# Visual Abstract Reasoning in Computational Imagery

**Yuan Yang**
Department of Computer Science, Vanderbilt University
yuan.yang@vanderbilt.edu

## Abstract

Despite current AI's human-like behavior, super efficiency, and unbelievable ability to handle complex games, we still complain that it shows no sign of creativity, originality, or novelty outside its training set, and that it fails to develop new insights into old experience or establish understanding of new experience. In short, it generates content from its training set, but does not invent content. A fundamental reason for this is that current AI is incapable of abstraction and reasoning in an abstract, generalizable, and systematic way. Think, for instance, of what AI systems we can build if we have a base system that can answer this simple question—when two things are the *same*. Instead of studying these high-level questions, I put my thesis in the context of visual abstract reasoning (VAR), a task widely used in human intelligence tests. A classical example of this task is Raven's Progressive Matrices (RPM, see Figure 1), a family of intelligence tests that was designed to measure eductive ability, i.e., the ability to make meaning out of confusion and generate high-level, usually nonverbal, schemata which make it easy to handle complexity. A similar concept to eductive ability is fluid intelligence, or the ability to discriminate and perceive complex relationships when no recourse to answers is stored in memory. Whether eductive ability or fluid intelligence, RPM points to the qualities that have been lacking in AI. To explore these qualities in AI, I propose the following research questions.

## Research Questions

**RQ 1** Before building any AI system , solving RPM, and claiming victory, it might be worthwhile to probe how RPM can measure eductive ability or fluid intelligence, which is apparently not as simple as weight and height of a person are measured. Moreover, the existing works of computational models for solving RPM, including cognitive systems and AI systems, would provide informative guidance for setting relevant research questions. Therefore, **the first set of research questions of my thesis** is:

- What does RPM measure exactly? And how does RPM measure them?
- What are the entire task domain that RPM represents? i.e., what are the other tasks in the task domain?
- How do current AI systems solve RPM and RPM-like tasks?



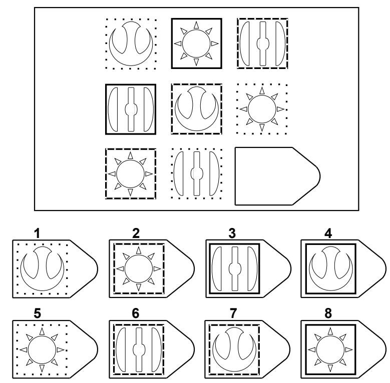


Figure 1: An Example Item of RPM.

To answer these questions, I did a comprehensive review (Yang and Kunda 2023) on the psychometric origin of RPM, other RPM-like tasks for both human and AI testing, and computational models for solving them since 1960s. The main takeaway of this review is that testing this intangible cognitive ability boils down to testing the subject's *learning potential in an unfamiliar environment in a real-time manner*. For example, given an unfamiliar visual puzzle, the subject needs to extract high-level patterns (i.e., abstract concepts) to explain and solve the puzzle; then a second visual puzzle, based on the knowledge obtained in the first puzzle, is presented but requires more intellectual efforts to learn more difficult high-level patterns; then a more difficult third puzzle, based on the knowledge obtained in the second puzzle, is presented; this process repeats until it covers all the difficulty levels that humans can handle. The whole test is a ladder of puzzles where each rung makes it possible for the subject to step on the next rung and the maximum height the subject can reach is a measure of the cognitive ability.

**RQ2** It is hard to imagine how we can build or train an AI that solves completely unknown intelligence tests as human subjects solve them in standard testing sessions. However, the review did bring out a key message that matches our current paradigms of AI—learning potential. In many AI studies, researchers build learning models and test their leaning ability, but in a much simpler setting than human intelligence testing. It seems impossible to directly build a

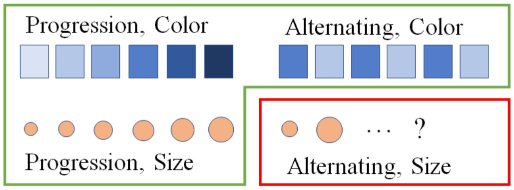


Figure 2: An Example of Nontrivial Generalization.

human-level AI that climbs the RPM ladder in the way human does. Thus, I focus on a single rung and fit our learning paradigm of AI into this rung. A typical way such a rung is constructed is through nontrivial generalization in VAR.

Figure 2 shows a diagram of how we can fit the learning paradigm of AI into the rung of nontrivial generalization. It is a sequence completion task. Learning models are to be trained on sequences like the ones in the green box and to be tested on sequences like the one in the red box. Each sequence is presented visually and characterized by an abstract concept and a perceptual element. The training set contains all the necessary abstract concepts and perceptual elements to solve the test set, but not in the same combinations as in the test set. Extra intellectual efforts are needed to solve the test set, rather than simply applying existing learned knowledge. This kind of generalization is nontrivial in that the test set actually requires to give new meanings to learned perceptual elements and new interpretations to learned abstract concepts—an obvious challenge for current training-testing paradigm. No learning models have shown such generalization ability on benchmark datasets (Barrett et al. 2018).

A possible solution is to implement a dynamic interplay between perceptual and conceptual processing. With the interplay, the perceptual and conceptual processing are encoding the underlying processes of applying abstract concepts on perceptual stimuli and deriving perceptual stimuli from abstract concepts. We initially implemented this solution and achieved satisfying performance on benchmark datasets of trivial generalization (Yang et al. 2023). Thus, **the second research question is to what extent the interplay mechanism between perceptual and conceptual processing can help realize nontrivial generalization in VAR in AI**.

**RQ3** An introspection of how we humans solve RPM would arguably tell that we are solving RPM, a visual abstract reasoning task, visually. In cognitive psychology parlance, we can use mental imagery to solve RPM. Mental imagery is an imagistic representation that can be manipulated mentally. The advocates of mental imagery argue that the way mental imagery functions in human brain is the biological/neural basis upon which other higher cognitive abilities are built. It is important for human cognition and for building AI systems because it allows abstract concepts to be incarnated and applied on visual stimuli to generate mental images. This is possible even when such application of abstract concepts is impossible or nonexistent yet in reality. Thus, mental imagery is an important cognitive ability for flexibility and generalizability in unfamiliar situations and for creativity. Given these advantages of imagery, **the third research question is whether imagery, when implemented computationally in AI, is sufficient for solving VAR tasks**. Initial works have been done to explore this direction (Yang, McGreggor, and Kunda 2020; Yang et al. 2022).

**RQ4** However, mental imagery is probably not the only basis of human cognition. In cognitive science, the ongoing debate over imagery has been lasting for decades but never got resolved. Another major competitor is the propositional representation. Beyond cognitive science, AI researchers are also facing a similar debate over different representations. No matter what representation (imagery or proposition, or both) human cognition uses, it supports robust cognitive ability in all situations, but choosing one representation against another in AI usually means significant limitations in certain scenarios. Thus, the tricky part of using imagery in AI system is how to achieve the human-level robustness.

A possible solution is to extend the pure imagery-based system to an imagery-based production system. A very prototypical example of this kind of solution is generative models, such as autoencoders and GAN. These models are prototypical in that they are flat and static without involving complex structure of abstract concepts, dynamic perceptual and perceptual processes, or multi-step reasoning ability. Consider the situation where a human solves RPM visually: she would inspect matrix entries row by row and column by column, apply the image operations, and generate many intermediate results; she would also make an analogy between rows and columns, which is at a higher level than the concepts embodied by rows and columns; when multiple concepts and perceptual elements were involved, she would iteratively process each of them, possibly retracting previous results and redoing them. Current generative models are not able to implement all these procedures and orchestrate them effectively. **The last research question is thus how to extend a generative model to a production system that is able to produce a flexible reasoning process in imagery**. No work has been done for this research question so far.

## References

Barrett, D.; Hill, F.; Santoro, A.; Morcos, A.; and Lillicrap, T. 2018. Measuring abstract reasoning in neural networks. In *International conference on machine learning*.

Yang, Y.; and Kunda, M. 2023. Computational Models of Solving Raven's Progressive Matrices: A Comprehensive Introduction. Forthcoming, arXiv:2302.04238.

Yang, Y.; McGreggor, K.; and Kunda, M. 2020. Not quite any way you slice it: How different analogical constructions affect Raven's Matrices performance. In *the Eighth Annual Conference on Advances in Cognitive Systems*.

Yang, Y.; Sanyal, D.; Ainooson, J.; Michelson, J.; Farhana, E.; and Kunda, M. 2023. A Cognitively-Inspired Neural Architecture for Visual Abstract Reasoning Using Contrastive Perceptual and Conceptual Processing. Forthcoming, arXiv:2309.10532.

Yang, Y.; Sanyal, D.; Michelson, J.; Ainooson, J.; and Kunda, M. 2022. An End-to-End Imagery-Based Modeling of Solving Geometric Analogy Problems. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.