Clarion and Raven's Matrices

Ron Sun & Can Mekik
Rensselaer Polytechnic Institute

Virtual International Symposium on Cognitive Architecture
June 2020

Clarion is an integrative, psychologically oriented cognitive architecture.

- Connectionist-symbolic hybrid architecture.
- Exhibits a dual-representational structure.
- Addresses learning, skills, reasoning, motivation, metacognition, social interaction, etc.
- Consists of a number of distinct, interdependent, and complementary subsystems with complex interactions.

Sun, R. (2016) *Anatomy of the Mind*. Oxford University Press.

Implicit knowledge is tacit knowledge operating outside of one’s awareness.

- Detected in experimental paradigms such as:
    - Artificial grammar learning (Reber, 1989)
    - Sequence learning (Cleeremans et al., 1998; Seger, 1994)
    - Dynamic system control (Seger, 1994)
    - Probability learning (Evans & Frankish, 2009; Seger, 1994)

Generally associated with the following observations:

- Subjects unable to verbally report certain task knowledge.
- Subjects exhibit such task knowledge in some circumstances: e.g., when presented with forced choices.

Cleeremans, A. and Destrebecqz, A. and Boyer, M. (1998) Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2(10)*, 406-416.
Evans, J. and Frankish, K. (eds.) (2009) *In two minds: Dual-processes and beyond.* Oxford Univresity Press.
Reber, A. S. (1989) Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118(3)*, 219-235.
Seger, C. (1994) Implicit learning. *Psychological Bulletin*, *115(2)*, 163-196.

There are two “levels” of representation in Clarion.

- Dual-representational structure present in each subsystem:
    - The top “level” encodes explicit knowledge
    - The bottom “level” encodes implicit knowledge
- The two “levels”:
    - Interact, for example, by cooperating in action selection, reasoning, and learning.
    - May encode knowledge in a redundant (dual representational) fashion.
- Essentially, it is a dual-process theory of mind.
- In particular, duality of representation has been extensively argued in Sun et al. (2005) and Sun (1994, 2002, 2016).

Sun, R. (1994) *Integrating Rules and Connectionism for Robust Commonsense Reasoning*. John Wiley and Sons.
Sun, R. (2002) *Duality of the Mind*. Lawrence Erlbaum Associates.
Sun, R (2016). *Anatomy of the Mind*. OUP.
Sun, R.; Slusarz, P. and Terry, C. (2005). The interaction of the Explicit and the Implicit in Skill Learning: A Dual Process Approach. *Psychological Review*, *112(1)*, 159-192.

Interaction between the levels may result in synergy.

- Interaction may capture a variety of synergistic psychological phenomena, for example:
  - Speeding up skill learning
  - Improving skill performance
  - Facilitating transfer of learned skills
  - Enabling similarity-based reasoning, including categorical inheritance
  - Supporting creative problem solving
  - And so on
- Some effects relevant to this presentation:
  - Verbalization effects (see, e.g., Sun & Zhang, 2004); and,
  - Performance pressure effects (i.e., choking under pressure, see Wilson & Sun, 2020)
  - Etc.

Sun, R (2016). *Anatomy of the Mind*. OUP.
Sun, R.; Slusarz, P. and Terry, C. (2005). The interaction of the Explicit and the Implicit in Skill Learning: A Dual Process Approach. *Psychological Review*, *112(1)*, 159-192.
Sun, R., & Zhang, X. (2004). Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research*, *5*(1), 63-89.
Wilson, N. R. & Sun, R. (2020). A Mechanistic Account of Stress-Induced Performance Degradation. *Cognitive Computation*, https://doi.org/10.1007/s12559-020-09725-5

Motivation is captured by goals and drives.

- Need to capture and explain why an agent does what it does.
- Simply saying that an agent chooses actions to maximize gains, rewards, reinforcements, or payoffs leaves open the question of what determines these things.
- Drives and their interactions lead to goals and actions (Murray, 1938; Toates, 1986):
    - Low-level primary drives (mostly physiological): hunger, thirst, physical danger, etc.
    - High-level primary drives (mostly social): affiliation and belongingness, power and dominance, fairness, etc.
    - There are also secondary ("derived") drives, which are changeable, and acquired mostly in the process of satisfying primary drives.
- Drive processes are implicit; goal processes are explicit.

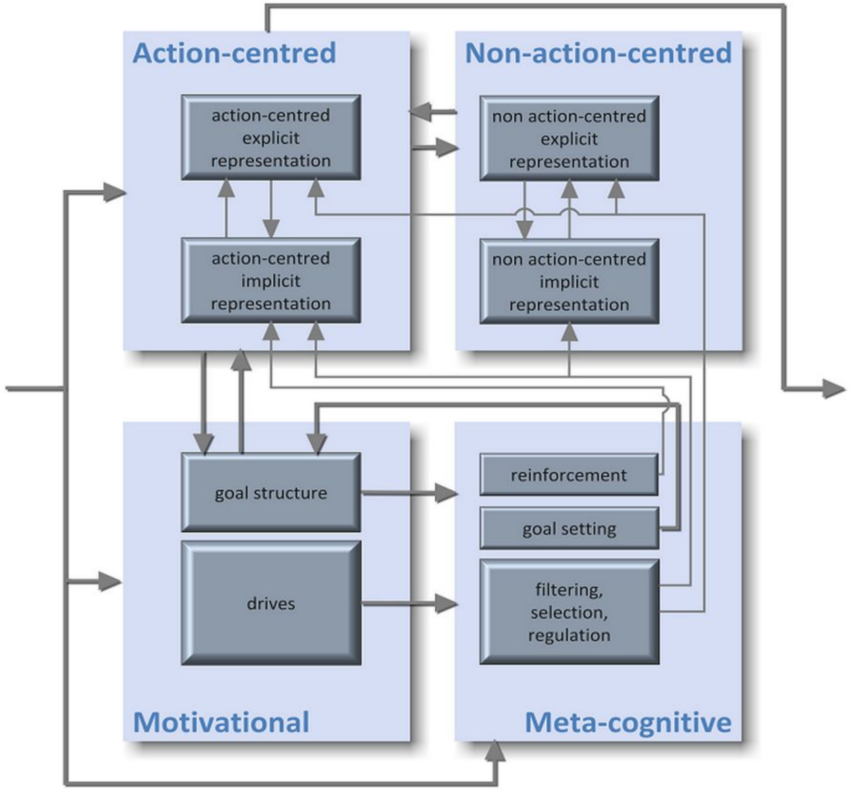
Murray, H. (1938). *Explorations in personality*. Oxford University Press.
Toates, F. (1986). *Motivational systems*. Cambridge University Press.

Meta-cognitive processes are necessary to monitor and regulate processing.

- Meta-cognition refers to one’s knowledge concerning one’s own cognitive processes and products and the control and regulation of them (Flavell, 1976).
- Since an agent may have many goals, drives, and cognitive mechanisms, there is often a need for mechanisms coordinating component processes.

Flavell, J. (1976). Metacognitive aspects of problem solving. In B. Resnick (Ed.), *The nature of intelligence*. Erlbaum Associates.

You can think of Clarion as a network of neural networks.


Action-centred
action-centred explicit representation
action-centred implicit representation
Non-action-centred
non action-centred explicit representation
non action-centred implicit representation
goal structure
drives
Motivational
reinforcement
goal setting
filtering, selection, regulation
Meta-cognitive

- 4 Subsystems:
  - Action-centered Subsystem
  - Non-action-centered Subsystem
  - Motivational Subsystem
  - Metacognitive Subsystem
- Light boxes are subsystems; dark boxes house one or more neural networks.
- Chunk nodes & rules at top level; (micro)feature nodes and implicit networks at bottom level.
- Subsystem outputs selected through competition.

Sun, R (2016). *Anatomy of the Mind*. Oxford University Press.

Chunk nodes encode explicit knowledge, (micro)feature nodes encode implicit knowledge.

- Chunk nodes:
  - Named/labeled
  - Localist representations: One node = one concept.
  - Supporting rule-based reasoning (RBR).
- Feature (microfeature) nodes:
  - Part of a distributed representation: each concept is represented by some combination of features (microfeatures).
  - Supporting implicit reasoning, as well as SBR.

Chunk nodes and (micro)feature nodes are linked.

- A *chunk* is a chunk node together with its links to (micro)feature nodes.
- Activations may flow:
  - In a bottom-up fashion, from (micro)feature nodes to chunk nodes
  - In a top-down fashion, from chunk nodes to (micro)feature nodes
- Activation flow is the main way in which the two levels interact.
- Top-down & bottom-up activation flows may behave differently in different subsystems.

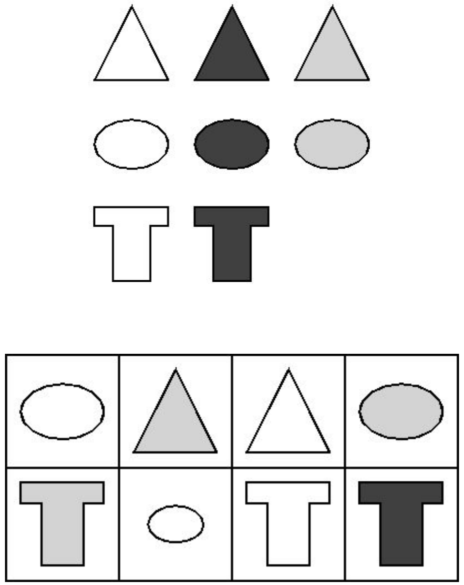
Top-down and bottom-up links support similarity-based reasoning in NACS.

$$S_{A\sim B} \approx N_{A\cap B} / N_B$$

- $S_{A\sim B}$ is similarity form chunk A to chunk B, N is number of features (for A∩B or B).
- Cross-level activation captures SBR: Top-down activations from A activate its features, which includes shared features that, in turn, partially activate B in bottom-up fashion.
- May capture complex human reasoning patterns in conjunction with rule-based reasoning (Sun, 1994; 1995).

Sun, R. (1994) *Integrating Rules and Connectionism for Robust Commonsense Reasoning*. John Wiley and Sons.
Sun, R (1995) Robust Reasoning: Integrating rule-based and similarity-based reasoning. *Artificial Intelligence*, *75(2)*, 241-296.

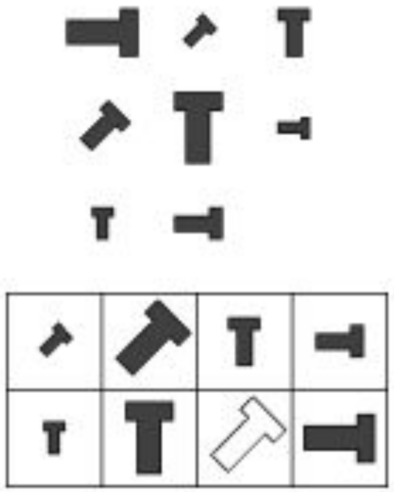
Our recent work involves modeling human performance in Raven's Matrices (RPM).

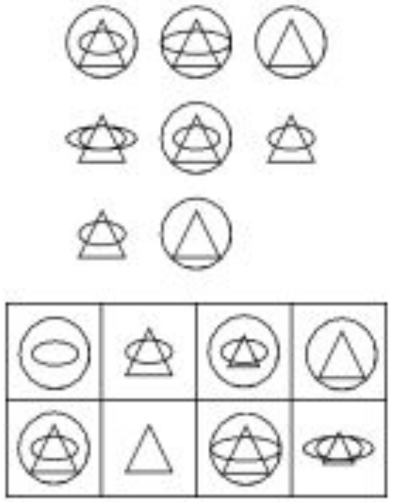


- Strong measure of fluid intelligence.
- Involves solving matrix problems.
- Typically viewed as an inductive or analogical task.
- Working with the Sandia Generated Matrices (Matzen et al., 2010).

Matzen, L. E. et al. (2010) Recreating Raven’s. *Behavior Research Methods*, *42(2)*, 525-541.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven’s Progressive Matrices and Vocabulary Scales: Section 1* (1998 Ed.).

Matrix problems can be quite complex.





Matzen, L. E. et al. (2010) Recreating Raven's. *Behavior Research Methods*, *42(2)*, 525-541.

Clarion may help develop a more detailed understanding of human performance on RPM.

- The dual-representational architecture offers an integrated and parsimonious way to capture:
  - The basic cognitive processes required to solve matrix problems, and
  - Important cognitive and motivational effects.
- In particular:
  - Analogical aspects of the task may be captured as a special case of similarity-based reasoning.
  - The interaction of the two levels may account for various important phenomena such as verbal overshadowing (DeShon et al, 1995), choking under pressure (Gimmig et al., 2006), etc.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices. *Intelligence*, *21*, 135−155.
Gimmig, D. et al. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence. *Psychon Bull Rev,* 13, 1005–1010.

Our work aims to capture cognitive and motivational phenomena related to RPM in an integrated model.

- We aim to capture:
  - The role of implicit versus explicit processes (such as verbal overshadowing, choking under pressure, etc.).
  - Cognitive aspects partially addressed in existing models (e.g., goal management, working memory, high-level visual processing, etc.).
  - Motivational processes/effects (e.g., self-efficacy, anxiety, etc.), not previously tackled.

Carpenter, P. A.; Just, M. A. & Shell P. (1990) What one intelligence test measures. *Psychological Review*, *97*, 404-431.

Kunda, M., McGreggor, 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*, 47–66.

Lovett, A., & Forbus, K. (2017). Modeling visual problem solving as analogical reasoning. *Psychological Review*, *124*(1), 60–90.

Ragni, M., & Neubert, S. (2014). Analyzing Raven’s intelligence test: Cognitive model, demand and complexity. In H. Prade & R. Gilles (Eds.), *Computational approaches to analogical reasoning: Current trends* (Vol. 548), Studies in Computational Intelligence.

Rasmussen, D., & Eliasmith, C. (2011). A Neural Model of Rule Generation in Inductive Reasoning. *Topics in Cognitive Science*, *3*, 140-153.

Matrix problems may be solved by combining implicit and explicit processing.

- Basic ideas:
  - Basic perception accomplished by visual module, as directed by ACS.
  - Visual relations detected by implicit processes in NACS, as directed by ACS.
  - Response selection primarily driven by SBR in NACS.
  - Motivation affects amount and explicitness of processing (through goal setting in MCS) based on drive strengths in MS.
- Work so far: two preliminary models based on these ideas and a third model, in development, within Clarion.

Model 1 selects responses based on similarities in pairwise attribute differences.

- Convolutional Siamese network; detects pairwise attribute differences.
- Model finds invariant features along principal and diagonal axes; computes violation score based on Manhattan distance.
- Works only on a subset of Sandia Matrices.
- Cognitively not particularly plausible.
- 78.7% correct (versus average human performance of 81.25% on the same subset).

Mekik, C. S, Sun, R, & Dai, D (2017). Deep Learning of Raven's Matrices. *ACS2017*.

Model 2 selects responses based on similarities among axial relational features.

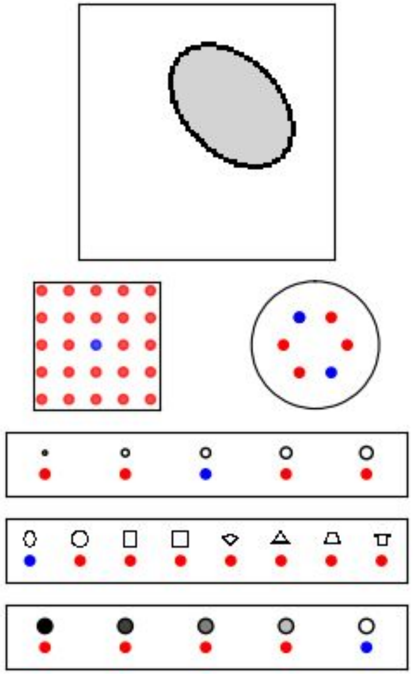
- Convolutional network detects ternary relational features in rows/columns (using 3 identical stacks of convolutional layers converging into fully connected layers).
- Similarity through KL divergence, row/column representations combined through geometric means.
- More general than Model 1.
- More psychologically plausible, yet simpler than Model 1.
- 85% correct (versus average human performance of about 65% on all Sandia Matrices).

Mekik, C. S, Sun, R, & Dai, D (2017). Similarity-based reasoning, Raven's Matrices and General Intelligence. *IJCAI2018*.

Model 3 is a more detailed model, more fully inspired by Clarion and (hopefully) more psychologically plausible.

- Main ideas:
  - Constructing chunks representing characteristic row/column features.
  - Finding the best match of these chunks among the candidates using SBR.
- Steps/Subgoals:
  - Detection of dimensions of interest (those exhibiting variation in values).
  - Selection of a dimension and detection of feature patterns (i.e., relations) in that dimension (accounting for dimensional variations).
  - Resolution of conflicts (among feature patterns).
  - Response selection (based on feature patterns, using SBR).

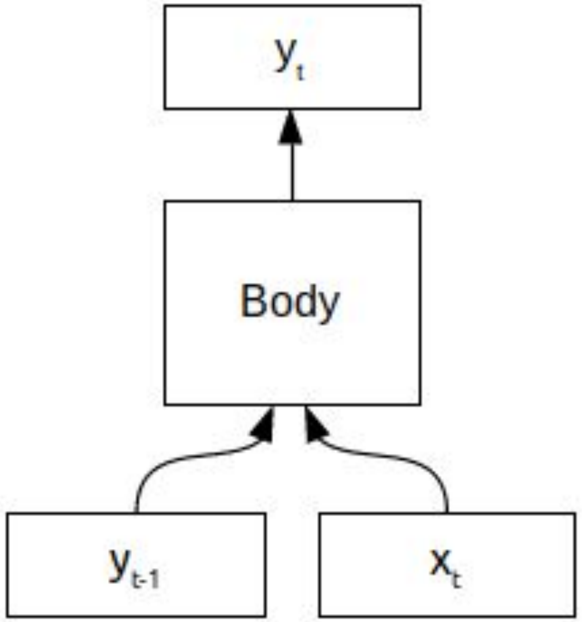
Visual processing: feature integration theory.



- Sequential processing of visual relations (Franconeri et al., 2012; Triesman & Gelade, 1980).
    - Basic features detected in parallel (during fixation): pose (i.e., position, size, orientation), form, and texture.
    - Relational features detected by sequences of fixations (along main axes; see next slide).
- Some simplifications:
    - One fixation point for each panel.
    - Form/texture representations.

Franconeri et al. (2012) Flexible visual processing of spatial relationships. *Cognition*, *122*, 210-227.
Triesman A. M. & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136.

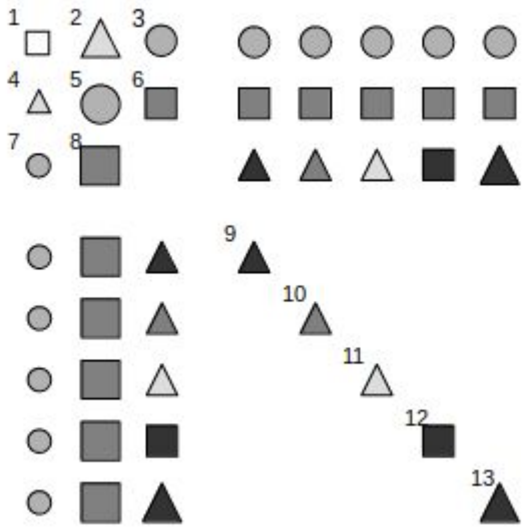
Relation detection is sequential and implicit.

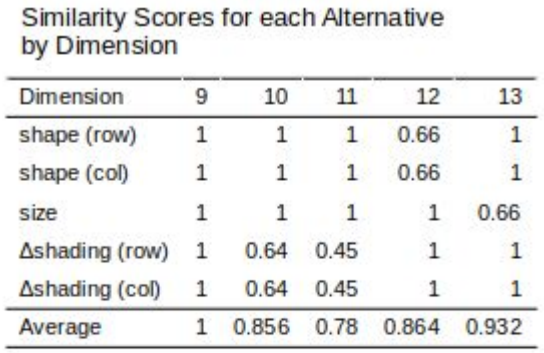


- Relational network operates in the bottom-level of NACS.
- I/O structure:
  - x_t: Object nodes (i.e., basic features).
  - y_t: Relational nodes (e.g. change in size).
  - y_{t-1}: Lagged relational nodes (i.e., previous output).
- Simple, flexible, & expressive; similar to classic recurrent architectures (Elman, 1991; Jordan, 1986).

Elman, J. L. (1991) Distributed Representations, Simple Recurrent Networks and Grammatical Structure. *Machine Learning*, *7*, 195-225.
Jordan, M. L. (1986) Serial Order: A Parallel Distributed Processing Approach.

Model 3 selects responses based on similarities among axial descriptor chunks.



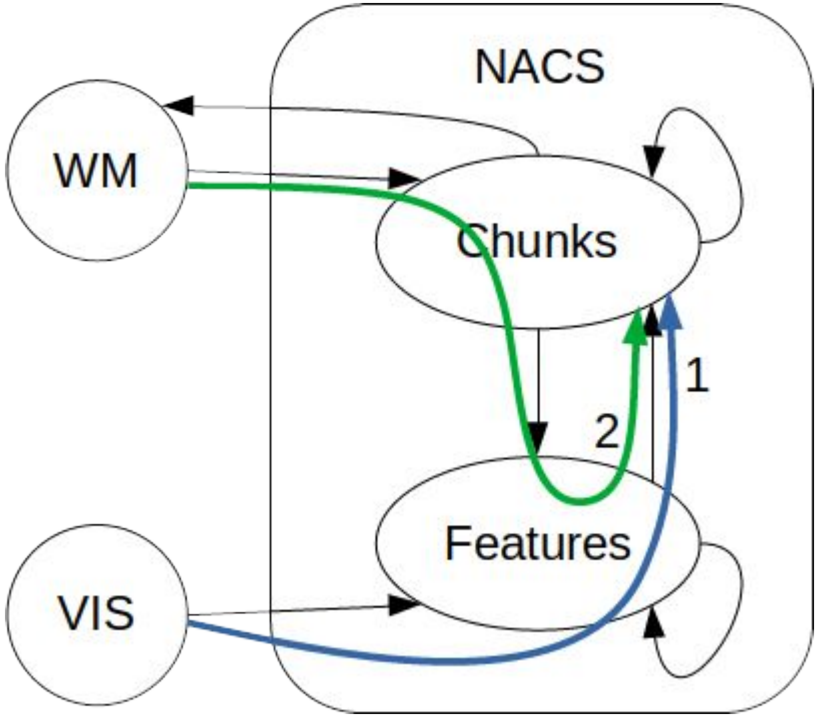


Similarity Scores for each Alternative by Dimension

Dimension	9	10	11	12	13
shape (row)	1	1	1	0.66	1
shape (col)	1	1	1	0.66	1
size	1	1	1	1	0.66
Δshading (row)	1	0.64	0.45	1	1
Δshading (col)	1	0.64	0.45	1	1
Average	1	0.856	0.78	0.864	0.932

- Lower arity relational features processed first.
- Response selection may be based on averaging or incremental elimination.
- Full development in progress.

The model accounts for the role of working memory and control.



- Two basic chunk creation processes:
  1. Chunking of basic & relational visual information.
  2. Creation of row/column descriptor chunks.
- The solution process requires intricate control:
  - WM management: crucial. Chunks must be maintained in NACS and working memory.
  - Chunks are deleted if base-level activation drops below threshold (e.g., due to disuse).
  - Number of steps/subgoals: varies with matrix complexity.
  - Input/output filtering (i.e., attention control): necessary to prevent interference.

Control & WM are recurring themes in the literature.

- Some notable high-level claims/findings:
    - Carpenter et al. (1990): Goal management improves performance by preventing interference and reducing WM load.
    - Engle et al. (1999): WM (but not STM) explains a large proportion of variance (50%) in *g* scores as measured by Raven's and other tests.
    - Embretson (1995): Control processing & WM estimates (from scores) together account for nearly all variance (92%) in a variant of Raven's Matrices. Notably, control processing & WM have both overlapping and independent contributions (70% and 50% respectively when considered alone).
- Our model captures these phenomena and others in a parsimonious way, as discussed in next two slides.

Carpenter, P. A.; Just, M. A. & Shell P. (1990) What one intelligence test measures. *Psychological Review*, *97*, 404-431.
Embretson, S. E. (1995) The role of working memory capacity and general control processes in intelligence. *Intelligence*, *20*, 169-189.
Engle et al. (1999) Working Memory, Short-Term Memory, and General Fluid Intelligence. *JEP*, *128(3)*, 309-331.

Incomplete analyses, due to WM, control, or motivational issues, may result in lower accuracy.

- Response selection does not necessarily require complete analysis. At any time, apparently best alternative may be picked (via softmax).
- Completeness of analyses may be impeded by control or WM issues, as well as motivational variables:
  - Interference among different dimensions of analysis;
  - Information loss due to high BLA decay and/or thresholds; or,
  - Low (or overly high) self-efficacy (through reduced effort, i.e., premature responding).
- The model copes by:
  - Incremental and structured processing through setting goals and subgoals (Carpenter et al., 1990);
  - Periodically reviewing and/or refreshing important information (WM management); or,
  - Regulating self-efficacy (e.g., through implementation intentions, Bayer & Gollwitzer, 2007; Wieber et al., 2010).

Carpenter, P. A.; Just, M. A. & Shell P. (1990) What one intelligence test measures. *Psychological Review*, *97*, 404-431.
Bayer, U. C. & Gollwitzer, P. M. (2007). Boosting Scholastic Test Scores by Willpower: The Role of Implementation Intentions. *Self and Identity*, *6*, 1-19.
Wieber, F., Odental, G., & Gollwitzer, P. (2010). Self-efficacy feelings moderate implementation intention effects. *Self and Identity*, *9(2)*, 177–194. doi:10.1080/15298860902860333

Overly explicit or overly implicit processing, due to control or motivational issues, may also result in lower accuracy.

- Feature patterns are detected primarily through implicit processes (visual module, relational network), but with explicit direction and intervention. This requires a balance of explicit versus implicit processing (Sun et al., 2001, 2005).
- Processing may become overly explicit or overly implicit due to:
  - Performance anxiety (i.e., choking under pressure, see Gimmig et al., 2006), which may impede explicit (i.e., goal-directed) control by encouraging an overly implicit cognitive mode.
  - Verbalization (i.e., verbal overshadowing, see DeShon et al, 1995), which may impede pattern detection by encouraging an overly explicit cognitive mode.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices. *Intelligence*, *21*, 135−155.
Gimmig, D. et al. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence. *Psychon Bull Rev,* 13, 1005–1010.
Sun, R.; Merrill, E. and Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, *25*, 203-244.
Sun, R.; Slusarz, P. and Terry, C. (2005). The interaction of the Explicit and the Implicit in Skill Learning: A Dual Process Approach. *Psychological Review*, *112(1)*, 159-192.

There are many interesting possibilities for extending the model.

- Constructive-matching solution strategy in addition to response-elimination strategy (Vigneau et al., 2006).
- Relaxed assumptions:
  - More generic form/texture representations (e.g., Peinecke et al, 2007).
  - Less restricted fixation options.
- Learn new relational features through, e.g., constructive learning algorithms (e.g., Fahlman & Lebiere, 1990; Kwok & Yeung, 1997)

Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In Advances in neural information processing systems (pp. 524-532).

Kwok, T.-Y.; Yeung, D.-Y. (1997) Constructive Algorithms for Structure Learning [...]. *IEEE Trans. Neural Networks*, *8(3)*, 630-645.

Peinecke, N.; Wolter, F.-E. & Reuter, M. (2007) Laplace spectra as fingerprints for image recognition. *CAD*, *39*, 460-476.

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic infulence on intelligence. *Intelligence*, *34*, 261–272.

The model promises a detailed, mechanistic understanding of human performance on RPM.

- The work is converging towards:
  - A parsimonious, integrative model of human performance on Raven's Matrices accounting for a wide variety of cognitive and motivational phenomena within Clarion.
  - A Clarion-based model of visual processing from pixel inputs to object detection.
  - A novel approach to relation detection/representation in Clarion, relying on implicit processes (as opposed to using complex chunks in Clarion as in Licato et al., 2014a, 2014b; also cf. Lovett & Forbus, 2017).

Licato, J., Sun, R. & Bringsjord, S. (2014a) Structural Representation and Reasoning in a Hybrid Cognitive Architecture. IJCNN2014.
Licato, J., Sun, R. & Bringsjord, S. (2014b) Using a Hybrid Cognitive Architecture to Model Children's Errors in an Analogy Task. Proceedings of CogSci2014.
Lovett, A., & Forbus, K. (2017). Modeling visual problem solving as analogical reasoning. *Psychological Review*, *124*(1), 60–90.

The model may even lead to a more detailed understanding of fluid intelligence.

- Emerging ideas on fluid intelligence from the model:
  - Fluid intelligence is the ability to process relations in a goal-oriented way.
  - Relational processing requires fine coordination among subsystems (e.g., ACS and NACS) and implicit/explicit processes.
  - As a result, fluid intelligence is determined by:
    - Relatively stable architectural features (e.g., modularization of subsystems, similarity through cross-level interaction, BLAs of chunks, etc.),
    - Motivational drive structure & processes (e.g., baseline curiosity and achievement drive activations)
    - (Meta-)cognitive skills.

Thank you for your attention!

- Many thanks to:
  - Selmer Bringsjord, Sergei Bugrov, Brett Fajen, Chris Sims, and Scott Steinmetz for advice and discussions.
  - Laura Matzen for providing Sandia Matrices materials.
- This work was/is supported by:
  - ARI Grant No. W911NF-17-1-0236
  - Rensselaer Polytechnic Institute HASS Graduate Fellowship 2017-2019.