not.

Table 1—Results from submission-1. ✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are higlighted.

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✗	✓	✓	✓	✓	✓	✈	✈	✈	✗	✈	6/12
Challenge	✈	✓	✈	✗	✈	✈	✈	✈	✈	✈	✈	✈	1/12

	Accuracy
Test	3/12
Ravens	2/12

Error Analysis: The agent produces a score of 6/12 on Basic and and high error across the rest of the sets. Analysis of errors reveal that for **CP-4**, ***even though problem identification through rotation relationship was correct, the process of solving was incorrect***, resulting in error (See Figure 3e). Further, the agent only solved CP-2, which adheres to diagonal flip property, suggesting better rules are needed to handle CP's.

Improvement Proposal: The limitations of the agent could be improved considerably by adding new rules to tackle skipped problems and **solving CP-4 through a different solution logic.**

2.2 Submission-2: Solving CP-4 (2019-09-26 15:50:48 UTC)

Improvement Intuition and Rule 3.2: Based on error analysis from submission-1 , rule 3.1 was modified to tackle CP-4. As seen in CP-4 (Figure 3e), in addition to satisfying rule-3.1, it can also be seen that ***Overlay(B,D) = Overlay(A,E)*** and ***Overlay(C,E)= Overlay(B,F)*** and ***Overlay(D,H) = Overlay(E,G)***. Among these patterns, for this submission, main diagonal relationship is considered and rule-

3.2 (Figure 3) i.e. ***LogicalRotate270*** is introduced. Problems are still skipped, when none of the rules fail to come through. So the agent has **rules 1,2,3.1,3.2 & 4** with **RMS similarity** to obtain results as shown in Table 2 with total execution time of **6.52 secs**. Code snippets of rules are in section **??** and description in Figure 3.

Cognitive Connection: After introducing rule 3.2, the agent still mimics the way a human would think about the solution of overlaying. ***Some extent***, because while identity and reflection follow human thinking, overlay of patterns through **Logical AND** operation, is different from human reasoning.

Table 2—Results from submission-2. ✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted.

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✗	✓	✓	✓	✓	✓	✈	✈	✈	✗	✈	6/12
Challenge	✈	✓	✈	✓	✈	✈	✈	✈	✈	✈	✈	✈	2/12

	Accuracy
Test	3/12
Ravens	3/12

Error Analysis and Improvement Proposal: Problems CP-4 was solved, without any other errors in BP's. However within BP's, the rule-1 also produced wrong answer for BP-11, where the images A,B,C are very similar to one another. Further, investigation on skipped problems reveals that ***BP-11 and CP's 1,5 & 7 satisfy XOR relationships*** as shown in (Figure 4). Modifications to accommodate previous observations should improve the results across CP set. Each of these improvements are addressed in upcoming submissions.

2.3 Submission-3: Solving BP-11, CP 5,7 & 1+ (2019-09-27 09:43:42 UTC)

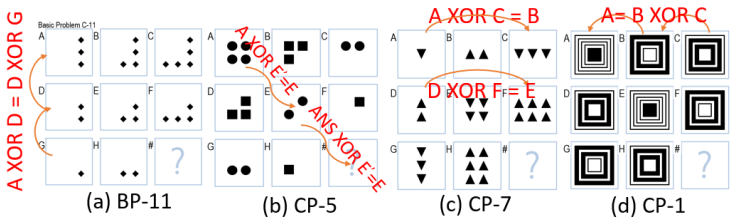


Figure 4—Patterns and relationship in BP-11 (a), CP-5 (b), CP-7 (c) & CP-1 (d).

Improvement Intuition and Rule: Based on analysis from submission-2, we can see that for BP-11 (Figure 4), the ***problem follows XOR relationship*** where ***XOR(A,D)=XOR(D,G)***, similar relationships are true for other columns as well. CP-7 satisfies row wise XOR relationship ***XOR(A,C)=B and XOR(D,F)=E*** and CP-5 satisfies XOR relationship diagonally where ***XOR(A,E)*** $\approx E' \approx$ ***XOR(E,Answer)***.

As such for these we include **rules 5, 6 and 7** respectively (Figure 5). Also CP-1 shows a unique property where ***XOR(B,C)=A***, this was added as rule-8. Overall for this submission, **rules 1-8** are used with a combination of **ED** and **RMS** metrics.

5	**AdjrowXOR**	If **XOR**(A, D)=**XOR**(D, G) and **XOR**(H,E)=**XOR**(B, E) then find answer S such **XOR**(C,F)=**XOR**(F,S)
6	**AdjcolXOR**	If **XOR**(A, C)=B and **XOR**(D, F)=E then find answer S such **XOR**(G,S)=H
7	**AdjDiaXOR**	If **XOR**(A, E')=E then find answer S such **XOR**(E', S)=E

Figure 5—Rules developed for submission-3.

Performance: The agent solves BP-11 & and also CP's 1, 5, & 7 leading to results in Table 3 with **total execution time of 13.43 secs**. Code snippets are in **??**.

Table 3—Results obtained from submission-3.✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted.

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy			Accuracy
Basic	✓	✗	✓	✓	✓	✓	✓	✈	✈	✈	✓	✈	7/12		Test	3/12
Challenge	✓	✓	✗	✓	✓	✈	✓	✈	✈	✈	✈	✈	5/12		Ravens	3/12

Cognitive Connection: Introducing XOR operation, causes the agents behavior and reasoning strategy to be logical than the previous submissions. Moreover, the agent is now a matured production system for solving BP's with production rules in long term memory for solving RPM's. Further this production system has a natural sync with human thinking (Axten, 1973) and the RPM solving process (Carpenter, Just, and Shell, 1990). At the same time, there is no inductive learning embodied in the agent for self-adaptation to newer unseen problems.

Analysis & Proposed Improvement: While the the rules does improve results across all the sets, it also adds in few errors especially **rule-6,** which **generates wrong answer for CP-3**. Moreover, it can be seen that since submission-1, BP-2 is still incorrect. Additionally multiple BP's and CP's are skipped. ***Analysis of BP-8, 10 and 12 reveals that these problems satisfy combination of LogicalRotate270 and XOR rules. Also from visual analysis of BP-2 (Figure 6), we can see the problem consists of same image under various scales (zoomed)***.

2.4 Submission-4 : Solving BP-2,8,10 & 12 (2019-09-27 14:53:21 UTC)

Improvement Intuition and Conflicts: Based on analysis from submission-3, we can see that for BP-8 and BP-12 (Figure 6), follows *Rotate270* along sub-diagonals where ***Rotate270(B)=D, Rotate270(F)=H* and *Rotate270(C)=G***. In case of BP-10,

the problem follows ***Rotate90*** along sub-diagonals. However, between BP-8 and BP-12 there is a conflict, where both the problems could be identified by *Rotate270* relationship, but logic of computing answer is different. For BP-2, the affine relationship is not straightforward. First look at BP-2, shows the problem looks uses scaling transformation, where the same object is scaled across the rows and columns with consistent ratio.

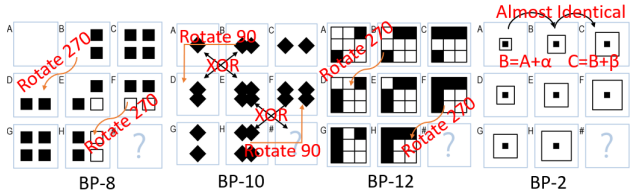


Figure 6—Patterns and relationship in BP-8 (a), CP-10 (b), CP-12 (c) & BP-2 (d).

Rules: The detailed explanation & code snippets is in section ??. To begin with, from BP-2, it can be seen that **i)** ***scale increases across the images in rows*** **ii)** ***images A,D,G is similar to A,B,C*** **iii)** ***images A,B,C is smaller than D,E & F*** **respectively**. These information together is used to create rule-9 a.k.a ***TowardsIdentity*** (See Figure 1) which checks for validity of previous observations, if yes then the solution is obtained by identifying answer choice closer to H. This rule was tested on BP's 8, 10 & was adapted with some minor changes (See section **??** for in depth explanation) to solve BP-8 leading to rule-10 a.k.a ***ATowardsIdentity*** and BP-10 leading to rule-11 a.k.a ***ComplexRule***. BP-12 was solved by exploiting diagonal rotate relationship resulting in rule-12 a.k.a ***DiagonalRotate270***. (See Figure 7)

9	**TowardsIdentity**	If **ED**(B,A)>**ED**(C,B) and **ED**(D,E)>**ED**(E,F) and **ED**(A,B,C)<=**ED**(A,D,G) and **ED**(D,E,F)>=**ED**(A,B,C) the find answer **S** such that **ED**(G,H)>=**ED**(H,**S**)
10	**ATowardsIdentity**	If **ED**(B-A)<=**ED**(C-B) and **ED**(D-E)>=**ED**(E-F) and **ED**(A,B,C)=**ED**(A,D,G) and **ED**(D,E,F)>**ED**(A,B,C) the find answer **S** such that ED(G,H)>=**ED**(H,**S**)
11	**ComplexRule**	If **ED**(A,B,C)> **ED**(A,D,G) and **ED**(A,D)<= **ED**(D,G) and **ED**(B,E) <= **ED**(E,H) and **ED**(A,D,G)> **ED**(B,E,H) and **ED**(C,F) > **ED**(B,E) > **ED**(A,B) and **ED**(E,F) > **ED**(E,H) and **ED**(F,H) > **ED**(B,F) and **ED**(B,D)> **ED**(C,E) and **ED**(BC)-**ED**(GH) <**threshold** then find choice **S** such that **ED**(**XOR**(A,E),**XOR**(E,S)) is minimum
12	**DiagonalRotate270**	If **Rotate270**(B) = D and **Rotate270** (F) = H and **Rotate270** (C) = G, then find answer **S** such that **ED**(**S**, H)=**ED**(F,H)

Figure 7—Rules developed in submission 4.

Performance & Cognitive Connection: The agent solves BP's 2,8, 10 & 12 leading to results in Table 4 with **total execution time of 9.6 secs**. The agent after submission, is still a production system behaving like human. Also we can see ad-

ditional behaviors, namely multiple hypothesis analysis where the agent solves RPM's through multiple hypothesis in individual or combination like humans, common sense reasoning where the agents design uses similar rules for problems BP-2 and BP-8 based on sense of similarity in problem types and case based reasoning where agent adapts the existing solution for new cases.

Table 4—Results obtained from submission-4.✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted.

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✓	✓	✓	✓	✓	✓	✓	✈	✓	✓	✓	11/12
Challenge	✓	✓	✗	✓	✓	✗	✓	✓	✈	✈	✓	✗	7/12

	Accuracy
Test	3/12
Ravens	4/12

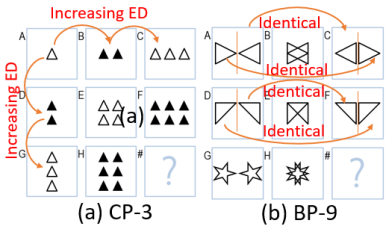
Analysis & Proposed Improvement: From Table 4 we can see that BP-9 is yet to be solved along with some of the erroneous and skipped CP's. Analysis of ***BP-9 (See Figure 7a) shows that problem requires part wise comparison of images***. Further for CP-3, we can see that its similar to BP-10, especially moving vertically from (A to G) the number of dark pixel increases meaning the **ED** with increase, as such rule-11 may be adapted for solving CP-3. Similarly complex rules could be adapted fro CP-6 and 7 as well.

2.5 Submission-5 : (2019-09-27 15:44:07 UTC)

Submission-5 was an incorrect submission, where agent of submission-4 was re-submitted again to bonnie by mistake. Hence no change in results and analysis.

2.6 Submission-6 : Solving BP-9 & CP-3 (2019-10-01 11:58:14 UTC)

Improvement Intuition: Based on analysis from submission-4, it can be seen in for BP-9 vertically dividing the images A and comparing as shown in Figure 7a should solve the problem. For CP-3 again there is increasing dark pixels as we move from A to G. Moreover all the variations previously (section 2.4) seen in BP-10 is equally valid for CP-3 as well.



13	Parts Compare	If **LeftPart**(A)=**RightPart**(C) and **RightPart**(A)=**LeftPart**(C) then find answer S such that **LeftPart**(G)=**RightPart**(S) and **RightPart**(G)=**LeftPart**(S)
14	AComplex Rule-1	If **ED**(A,B,C)> **ED**(A,D,G) and **ED**(A,D)<= **ED**(D,G) and **ED**(B,E) <= **ED**(E,H) and **ED**(A,D,G)> **ED**(B,E,H) and **ED**(C,F) > **ED**(B,E) > **ED**(A,D) and **ED**(E,F) > **ED**(E,H) and **ED**(F,H) > **ED**(B,F) and **ED**(B,D)> **ED**(C,E) and **ED**(BC)-**ED**(GH) <threshold then find choice S such that **ED**(**XOR**(A,E),**XOR**(E,S)) is minimum

(c)

Figure 8—CP-3 (a), BP-9 (b) & Rules (c) used in submission 6.

Solving BP-9: To solve BP-9, in this submission we introduce rule-13 a.k.a ***PartsCompare*** which works by separating the images into two halves and comparing the alternative halves as shown in Figure 7a.

Solving CP-3: As a sanity check ***rule-11 was run on CP-3. Analysis revealed that most parts of the rules are valid except for comparison between A and D.*** Previously in problem BP-10, B and D were same, where as in CP-3, this is not true leading to failure of the comparison. As such rule-11 was modified (red highlighted) to accommodate this change, resulting in rule 14 a.k.a ***AComplexRule-1*** as shown in Figure 6c. Detailed explanation in section **??**.

Table 5—Results obtained from submission-6.✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted.

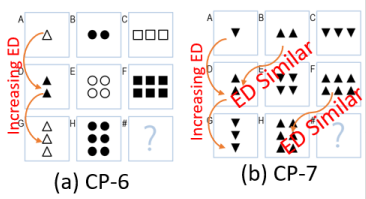
	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
Challenge	✓	✓	✓	✓	✈	✗	✗	✓	✈	✈	✓	✗	6/12

	Accuracy
Test	3/12
Ravens	4/12

Performance: The agent had **rules 1-14** & solves all the problems in Basic set. Test and Raven set show no improvement. Consolidated results are in Table 5.

Cognitive Connection and Error Analysis: No changes in cognitive connection since section 2.4 and no error analysis was done in this submission. Instead, solution to CP's 6 & 7 was developed and submission-7 (Section 2.7) was executed.

2.7 Submission-7 : Solving CP-6 & CP-7 (2019-10-09 13:43:56 UTC)



(a) CP-6 (b) CP-7

15	AComplex Rule-2	If ED(B,C)< ED(A,B) and ED(D,E)<= ED(E,F) and ED(B,E) <= ED(E,H) and ED(A,D,G)< ED(B,E,H) and ED(A,B,C)< ED(D,E,F) and ED(E,F) > ED(E,H) and ED(G,H) < ED(E,F) and ED(A,C) > ED(A,G) and ED(C,F)> ED(G,H) and ED(A,G) <threshold
16	AComplex Rule-3	If ED(A,D)< ED(D,G) and ED(E,F)> ED(E,H) and ED(B,E) <= ED(E,H) and ED(A,D,G)< ED(A,B,C) and ED(B,E,H)> ED(A,D<G) and ED(F,H) > ED(B,F) and ED(B,D) < ED(C,E) and ED(A,D) > ED(D,G) and ED(B,E)> ED(E,H) and ED(C,F) > ED(B,E) > ED(A,D) and ED(BC)-ED(GH) <threshold

(c)

Figure 9—CP-6 (a), BP-7 (b) & Rules 15,16 used in submission 7.

Improvement Intuition and rules: ***Based on analysis from submission-6 for both CP-6 & 7 again it can be seen that there is increase in dark pixels as we move from A to G, moreover some of the variations seen in BP-10 is equally valid for CP-6 & 7 as well. Hence rules 11,14 were oracle tested and adapted*** as shown in Figure 8c & 8d respectively. These are rules 15 (AComplexRule-2) and 16

(AComplexRule-3) respectively in Figure 1. Totally for this submission, we had rules 1-16, with both **RMS** and **ED** metric. Also the rules were re-sorted such that rules with hard thresholds for RMS and ED were given priority in processing.

Table 6—Results obtained from submission-7.✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted.

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
Challenge	✓	✓	✓	✓	✈	✓	✓	✓	✈	✈	✓	✗	8/12

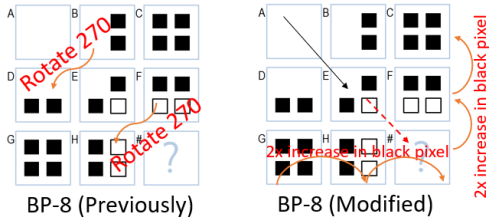
	Accuracy
Test	3/12
Ravens	4/12

Rules and Performance: The agent solves CP-6 and CP-7, resulting in improvement on challenge set, however the test set performance remained same. Final results are as shown in Table 6. **Total execution time is 12.5 secs**.

Cognitive Connection & Error Analysis: At this point of time, we can see the approach takes into account the idea of partial order planning where the rules are sorted such that the net performance is unhurt due to some hard coded condition, thereby helping agent select solution for the problem in hand. As we can see from results, CP 5 and 12 are still erroneous while CP 9,10 are skipped. *Analysis of problems reveal that rule **ATowardsIdentity** developed for BP-8 is responsible for errors in case of CP-6 & CP-12.*

Improvement Proposals: Since ***ATowardsIdentity*** is core cause of the error, the improvement could be achieved either by changing ***ATowardsIdentity*** rule for BP-8, such that the rule developed should be useful only for BP-8 or by building a rule common, valid across the incorrect and correct problems and in turn solves all of them correctly. However, as mentioned earlier ***ATowardsIndentity*** solves BP-8, 12 and CP-12 respectively.

2.8 Submission-8 : Solving CP-8 (2019-10-16 18:23:12 UTC)



17	Pixel Rule	Find answer choice S that has maximum similarity such that **2*NoofBlackPixels**(H)=G and **2*NoofBlackPixels**(F)=C is True

Figure 10—Relationship considered for BP-8 in ***ATowardsIdentity*** (a), ***PixelRule***(b) & Pixel rule description (c)

Improvement Intuition & Rule: Based on analysis from submission-7, the idea

of creating a unique solution to BP-8 was attempted. In that sense, for building unique solution to BP-8, we now move towards pixel based analysis as seen in Figure 10. As it can be seen in Figure 10, while previously we used rotation relationship (Figure a), for this submission we revert to pixel ratio analysis, where analysis of BP-8 reveals that there is an ***increase in black pixel by two times as we move from H to G and F to C***, further we can see the correct answer follows relationship ***(2*NoofBlackPixels(H)-NoofBlackPixels(G)≈ NoofBlackPixels(ANS)*** and ***2*NoofBlackPixels(F)-NoofBlackPixels(C)≈NoofBlackPixels(ANS)*** which forms the basis of the ***PixelRule***. The ***PixelRule*** (See section **??**) tries to identify answer choice that satisfies the mentioned relationship.

Rules and Performance: For this submission, we use rules 1-17 except rule 10 with combination of RMS and ED similarity metrics, resulting in performance as shown in Table 7. Most importantly the **performance on test set and ravens sets improved to 6/12** and the developed rule also solved BP-12 (See Figure 9c).

Table 7—Results obtained from submission-8.✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted. **To be modified**

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	10/12
Challenge	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✗	9/12

	Accuracy
Test	6/12
Ravens	6/12

Cognitive Connection & Error Analysis: Previously, till submission-7 the rules used affine transformations and euclidean distances between pixels to identify the problems and its solution. However, in this submission, the agents solution strategy is expanded to consider pixel ratio information, which can be viewed as **performing incremental concept learning**. Also we can see from results, CP's 7,10 & 12 is still erroneous. More over, the ***PixelRule*** shows a unique behaviour where at the expense of performance over basic and challenge sets, there is an improvement in performance of test set. This suggest that the rules developed are highly tuned towards BP's and CP's and ***PixelRule*** could be reordered and executed first followed by rules with hard thresholds for RMS and ED.

2.9 Submission-9 : Reordering Rules (2019-10-18 13:37:44 UTC)

Improvement Intuition, Rules & Performance: Based on analysis from submission-8, for this submission, ***the rules were reordered*** such that ***PixelRule*** was executed before rules 12-17 and rule 10 was removed. The results obtained are as shown in Table 8. More specifically, reordering rules improved performance of Basic and

Challenge set, without affecting test sets. Results in Table 8.

Table 8—Results obtained from submission-9.✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted.

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
Challenge	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	10/12

	Accuracy
Test	6/12
Ravens	6/12

Cognitive Connection, Error Analysis & Improvement Proposal: No changes in cognitive connection since previous submission and no error analysis was done in this submission. Instead, an investigation was done to see, number of problems that are solve only by the ***Pixel Rule***. Results revealed that it can alone solved 6 BP's and 3 CP's suggesting that the rule is generic and can be even higher preference. As such reordering of rules was done again to see if there is any improvement in solution and submission-10 (Section 2.10) was executed.

2.10 Submission-10 : Reordering Rules Again (2019-10-18 13:37:44 UTC)

Improvement Intuition, Rules & Performance: Based on analysis from submission-9, here the rules were reordered again where the ***PixelRule*** was given preference to execute right after rule-8. Also rest of the rules were kept as it is from previous submission to achieve results as shown in Table 9 and no improvement in result was possible.

Cognitive Connection & Error Analysis: No changes in Cognitive Connection from previous submission and the challenge problems CP-10 and CP-12 are still erroneous. Test set results still needs significant report.

Table 9—Results obtained from submission-9.✓, ✗ & ✈ indicates correct, incorrect and skipped answers. WP & IC are highlighted.

	B-01	B-02	B-03	B-04	B-05	B-06	B-07	B-08	B-09	B-10	B-11	B-12	Accuracy
Basic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12/12
Challenge	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	10/12

	Accuracy
Test	6/12
Ravens	6/12

3 CONCLUSION

Efficiency and Generality: Section 2.1-2.10 incrementally presents various rules designed, modified & adapted to improve results across all the sets. The developed agent achieved 12/12 on Basic, 10/12 on Challenge sets with 6/12 on Test/Raven sets.

The final agent used total of 16 rules that are specific to group of problems as the

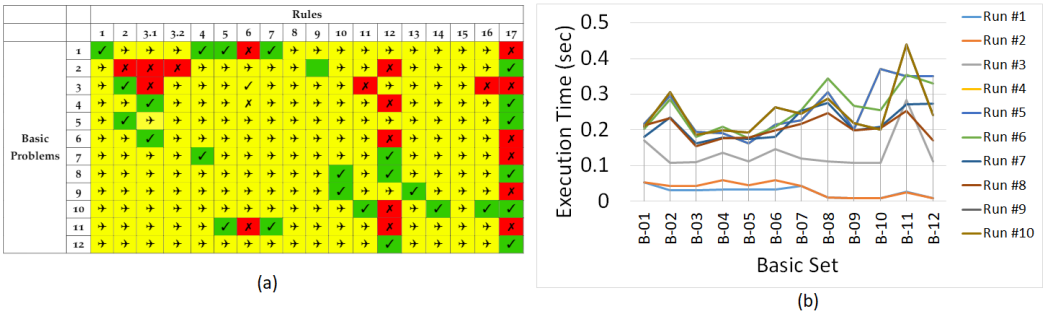


Figure 11—Generality (a) and Efficiency (b) of the agent.

agent was able to address 12 BP's and 10 CP's, with fair amount of success on Test and Raven's set. Table 7 shows the generalization of each rule, across the basic set problems. As we can see, incrementally adding new rules, increases overall coverage of the system, more specifically the ***PixelRule*** covers maximum of 6 BP's, while rest are very specific to each of the problems. Similar behavior was seen even in case of challenge set (See section **??**). Also as we can see, increasing problem complexity increases the agent's time consumption (Figure 11b), the final agent consumes 12 secs to solve all the problems.

Human Cognition and AI: To begin with, the agent's design simulates human thinking while solving RPM's. Typically, humans start to solve by **discerning various relationships between the images in each problem and then apply and adapt these to solve the problems incrementally** by observing each problem, this can be observed across the submissions of the developed production system, where it incrementally solves RPM's by exploiting various relationships.

Considering execution time, human's initial RPM's in negligible time and **as problems become harder, the time consumption also increases**. Such a behavior can be seen in the agent where it spends lesser time on BP-(1,2,3) and more time on BP-(9,10,12). While the agent is not as fast as human's solve RPM's, the pattern of time consumption is very similar.

Further, we can see the agent uses more logical operations such as ***AND, XOR*** in the rules for reasoning on RPM's, this is unlike human behaviour which appears to be logical but doesn't use any logic as part of the reasoning strategy.

Multiple aspects of the agent is similar to human reasoning strategy, firstly **multiple hypothesis analysis** where the agent solves RPM's through combination of multiple strategies like humans, **Common sense reasoning** - where the agents

design uses similar rules for for problems based on commonsense of similarity in problem types and **case based reasoning** where agent adapts the existing solution for new RPM's incrementally.

The **design includes concepts learned during the class**, where the system designed is a **production system** with series of rules with **case-based reasoning** where multiple different thresholds are devised using heuristics to adapt for newer problems, **partial order planning** where the rules as sorted to avoid wrong solutions. Also, similar to project-1 the **errors made by agent are contradictory to human** testing, where most errors made by humans are repetition type (Kunda et al., 2016), while the **agent makes mostly wrong principle type errors**.

Finally, designing the agent based on visual representation, closely relates the agent to human's (Soulières et al., 2009), especially with **human relying on image relationship to solve the problems**.

Design Rectifications and Improvements: With availability of unlimited time and resources, following are the possible changes that can be done, to achieve more accurate results even more efficiently.

- **Threshold:** Inline with project-1, the problem of selecting a threshold still persists, previously it was selected manually however for this submission the threshold selection was empirical. Also at the moment, multiple threshold's are are used to obtain best result. Instead of this, the images could be realigned to get an closed interval of thresholds.
- **Pixel Analysis:** Coverage analysis reveals that ***PixelRule*** alone solves 10 BP's and 4 CP's, erasing the effort on other developed rules, hence pixel analysis based strategies for upcoming sets could be explored to reduce overall development time.
- **Analysis of Generalization:** The rules for CP's were very specific causing significant test set errors, warranting generalizable rules in the future.

4 REFERENCES

[1] Axten, Nick (1973). "Human Problem Solving". In: *Contemporary Sociology* 2.2, pp. 169–170. ISSN: 00943061, 19398638. URL: http://www.jstor.org/stable/2063712.

[2] Carpenter, Patricia, Just, Marcel Adam, and Shell, Peter (1990). "What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test." In: *Psychological review* 97 3, pp. 404–31.

[3] Kunda, Maithilee, McGreggor, Keith, and Goel, Ashok K. (2013). "A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations". In: *Cognitive Systems Research* 22–23, pp. 47–66.

[4] Kunda, Maithilee, Soulières, Isabelle, Rozga, Agata, and Goel, Ashok K. (2016). "Error patterns on the Raven's Standard Progressive Matrices Test". In: *Intelligence* 59, pp. 181–198. ISSN: 0160-2896. DOI: https://doi.org/10.1016/j.intell.2016.09.004. URL: http://www.sciencedirect.com/science/article/pii/S0160289616300149.

[5] Raven, John C. (1962). "Manual for Raven's progressive matrices and vocabulary scales". In: San Antonio, Texas: Pearson.

[6] Soulières, Isabelle, Dawson, Michelle, Samson, Fabienne, Barbeau, Elise Brochu, Sahyoun, Chérif P., Strangman, Gary, Zeffiro, Thomas A., and Mottron, Laurent (2009). "Enhanced visual processing contributes to matrix reasoning in autism." In: *Human brain mapping* 30 12, pp. 4082–107.