Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Behavioral Sciences

Same/different in visual reasoning

Kenneth D Forbus[1] and Andrew Lovett[2]

Visual reasoning tasks involving comparison provide interesting insights into how people make similarity and difference judgments. This review summarizes work that provides evidence that the same structure-mapping comparison processes that appear to be used elsewhere in cognition can also be used to model comparison in human visual reasoning tasks. These models rely on qualitative representations, which provide symbolic descriptions of continuous properties, an important kind of relational representation. Cognitive simulations of multiple human visual reasoning tasks, using the same model of high-level vision to compute relational representations, achieve human-like performance, both in terms of accuracy and estimating the relative difficulty of problems.

Addresses
[1] Northwestern University, United States
[2] US Naval Research Laboratory, United States

Corresponding author: Forbus, Kenneth D (forbus@northwestern.edu)

Current Opinion in Behavioral Sciences 2020, **37**:63–68

This review comes from a themed issue on **Same-different conceptualization**

Edited by **Edward A Wasserman**, **Jean-Rémy Hochmann** and **Susan Carey**

https://doi.org/10.1016/j.cobeha.2020.09.008

Comparison in visual reasoning tasks

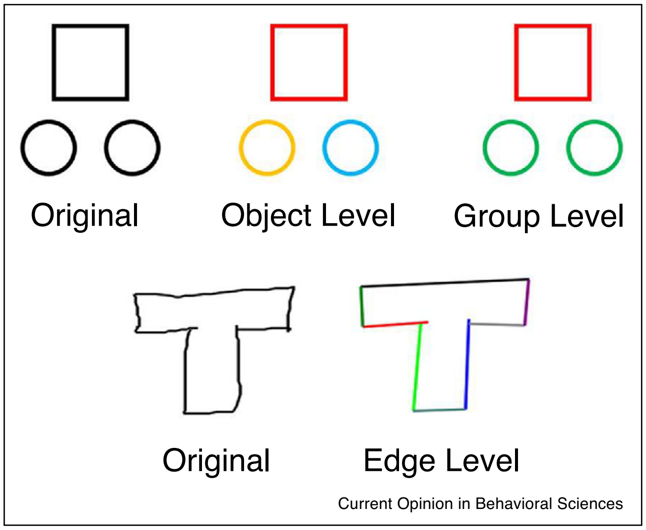
Many visual tasks have been explored in psychology and other areas of cognitive science. We focus here on visual reasoning tasks that involve comparison, since those shed light on how concepts such as same and different are conceptualized. We begin with a summary of hypotheses about visual representations, since the contents of visual representations help determine what are more or less similar, the spectrum for which same and different provide end-points. We then summarize key ideas of structure-mapping theory and the Structure-Mapping Engine, which is used in many of the models below, as well as examine alternatives. We then outline models for three different visual reasoning tasks: Geometric analogies, an Oddity task, and Ravens' Progressive Matrices. For each task we compare and contrast models that have been proposed for them. Finally, we summarize some lessons that can be drawn from these models. A key conclusion is that structure-mapping notions of same/different apply in visual reasoning as well as other areas of cognition.

Human visual representations

The nature of human visual representations is still very much under investigation. However, there is already ample evidence that our visual systems compute *structural* representations, consisting of a set of entities with relationships between them [1–3]. These visual relationships include *qualitative representations* [4••], which provide symbolic abstractions of continuous information. These qualitative representations provide a bridge between perception and cognition. For example, it has long been known that edges are an important intermediate representation computed by our visual cortex [5,6]. Edges, as Marr argued, constitute a primitive form of symbolic representation, where discrete properties are grounded in metric information (i.e. the specific locations construed as part of an edge). Edges themselves can be described in terms of qualitative properties, for example, straight/curved, concavity/convexity, and positions relative to each other. Moreover, networks of edges appear to be important for human shape perception, as indicated by the results of Scholl and his collaborators who studied the role of the medial axis transform in people's shape judgments [7•] and modulate perceived similarity [8]. Networks of edges help segment scenes into regions, which can then be grouped into hypothesized entities. Additional relationships can then be computed between these entities, for example, people are sensitive to qualitative relations in visual same/different tasks [9].

As argued by Palmer *et al.* [10] and others, visual representations are computed at multiple levels, such as grouping into entities based on gestalt principles. The nature of visual entities can vary along a hierarchy [5,11,12], from parts within an object—for example, the edges making up the 'T' shape in Figure 1—to objects, to larger-scale groupings of objects. Critically, as perceivers, humans can strategically focus on different levels in the hierarchy, depending on the task they are performing. For example, if we want to compare two shapes, we can focus on the relationships between each shape's edges, noting parallel edges, convex and concave corners, and relative length. These types of qualitative relationships—which abstract away precise, quantitative details—remain largely constant as shapes rotate and transform in space, making it easier to find commonalities during comparison [1].

Figure 1



Examples of levels of shape representation in CogSketch. Color is used to show CogSketch's decomposition into elements at that level. The top illustrates how a basic object level description can be re-represented into groups based on internal similarity. The bottom illustrates how CogSketch's edge-finding and description capabilities can decompose a shape into edges.

The most comprehensive computational model for such visual representations is CogSketch [13], which automatically computes properties of edges and shapes, given initial networks of edges computed automatically from other inputs. (These inputs include digital ink, stimuli copy/pasted from vector graphics programs such as PowerPoint, and vision systems operating on images (e.g. Ref. [14]). CogSketch uses algorithms from computer vision and the qualitative reasoning community to construct these representations. These include relationships between entities, for example, positional relations (e.g. above, left) and qualitative topological relations (e.g. inside, touching, disjoint). Where possible, these are constrained from results from the vision science literature (e.g. its representational repertoire includes medial axis transforms, as cited above). As the studies of visual reasoning below illustrate, flexible re-representation is an important capability to explain human performance. CogSketch is capable of representing the same stimulus at the level of edges, objects, or groups, as Figure 1 illustrates, and at each level it can compute a structural representation by identifying qualitative relationships between the elements. A catalog of CogSketch's visual representations is beyond the scope of this survey, see Ref. [15] for details.

Visual comparison

One of the interesting surprises in research on visual problem solving is that a general purpose model of analogy and similarity, Gentner's [16] structure-mapping theory, can account for many visual similarity judgments. In structure-mapping, same/different judgments are performed by computing *mappings* between two descriptions, placing them into alignment. A mapping consists of three things. It includes a set of *correspondences* indicating what goes with what, for example, comparing two visual arrays would involve correspondences between visual objects within the arrays, while comparing two edge-level descriptions would involve correspondences between edges. Mappings include *candidate inferences* that suggest, based on the correspondences, how the patterns compared might be completed. For example, a candidate inference might propose a new entity in an array which would make it more similar to another, or that a corner which isn't quite 90 degrees might be 90 degrees, to increase visual similarity with another edge-level description. And finally, mappings include a numerical *score*, indicating how similar the items are. Thus mappings provide valuable information for making same/different judgments: the correspondences specify how two things are similar, and the candidate inferences specify how they differ. There is evidence that these *alignable differences* are psychologically salient [3,17].

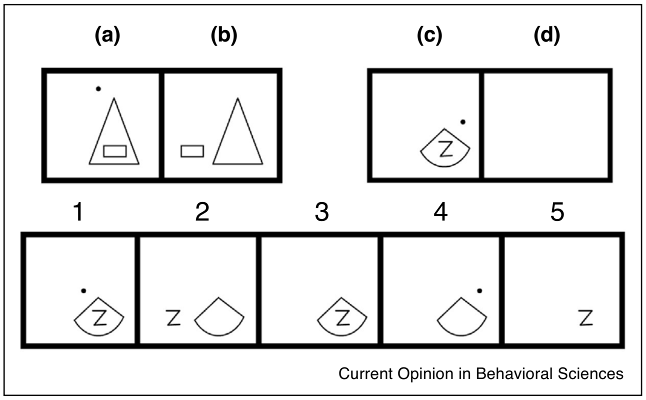
The Structure-Mapping Engine (SME; [18]) provides a computational model of the structure-mapping comparison process. SME works via a middle-out process, first constructing local hypotheses about correspondences in parallel between two symbolic representations with relational and attribute information, and then coalesces these into structurally consistent mappings. This two stage process explains an important finding in same/different judgments, namely that it is easier to determine that two things are different when they are very different, but easier to say how they are different when they are very similar [19]. When two things are very different, the number of hypothesized correspondences is very small. When two things are very similar, the candidate inferences provide alignable differences.

There are many computational models of analogy [20]. For example, LISA [21], DORA [22] and DRAMA [23] all explore neural models of versions of structure-mapping. It is unclear whether their relational capacity is sufficient to support visual reasoning. DORA, for example, only handles up to 2.5 relations at once [22], whereas existing models of geometric analogies and oddity tasks that provide human-level performance require, on average, 16 and 20 relations, respectively [18]. Some argue (e.g. Ref. [24]) that different domains require different models of analogy, but the domain-general nature of SME, which handles visual representations, linguistic representations, causal representations, problem-solving, and moral reasoning [18] suggests otherwise.

Visual reasoning tasks and their models

This section outlines three visual comparison tasks, and describes how they have been modeled. Critically, these

Figure 2



A geometric analogy problem.

models highlight the close relationship between visual comparison and the hierarchy of visual representations (edges, objects, and groups). Any comparison necessarily must be performed at a particular level in this hierarchy. If the comparison fails to reveal a good solution to a problem, then the reasoner must try again, using a different level. Thus, a key component of same/different detection in visual reasoning is finding the appropriate level of representation for the things being compared.

Geometric analogies

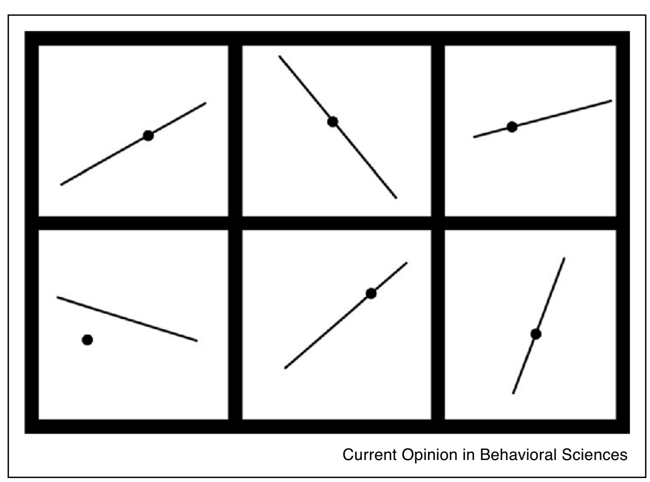
The earliest computational model of analogy was by Evans [25•], who studied geometric analogy problems of the kind that used to be on the Miller Analogy Test, a psychometric instrument. Given a pair of visual stimuli *A* and *B* and a third, *C*, the goal is to identify which of four alternatives best completes the analogy 'A is to B as C is to D' (Figure 2). Evans cast the problem as defining transformations from one figure to another, and had domain-specific algorithms for computing those transformations. The CogSketch model of geometric analogies [26,27•], by contrast, uses SME, a domain-general model of comparison. The CogSketch model has two methods for deriving answers. The first is *second-order comparison*, where A and B are compared, and C is compared to each of the four answers. The mappings themselves are then compared, via the same structure-mapping process, with the most similar 2nd order analogy selected as the answer. In solving the problem of Figure 2, for example, the first-stage comparison process includes decomposing objects into edges and comparing their edges, which enables the system to determine that the rectangle in the top left elements of Figure 2 is inside the triangle (A) versus left of the triangle (B). These differences in relationships are part of what is output from the first-stage processing. The differences between C and element 2 (e.g. inside to left, vanishing dot) are the most similar to the differences between A and B, hence that answer is chosen. The second method is *projection*, for example, it compares A to B, then compares A to C, using candidate inferences from the A:B comparison to construct what should be the correct answer. This answer is then compared against the four candidates, and the closest one is picked. Projection requires fewer comparisons but is not always possible. Attempting projection, and then reverting to second-order comparison when needed, helps explain the distribution of human response times across problems [27•], and is consistent with evidence that humans prefer projection but can adjust their strategy depending on problem difficulty [28].

Since Evans' original program, several other computational models of geometric analogy have been developed [29–31]. These models rely exclusively on a projection strategy, limiting their ability to explain human response times. In addition, none of them possess the CogSketch model's ability to flexibly move between levels of abstraction, making it difficult, for example, to compare shapes and identify spatial transformations between them, such as rotations.

Oddity task

An oddity task asks participants to select, from an array of stimuli, which one is 'different' or 'doesn't fit'. An example from Dehaene *et al.* [32] is illustrated in Figure 3, who tested participants from two cultures, majority-culture Americans and Munduruku, from the Amazon. In finding what is different, participants must find ways to construe the other elements as the same, hence same/different judgments are at the heart of the task. Lovett and Forbus' [33••] CogSketch model divides the array into two subsets, randomly. For each subset, a generalization is computed by using SME to compare the three stimuli, keeping their commonalities. Then each of the remaining

Figure 3



An oddity task problem.

stimuli is compared against that generalization, to see whether the similarity score drops for one of them more than the others. If so, that is the oddity. If no oddity is identified, then the model re-represents the stimuli at a different level in the visual hierarchy and tries again.

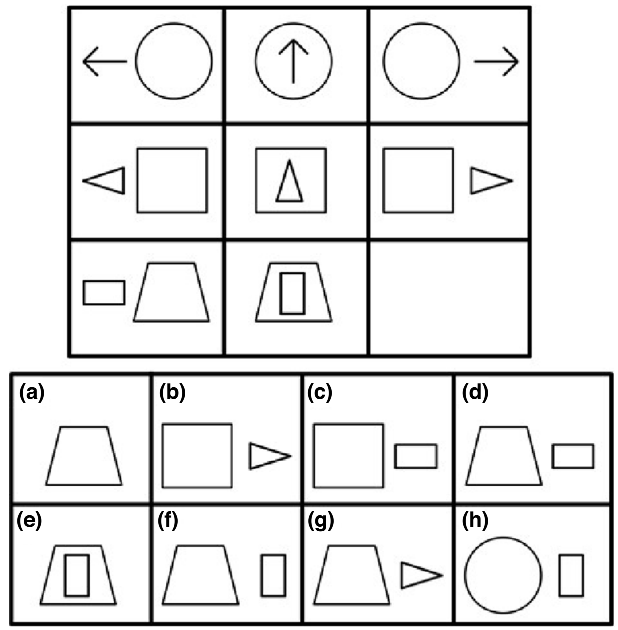
The model achieves accuracy comparable to people, and the problems that are difficult for people are also the ones the simulation fails on. Moreover, by doing ablation studies on the model, hypotheses about differences between the two cultural groups can be generated. Notably, it appears that the Americans perform well with groups of objects but struggle to reason about the edges within each object, whereas the Munduruku show the opposite pattern. The hypotheses suggested by the model can be tested by future behavioral studies.

Another class of model uses a fractal approach to solve oddity problems: each image is divided into subimages, and these subimages are exhaustively compared at the pixel level to determine the degree of commonalities in two images [34]. Then, an oddity is identified as an image that has a lower degree of commonalities with its neighbors. Although the fractal approach achieves a high degree of accuracy on the task, its error patterns look less human-like than the CogSketch model's, and it is unclear what the approach can tell us about representation and comparison in humans.

Ravens' Progressive Matrices

The Ravens' Progressive Matrices test ([35]; hereafter RPM) is one of the best predictors of human fluid intelligence. This is fascinating because it is essentially a suite of visual geometric analogy problems, of increasing difficulty (hence progressive). Figure 4 illustrates a Ravens-style problem. There is a long history of attempts to model the RPM. An early influential model was developed by Carpenter *et al.* [36], who identified several important strategies for solving such problems. However, their computational model relied on hand-generated input representations, rather than automatic encoding. In contrast, the CogSketch model [37••], automatically generates representations from PowerPoint images and solves problems using projection and second-order comparison strategies, similar to the geometric analogy model. The model achieves human-level performance on the test, and problems that are difficult for the model are also difficult for humans. Furthermore, ablation studies on the model suggest a key factor that makes RPM problems challenging: problems that require flexibly reorganizing one's visual representation, for example, by breaking objects down into their edges, comparing images at the edge level to find correspondences, and then grouping the corresponding edges to form new objects, are particularly difficult for human test-takers. Such representational challenges can only be identified by models that must generate their own visual representations.

Figure 4



Current Opinion in Behavioral Sciences

An RPM problem.

Several other RPM models rely on hand-generated representations, either for the overall images or for the shapes within each image [38,39]. McGreggor *et al.* [40] developed a fractal model that operates directly on the images, similar to their fractal model for the oddity task. However, again it is unclear what the model can tell us about representation and comparison in humans. Kunda *et al.* [41] developed a simpler model that computes affine transformations between images, but the model is unable to solve many of the more difficult RPM problems.

More recently, several deep learning models have proven effective at solving RPM problems (e.g. Refs. [42,43]). Typically, such models can be difficult to interpret because they lack explicit knowledge structures or processes that align with psychological theories—they simply learn a mapping from RPM problem to solution. Interestingly, one recent model was able to solve problems after training on a set of general images, rather than training directly on RPM examples [44]. However, it was primarily able to solve easier problems that involve filling in the missing piece of a pattern.

Summary

There are multiple sources of evidence suggesting that qualitative representations, computed over entities organized into a visual hierarchy, are used in visual reasoning. Structure-mapping over such representations provides a modeling framework that has been shown to handle a variety of visual reasoning tasks, achieving human-level performance with predictions of other properties of human behavior. Moreover, the structure-mapping notions of same/different apply in visual reasoning, as well as other areas of cognition.

A clear lesson from visual problem-solving is that finding the appropriate level of representation is crucial to making same/different determinations. We expect that this is true for same/different judgments more broadly, which suggests that one path to a better understanding of such judgments might be computational modeling of encoding and re-representation processes more broadly.

Conflict of interest statement

Nothing declared.

Acknowledgements

This research was supported by the NSF-funded Spatial Intelligence Learning Center and the Office of Naval Research.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Biederman I: **Recognition-by-components: a theory of human image understanding**. *Psychol Rev* 1987, **94**:115-147.

2. Kosslyn SM, Koenig O, Barrett A, Cave CB, Tang J, Gabrieli JDE: **Evidence for two types of spatial representations: hemispheric specialization for categorical and coordinate relations**. *J Exp Psychol Hum Percept Perform* 1989, **15**:723-735.

3. Markman AB, Gentner D: **Commonalities and differences in similarity comparisons**. *Mem Cogn* 1996, **24**:235-249.

4. •• Forbus K: *Qualitative Representations: How People Reason and Learn about the Continuous World*. MIT Press; 2019.
Provides an overview of qualitative representations from a cognitive science perspective, including a survey of qualitative spatial representations.

5. Palmer S: **Hierarchical structure in perceptual representation**. *Cogn Psychol* 1977, **9**:441-474.

6. Marr D: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman; 1982.

7. • Lowet A, Firestone C, Scholl B: **Seeing structure: shape skeletons modulate perceived similarity**. *Attent Percept Psychophys* 2018, **80**:1278-1289.
This paper provides evidence that medial axis transforms are useful in modeling human shape perception.

8. Firestone C, Scholl B: **"Please Tap the Shape, Anywhere You Like": shape skeletons in human vision revealed by an exceedingly simple measure**. *Psychol Sci* 2014, **25**:377-386.

9. Lovett A, Franconeri S: **Topological relations between objects are categorically coded**. *Psychol Sci* 2017:1-11 http://dx.doi.org/10.1177/0956797617709814.

10. Palmer S, Brooks L, Nelson R: **When does grouping happen?** *Acta Psychol* 2003, **114**:311-330.

11. Hochstein S, Ahissar M: **View from the top: hierarchies and reverse hierarchies in the visual system**. *Neuron* 2002, **36**:791-804.

12. Marr D, Nishihara HK: **Representation and recognition of the spatial organization of three-dimensional shapes**. *Philos Trans R Soc Lond Ser B Biol Sci* 1978, **200**:269-294.

13. Forbus K, Usher J, Lovett A, Lockwood K, Wetzel J: **CogSketch: sketch understanding for cognitive science research and for education**. *Top Cogn Sci* 2011, **3**:648-666.

14. Chen K, Rabkina I, McLure M, Forbus K: *Human-like Sketch Object Recognition via Analogical Learning*. AAAI; 2019. 2019.

15. Forbus KD, Chang M, McLure M, Usher M: **The cognitive science of sketch worksheets**. *Top Cogn Sci* 2017, **9**:921-942 http://dx.doi.org/10.1111/tops.12262.

16. Gentner D: **Structure-mapping: a theoretical framework for analogy**. *Cogn Sci* 1983, **7**:155-170.

17. Gentner D, Gunn V: **Structural alignment facilitates the noticing of differences**. *Mem Cogn* 2001, **29**:565-577.

18. Forbus K, Ferguson R, Lovett A, Gentner D: **Extending SME to handle large-scale cognitive modeling**. *Cogn Sci* 2017, **41**:1152-1201 http://dx.doi.org/10.1111/cogs.12377.

19. Sagi E, Gentner D, Lovett A: **What difference reveals about similarity**. *Cogn Sci* 2012, **36**:1019-1050.

20. Gentner D, Forbus K: **Computational models of analogy**. *WIREs Cogn Sci* 2011, **2**:266-276.

21. Hummel JE, Holyoak KJ: **Distributed representations of structure: a theory of analogical access and mapping**. *Psychol Rev* 1997, **104**:427-466.

22. Doumas LAA, Hummel JE, Sandhofer CM: **A theory of the discovery and predication of relational concepts**. *Psychol Rev* 2008, **115**:1-43.

23. Eliasmith C, Thagard P: **Integrating structure and meaning: a distributed model of analogical mapping**. *Cogn Sci* 2001, **25**:245-286.

24. Holyoak K, Lee H, Lu H: **Analogy and category-based inference: a theoretical integration with Bayesian causal models**. *J Exp Psychol Gen* 2010, **139**:702-727.

25. • Evans T: **A program for the solution of a class of geometric-analogy intelligence-test questions**. In *Semantic Information Processing.* Edited by Minsky M. Cambridge, MA: MIT Press; 1968.
This paper was the first computational model of analogy, and included, unlike many later models, automatic computation of input representations, a landmark for its time.

26. Lovett A, Tomai E, Forbus K, Usher J: **Solving geometric analogy problems through two-stage analogical mapping**. *Cogn Sci* 2009, **33**:1192-1231.

27. • Lovett A, Forbus K: **Modeling multiple strategies for solving geometric analogy problems**. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*; *Sapporo, Japan: 2012*.
Describes the CogSketch model for geometric analogies that uses both visual projection and second order analogies, more closely matching human performance.

28. Bethell-Fox C, Lohman D, Snow R: **Adaptive reasoning: componential and eye movement analysis of geometric analogy performance**. *Intelligence* 1984, **8**:205-238.

29. Schwering A, Gust H, Kühnberger K, Krumnack U: **Solving geometric proportional analogies with the analogy model HDTP**. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* 2009.

30. O'Donoghue DP, Bohan A, Keane MT: **Seeing things - inventive reasoning with geometric analogies and topographic maps**. *New Gener Comput* 2006, **24**:267-288.

31. Ragni M, Schleipen S, Steffenhagen F: **Solving proportional analogies: a computational model**. *Workshop on Analogies: Integrating Multiple Cognitive Abilities, CogSci' 07* 2007.

32. Dehaene S, Izard V, Pica P, Spelke E: **Core knowledge of geometry in an Amazonian indigene group**. *Science* 2006, **311**:381-384.

33. Lovett A, Forbus K: **Cultural commonalities and differences in
•• spatial problem-solving: a computational analysis**. *Cognition* 2011, **121**:281-287.
Describes the CogSketch model of oddity tasks.

34. McGreggor K, Goel A: **Fractal representations and core geometry**. *Proceedings of the Second Annual Conference on Advances in Cognitive Systems* 2013.

35. Raven J, Raven JC, Court JH: *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 1. General Overview*. Oxford: Oxford Psychologists Press; 1998.

36. Carpenter PA, Just MA, Shell P: **What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices test**. *Psychol Rev* 1990, **97**:404-431.

37. Lovett A, Forbus K: **Modeling visual problem-solving as
•• analogical reasoning**. *Psychol Rev* 2017, **124**:60-90.
Describes the CogSketch model of Ravens' Progressive Matricies.

38. Rasmussen D, Eliasmith C: **A neural model of rule generation in inductive reasoning**. *Top Cogn Sci* 2011, **3**:140-153.

39. Cirillo S, Ström V: *An Anthropomorphic Solver for Raven's Progressive Matrices (No. 2010:096)*. Goteborg, Sweden: Chalmers University of Technology; 2010.

40. McGreggor K, Kunda M, Goel AK: **Fractals and Ravens**. *Artif Intell* 2014, **215**:1-23.

41. Kunda M, McGreggor K, Goel AK: **A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations**. *Cogn Syst Res* 2013, **22-23**:47-66.

42. Barrett DG, Hill F, Santoro A, Morcos AS, Lillicrap T: **Measuring abstract reasoning in neural networks**. *Proceedings of the 35th International Meeting on Machine Learning* 2018.

43. Zhang C, Gao F, Jia B, Zhu Y, Zhu S-C: **Raven: a dataset for relational and analogical visual reasoning**. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2019.

44. Hua T, Kunda M: *Modeling Gestalt Reasoning on Raven's Progressive Matrices Intelligence Test Using Generative Image Inpainting Techniques*. . Retrieved from the Arxiv database 2019.