



AI, visual imagery, and a case study on the challenges posed by human intelligence tests

Maithilee Kunda^{a,1}

^aElectrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235-1679

Edited by Richard M. Shiffrin, Indiana University Bloomington, Bloomington, IN, and approved August 19, 2020 (received for review December 16, 2019)

Observations abound about the power of visual imagery in human intelligence, from how Nobel prize-winning physicists make their discoveries to how children understand bedtime stories. These observations raise an important question for cognitive science, which is, what are the computations taking place in someone's mind when they use visual imagery? Answering this question is not easy and will require much continued research across the multiple disciplines of cognitive science. Here, we focus on a related and more circumscribed question from the perspective of artificial intelligence (AI): If you have an intelligent agent that uses visual imagery-based knowledge representations and reasoning operations, then what kinds of problem solving might be possible, and how would such problem solving work? We highlight recent progress in AI toward answering these questions in the domain of visuospatial reasoning, looking at a case study of how imagery-based artificial agents can solve visuospatial intelligence tests. In particular, we first examine several variations of imagery-based knowledge representations and problem-solving strategies that are sufficient for solving problems from the Raven's Progressive Matrices intelligence test. We then look at how artificial agents, instead of being designed manually by AI researchers, might learn portions of their own knowledge and reasoning procedures from experience, including learning visuospatial domain knowledge, learning and generalizing problem-solving strategies, and learning the actual definition of the task in the first place.

artificial intelligence | computational modeling | mental imagery | Raven's Progressive Matrices | visuospatial reasoning

I think in pictures. Words are like a second language to me. I translate both spoken and written words into full-color movies, complete with sound, which run like a VCR tape in my head. . . . Language-based thinkers often find this phenomenon difficult to understand, but in my job as an equipment designer for the livestock industry, visual thinking is a tremendous advantage.

Temple Grandin, professor of animal science and autism advocate (ref. 1, p. 3)

What I am really trying to do is bring birth to clarity, which is really a . . . thought-out pictorial semivision thing. I would see the jiggle-jiggle-jiggle or the wiggle of the path. Even now when I talk about the influence functional, I see the coupling and I take this turn—like as if there was a big bag of stuff—and try to collect it away and to push it. It's all visual. It's hard to explain.

Richard Feynman, Nobel laureate in physics (ref. 2, p. 244)*

Temple Grandin is a well-known animal scientist who is on the autism spectrum. She has had incredible professional success in the livestock industry, and she credits her success to her strong visual imagery skills, that is, abilities to generate, transform, combine, and inspect visual mental representations. (1).

Many physicists such as Richard Feynman (2), Albert Einstein (3), and James Clerk Maxwell (4) used imagery in their creative discovery processes, and similar patterns emerge in accounts by and about mathematicians (5), engineers (6), computer program-

mers (7), product designers (8), surgeons (9), memory champions (10), and more. People also use visual imagery in everyday activities such as language comprehension (11), story understanding (12), and physical (13) and mathematical reasoning (14).

These observations raise an interesting scientific question: What are the computations taking place in someone's mind when they use visual imagery? This is a difficult question that continues to receive attention across cognitive science disciplines (15).

Here, we focus on a related, more circumscribed question from the perspective of artificial intelligence (AI): If you have an intelligent agent that uses visual imagery-based knowledge representations and reasoning operations, then what kinds of problem solving might be possible, and how would it all work?

In this paper, we discuss progress in AI toward answering this question in the domain of visuospatial reasoning—reasoning about the geometric and spatial properties of visual objects (16). This discussion necessarily leaves out such intriguing and important complexities as nonvisual forms of spatial reasoning, for example, in people with visual impairments (17); the role of physics and forces in imagery (18); imagery in other sensory modalities (19); etc.

As a case study, we focus on visuospatial reasoning for solving human intelligence tests like Raven's Progressive Matrices. While many AI techniques have been developed to solve many different tests (20), we are still quite far from having an artificial agent that can “sit down and take” an intelligence test without specialized algorithms having been designed for that purpose. Contributions of this paper include discussions of 1) why intelligence tests are such a good challenge for AI; 2) a framework for artificial problem-solving agents with four components: a problem definition, input processing, domain knowledge, and a problem-solving strategy or procedure; 3) several imagery-based agents that solve Raven's problems; and 4) how an imagery-based agent could learn its domain knowledge, problem-solving strategies, and problem definition/input processing components, instead of each being manually designed.

Why the Raven's Test Is (Still!) a Hard AI Challenge

Take a look at the problems in Fig. 1. Can you solve them?

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Brain Produces Mind by Modeling,” held May 1–3, 2019, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. NAS colloquia began in 1991 and have been published in PNAS since 1995. From February 2001 through May 2019, colloquia were supported by a generous gift from The Dame Jillian and Dr. Arthur M. Sackler Foundation for the Arts, Sciences, & Humanities, in memory of Dame Sackler's husband, Arthur M. Sackler. The complete program and video recordings of most presentations are available on the NAS website at <http://www.nasonline.org/brain-produces-mind-by>.

Author contributions: M.K. wrote the paper.

The author declares no competing interest.

This article is a PNAS Direct Submission.

Published under the [PNAS license](#).

¹ Email: mkunda@vanderbilt.edu.

First published November 23, 2020.

*Feynman's quote includes a mild profanity that has been omitted due to PNAS editorial policy. The full quote can be found in many places online.

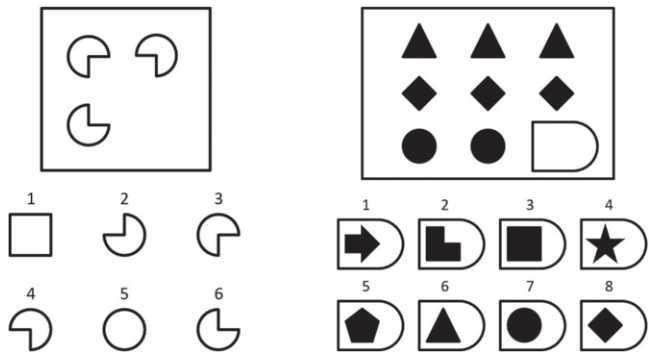


Fig. 1. Sample problems like those from the Raven's intelligence test, comparable to ones of easy-to-middling difficulty on the standard version of the test.

While these problems may seem straightforward, consider for a moment the complexity of what you just did. As you were solving each problem, some executive control system in your mind was planning and executing a series of physical and cognitive operations, including shifts of gaze from one element of the problem to another, storing extracted features in working memory, computing and storing the results of intermediate calculations, and so on. And, you did all of this without any explicit instructions as to what cognitive operations to use, or in what order to apply them.

At a deeper level, you may notice that no one actually even told you what these problems were about. Typically, Raven's test-takers are instructed to solve each problem by selecting the answer from the bottom that best completes the matrix portion on top (21). However, even if you hadn't seen problems quite like these before, it is likely that you were able to grok the point of the problems just by looking at them, no doubt due to a lifetime of experience with pattern-matching games and multiple choice tests.

From a general AI perspective, intelligence tests like the Raven's have been "solved" in the sense that we do have computational programs that, given a Raven's problem as input, can often produce the correct answer as an output. In fact, some of the earliest work in AI was Evans' classic ANALOGY program from the 1960s—at the time, the largest program written in LISP to date!—that solved geometric analogy problems from college aptitude tests (22).

However, all of these programs have essentially been hand-crafted to solve Raven's problems in one way or another. Humans (at least in theory) are supposed to take intelligence tests without having practiced them beforehand. Thus, intelligence tests like the Raven's are still an "unsolved" challenge for AI when treated as tests of generalization, that is, generalizing previously learned knowledge and skills to solve new and unfamiliar types of problems.

At an even higher level, the notion of "taking a test" is itself a sophisticated social and cultural construct. In people, for example, crucial research on stereotype threat has observed how stereotypes about race and gender can influence a person's performance on the exact same test depending on whether they are told it is a "test" or a "puzzle" (23). If we assume that human cognition can be explained in computational terms, then, someday, we ought to be able to have AI agents that model these effects.*

*Perhaps ironically, early AI research studied what we thought were the hard problems, like taking tests and playing chess. The next wave of research recognized that the real hard problems were, in fact, the ones that were easy for many people, like walking around or recognizing cats (24). Now, we are realizing that the original hard problems of taking tests and playing chess are quite hard after all—but only if you really consider the full work of the agent, which includes figuring out what to do and understanding why you are doing this thing in the first place. In other words, many animals can walk around and pick up rocks, but only humans play good chess and take difficult tests.

The Raven's test and similar tests of matrix reasoning and geometric analogy are particularly interesting for AI for several reasons. First, the Raven's test, originally designed to measure "eductive ability," or the ability to extract and understand information from a complex situation (21), occupies a unique niche among psychometric instruments as being the best single-format measure of a person's general intelligence (25). In other words, the Raven's test seems to tap into fundamental cognitive abilities that are very relevant to many other things a person tries to do.

Second, there are several Raven's tests that span a very wide range of difficulty levels, from problems that are easy for young children to problems that are difficult for most adults. The developmental trajectories of performance that people show offer a motivating parallel for studying AI agents that meaningfully improve their problem-solving abilities through various learning experiences.

Third, there is evidence that many people use multiple forms of mental representation while solving Raven's problems, including inner language as well as visual imagery (26, 27). Interestingly, many people on the autism spectrum show patterns of performance on the Raven's test that do not match patterns seen in neurotypical individuals (28), and neuroimaging findings suggest that many individuals on the spectrum rely more on visual brain regions than neurotypicals do while solving the test (29). Thus, the Raven's test is a fascinating testbed for AI research on visual imagery in particular and multimodal reasoning more generally.

A Framework for Artificial Agents That Solve Problems

Many approaches in AI can usefully be decomposed according to the framework shown in Fig. 2. The agent is given a problem as input and is expected to produce a correct solution as output.

The "problem definition" refers to the agent's understanding of what the problem is actually asking, that is, what constitutes a valid format of inputs and outputs ("problem template") and what the goal is in terms of desired outputs ("solution criteria"). For example, for a generic Raven's problem, the problem template might specify a two-dimensional matrix M of images m_i , with one entry in the matrix missing, and an unordered set A of answer images a_i , and that a valid answer consists of selecting one (and only one) answer $a_i \in A$. The solution criterion is that the selected answer should be the one that "best fits" in the missing slot in M .

The "input processing" component refers to how an agent takes raw or unstructured inputs from the "world" and converts them into a usable internal problem representation. For example, what the Raven's test actually provides is a pattern of ink on paper. At some point, this visual image needs to be decomposed into the matrix M and answer choice A elements in the problem template. For many artificial agents, input processing is performed outside the agent, either manually or by some other system. For example, most chess-playing agents do not operate using a video feed of a chess board, but rather using an explicit specification of where all of the pieces are on the board. While

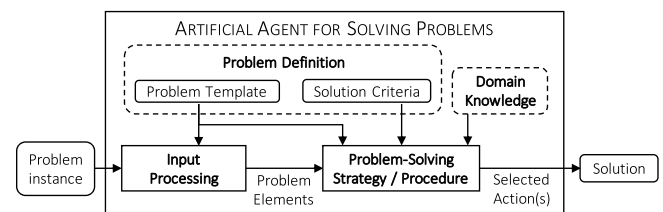


Fig. 2. Framework for artificial agents. Pushing the boundaries of what artificial agents can do often involves deriving more and more of the internal structure and knowledge of the agent through learning instead of programming.

this is a reasonable assumption to make in many AI applications, it does mean that the agent relies on having a simplified and preprocessed set of inputs.

“Domain knowledge” refers to whatever knowledge an agent needs to solve the given type of problems. The Raven’s test can be tackled using visuospatial knowledge about symmetry, sequential geometric patterns, rows and columns, etc.

Finally, the “problem-solving strategy” encompasses what the agent actually does to solve a given problem, that is, the algorithm that churns over the problem definition, domain knowledge, and specific problem inputs in order to generate an answer.

Given this framework, what would it mean for an agent to use visual imagery to solve problems? We offer one formulation: Anywhere beyond the input processing step, the agent needs to use or retain representations of problem information that count as “images” in some way. This includes image-like representations occurring in the problem definition, domain knowledge, problem-solving strategy, and/or the specific problem representations generated by the input processing component.

What counts as an image-like representation? Previous research on computational imagery often distinguishes between spatial representations, that is, those that replicate the spatial structure of what is being represented, versus visual/object representations, that is, those that replicate the visual appearance of what is being represented (30). These categories correspond to findings about spatial versus object imagery in people (31). Thus, we label agents using either type of representation as using visual imagery or being imagery based. The imagery-based Raven’s agents discussed later in this paper primarily use visual/object imagery and not spatial imagery, although, certainly, many other AI research efforts have developed agents that use spatial imagery (32).

Note that imagery here refers to the format in which something is represented, not the contents of what is represented. Many artificial agents reason about visuospatial information using nonimagery-based representations (33); for example, visuospatial domain knowledge can be encoded propositionally, such as the rule left-of (x,y) \implies right-of (y,x).

Different Types of Raven’s Problem-Solving Agents

Different paradigms of AI agents can now be described according to components in this framework.

Knowledge-based approaches, also associated with terms like cognitive systems (34) or symbolic AI, traditionally rely on manually designed domain knowledge and flexible problem-solving procedures like planning and search to tackle complex problems. The first wave of “propositional Raven’s agents” used manual or automated input processing to convert raw test problem images into amodal, propositional representations, such as lists of attribute–value pairs, and then problem-solving procedures would operate over these propositional representations (33, 35–37). Visuospatial domain knowledge in these agents included predefined types of relationships among elements, like similarity or containment, and methods for extracting and defining relationships.

As foreshadowed in early writings about possible representational and algorithmic strategy differences on the Raven’s test (38), a second wave of “imagery-based Raven’s agents” were also knowledge-based, but their internal representations of problem information remained visual, that is, the problem-solving procedures directly accessed and manipulated problem images, and even often created new images during the course of reasoning (39–43). Visuospatial domain knowledge in these agents included image functions like rotation, image composition, visual similarity, etc.

More recently, a wave of “data-driven Raven’s agents” aims to learn integrated representations of visuospatial domain knowl-

edge and problem-solving strategies by training on input–output pairs from a large number of example problems (44–49).

Which approach is correct? This is a bad question, as different types of agents are used for very different lines of scientific inquiry. Referring again to Fig. 2, most knowledge-based Raven’s agents are used to study problem-solving procedures and assume a relatively fixed set of domain knowledge (although some of these agents certainly include forms of learning as well). Most of the data-driven Raven’s agents are used to study how domain knowledge about visuospatial relationships can be learned from examples, and the problem-solving procedure is often (although not always) fixed.

All of these Raven’s agents have many hand-built components, although the parts that are hand-built differ from one agent to another. Many open AI challenges remain, even within the one task domain of the Raven’s test, to gradually convert the components in Fig. 2 from being manually programmed to being learned or developed by the agents themselves. Next, we discuss how knowledge-based agents can use imagery to solve Raven’s problems in several different ways, and then we examine emerging methods for agents to learn their own 1) domain knowledge, 2) problem-solving strategies, and, finally, 3) problem definitions.

Imagery-Based Strategies for Solving Raven’s Problems

Within the category of imagery-based Raven’s agents, many different formulations are possible, in terms of the problem-solving strategy that is used, the representation and contents of domain knowledge, and even the problem definition.

We describe five imagery-based strategies along with results from research by the author and colleagues. Results are reported for the Raven’s Standard Progressive Matrices test, scored out of 60 problems (21). For comparison, human norm data suggest that average children in the United States would score around 26/60 as 8-y-olds, 40/60 as 12-y-olds, and 49/60 as 16-y-olds.

At a high level, the following strategies are described in terms of two strategy types observed in psychology research (50): In “constructive matching,” the test-taker looks at the problem matrix, generates a guess for the missing element, and then chooses an answer most similar to its generated guess. In “response elimination,” the test-taker looks at each answer in turn, plugging it into the problem matrix, and choosing the one that produces the best overall matrix.

Strategy 1 (Fig. 3A). We developed an imagery-based agent that solves Raven’s problems through multistep search, using a constructive matching strategy (39, 43, 51): 1) Using elements from complete rows/columns of the matrix, search among known visual transformations for the one that best explains image variation across parallel rows/columns. 2) Apply this transformation to elements in a partial row or column to predict a new answer image. 3) Search among the answer choices to find the one that is most similar to the predicted answer image.

More formally, problem inputs include a set M of images m_i representing sections of the problem matrix, and a set A of answer choice images a_i . Let C be the set of all collinear subsets c of M , with c_x referring to the first element(s), and c_y referring to the last element. Each c contains matrix elements along rows, columns, or diagonals. We define an analogy g as a pairing of a single complete collinear subset c_1 with an incomplete collinear subset c_2 (i.e., $g = [c_{1,x} : c_{1,y} :: c_{2,x} : c_{2,y}]$, where $c_{2,y}$ is the missing element in the matrix). All such analogies that share the same c_2 are further aggregated into sets $G_i \in G$.

In addition, let T be the agent’s predefined set of visual transformations. Also, let $\text{sim}(I_1, I_2)$ be a function that returns a

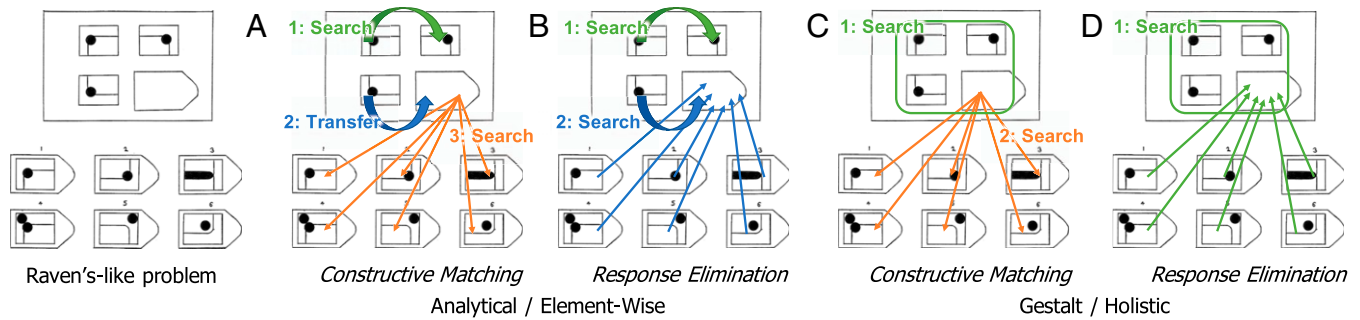


Fig. 3. Raven's-like problem and four different imagery-based strategies for solving it. A problem consists of matrix M of elements m_i and set A of answer choices a_j . (A) First strategy begins with search for transformation t that best transforms m_1 into m_2 , then applies t to m_3 to produce an image candidate for m_4 , and finally searches for answer a_i most similar to m_4 . (B) Second strategy also begins with search for t that best transforms m_1 into m_2 , then conducts similar searches for transformations t_{3i} that transform m_3 into each a_{3i} , and finally searches for answer a_i that yields t_{3i} most similar to t . (C) Third strategy begins with search for image m_4 that maximizes Gestalt metric for matrix M , and then searches for answer a_i most similar to m_4 . (D) Fourth strategy involves search for answer a_i that maximizes Gestalt metric for matrix M .

real-valued measure of similarity between images I_1 and I_2 . First, the agent finds the best-fit transformation,

$$(t_{max}, g_{max}) = \operatorname{argmax}_{t \in T, G_i \in G} \left(\operatorname{mean}_{g \in G_i} (\operatorname{sim}(t(g \cdot c_{1,x}), g \cdot c_{1,y})) \right),$$

Second, the agent computes a predicted answer image as $a_{pred} = t_{max}(g_{max} \cdot c_{2,x})$. Third, the agent returns the most similar answer choice: $a_{final} = \operatorname{argmax}_{a_i \in A} (\operatorname{sim}(a_{pred}, a_i))$. Hand-coded domain knowledge is provided in the form of the set T of visual transformations, including eight rectilinear rotations and reflections (including identity) and three to six image composition operations (union, intersection, subtraction, and combinations of these) as well as visual similarity and other image processing utility functions. Steps 1 and 3 above used exhaustive search.

Successive versions of the agent, using more transformations T and more varied ways to optimize over matrix entries in step 1, have achieved scores of 38/60 (39), 50/60 (51), and 57/60 (43) on the Raven's Standard Progressive Matrices test.

Strategy 2 (Fig. 3B). In a related line of research, colleagues developed a different imagery-based agent that adopted a response elimination type of strategy (Fig. 3B). In this work (40), a smaller set of visual transformations (rotation and reflection) was used to compute "fractal image transformations," that is, a representation of one image in terms of another, using techniques from image compression (52).

In particular, to compute a fractal transformation between source image A and target image B , B is first partitioned into a set of subimages b_i . Then, for each b_i , a fragment $a_i \in A$ is found such that b_i can be expressed as an affine transformation t_i of a_i . The fragments a_i are twice the size of b_i , resulting in a contractive transformations. The set T of all t_i is the fractal transformation of A into B .

To solve a Raven's problem, a fractal transformation T is computed using elements from each complete row/column j in the matrix, and then similar transformations T'_{ij} are computed for each of the answer choices plugged into the incomplete rows/columns of the matrix. Finally, the selected answer is the one yielding the fractal transformations most similar to those computed for the original rows/columns of the matrix. Formally, if we let T_{sim} be a similarity metric across fractal transformations, the final answer is given by

$$a_{final} = \operatorname{argmax}_{a_i \in A} \sqrt{\sum_j T_{sim}(T_j, T'_{ij})^2}.$$

Results using this fractal method were also 50 out of 60 correct on the Raven's Standard Progressive Matrices test, allowing for some ambiguous detections of the answers, or 38 out of 60 correct with a specific method for resolving these ambiguities (40).

Strategy 3 (Fig. 3C). The first two strategies consider each matrix element individually. However, people can also use a "Gestalt" strategy to consider the entire matrix as a whole (38, 53). For instance, for the problem in Fig. 3, if one looks at the matrix as a single image, an answer might just "appear" in the blank.

In recent work (42), we attempted to model this kind of strategy using neural networks for image inpainting, trained to fill in the missing portions of real photographs. We used a recently published image inpainting network consisting of a variational autoencoder combined with a generative adversarial network (54), and we tested several versions of the network trained on different types of photographs, such as objects, faces, scenes, and textures. Given an image of the incomplete problem matrix, the network outputs a guess for what image should fill in the missing portion. This guess is then used to select the most similar answer.

Formally, let F be the learned encoder network that converts an image into a representation in a learned feature space, and let G be the learned decoder network that converts a feature-based image back into pixel space, including inpainting to fill in any missing portions. Then, our agent first computes $M' = G(F(M))$ to obtain a new, filled-in matrix image, with m_x denoting the new, filled-in portion of M' . Let $L2dist$ represents the L2 norm of a vector in the learned feature space. Then, the final answer is

$$a_{final} = \operatorname{argmin}_{a_i \in A} (L2dist(F(m_x) - F(a_i))).$$

Fig. 4 shows examples of inpainting results on several example problems, some of which are filled in more effectively than others. The best version of this agent, trained on photographs of objects, answered 25 out of 60 problems on the Raven's Standard Progressive Matrices test. While this score may seem low, it is quite astonishing given that there was no Raven's-specific information fed into or contained in the inpainting network, and, in fact, the network had never before "seen" line drawings, only photographs.

Strategy 4 (Fig. 3D). The fourth strategy combines a Gestalt approach with response elimination. We have not yet implemented this strategy, nor do we know of other AI efforts that have, but we present a brief sketch here. Essentially, this strategy

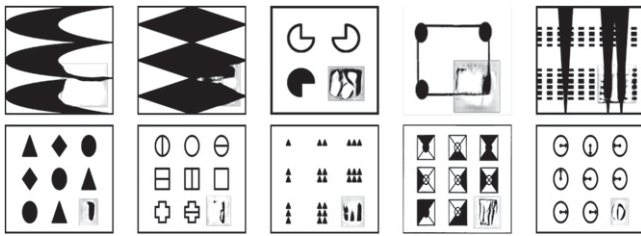


Fig. 4. Images generated using an inpainting neural network (54) for Raven's-like problems (42). The network was trained only on real-world photographs of objects.

works by plugging in answers to the matrix, and choosing the one that creates the “best” overall picture, for some notion of best.

Assume a Gestalt metric S that measures the Gestalt quality of any given image. Images that are highly symmetric, contain coherent objects, etc., would score highly, and images that are chaotic or broken up would score poorly. Then, the agent chooses the answer that scores highest when plugged into the matrix M ,

$$a_{final} = \operatorname{argmax}_{a_i \in A} (S(M \cup a_i)).$$

Strategy 5 (Not Shown in Figure). The above four strategies treat Raven's matrix elements as single images. However, previous computational and human studies have suggested that it can be helpful to decompose Raven's problems into multiple subproblems, by breaking up a single matrix element into subcomponents (35).

In previous work, we have also explored imagery-based techniques for decomposing a geometric analogy into subproblems, solving each separately, and then reassembling the subsolutions back together to choose the final answer (55), although this method has not yet been tested on the actual Raven's tests.

Open Questions. From this small survey, it is clear that there is no single imagery-based Raven's strategy. Imagery-based agents are like logic-based agents or neural network-based agents; there are a set of generally shared principles of representation and reasoning, but then individual agents are designed to use specific instantiations of these and combine them in different ways to produce very diverse problem-solving behaviors.

Exploring the space of imagery-based agents is valuable, not to find the “best” one but rather to characterize the space itself. Each agent, as a data point in this space of possible agents, is an artifact that can be studied in order to understand something about how that particular set of representations and strategies can produce intelligent task behaviors (56). Future work should continue to add data points to this space and also investigate the extent to which these strategies overlap with human problem solving.

Learning Visuospatial Domain Knowledge

Imagery-based agents use many kinds of visuospatial domain knowledge, including visual transformations like rotation, scaling, and composition; hierarchical representations of concepts in terms of attributes like shape and texture; Gestalt principles like symmetry, continuity, and similarity; etc. These types of knowledge can be leveraged by an agent to solve problems from the Raven's test as well as many other visuospatial tests (32).

Visuospatial domain knowledge also includes more semantically rich information such as what kinds of objects go where in a scene (57); we do not further discuss this type of semantic knowledge here, although it certainly plays an important role in

imagery-based AI, especially for agents that perform language understanding or commonsense reasoning tasks (32).

How is visuospatial domain knowledge learned? One hypothesis suggests that agents learn such knowledge through prior sensorimotor interactions with the world. Under this view, the precise nature of the representations and learning mechanisms involved are important open questions. For brevity, we discuss here AI research on learning two types of visuospatial domain knowledge—visual transformations and Gestalt principles.

Learning Visual Transformations. In humans, many reasoning operators used during visual imagery (e.g., transformations like mental rotation, scaling, etc.) are hypothesized to be learned from visuomotor experience, for example, perceiving the movement of physical objects in the real world (58). As with the well-known kittens-in-carousel experiments (59), learning visual transformations may rely on the combination of active motor actions coupled with visual perception of the results of those actions. Studies in both children and adults have indeed found that training on a manual rotation task does improve performance on mental rotation (60, 61).

Computational efforts to model the learning of visual transformations have generally represented each transformation as a set of weights in a neural network. In early work, distinct networks were used to learn each transformation individually (62). More recent work combines the visual and motor components of inputs for learning mental rotation (63). While many of these approaches implement visual transformations as distinct operations, a more general approach might represent continuous visual operations as combinations of basis functions that can be combined in arbitrary ways (64). Along these lines, other recent work uses more complex neural networks to represent transformations as combinations of multiple learned factors, although this work still focused on relatively simple transformations like rotation and scaling (65, 66).

People certainly do not learn visual transformations from specialized training on rotation, scaling, etc., taken as separate transformations. More generally, we have access to a very robust and diverse machinery for simulating visual change, and the simple “mental rotation” types of tasks often used in studies of visual imagery tap into only very tiny slices of this knowledge base. In line with evidence of the importance of motor actions and forces on our own imagery abilities (18), we expect that work in AI to model physical transformations—especially work in robotics that combines visual and motor inputs/outputs—will be essential for producing the kinds of capabilities agents need for visual imagery.

There is starting to be a wave of relevant work in AI in the area of “video prediction,” which involves learning representations of the appearance of objects as well as their dynamics (67–69), including for increasingly complex forms of dynamics, as with a robot trying to manipulate a rope (70). Importantly, these efforts focus on learning and making inferences about object dynamics directly in the image space, as opposed to computational approaches that rely on explicit physics simulations and then project predictions into image space. Thus, these new approaches offer intriguing possibilities as potential models for how humans might learn naive physics as a form of imagery-based reasoning.

Learning Gestalt Principles. Many visuospatial intelligence tests rely on a person's knowledge of visual relationships like similarity, continuity, symmetry, etc. Simple tests like shape matching require the test-taker to infer first-order relationships among visual elements, while more complex tests like the Raven's often progress into second-order relationships, that is, relations over relations.

In one sense, a test like the Raven's ought to be agnostic with respect to the specific choice of first-order relationships, and, indeed, in many propositional AI agents, a relation like $\text{CONTAINS}(X, Y)$ can be replaced with any arbitrary label, and the results will stay the same. However, for people, the actual visuospatial relationships at play do deeply influence our problem-solving capabilities. For example, isomorphs of the Tower of Hanoi task are more difficult if task rules are less well aligned with our real-world knowledge about spatial structure and stacking (71). Similarly, the perceptual properties of Raven's problems have been found to be a strong predictor of item difficulty (72).

A person's prior knowledge about visuospatial relationships is closely tied to Gestalt perceptual phenomena. In humans, Gestalt phenomena have to do, in part, with how we integrate low-level perceptual elements into coherent, higher-level wholes (73), as shown in Fig. 5. Psychology research has enumerated a list of principles (or laws, perceptual/reasoning processes, etc.) that seem to operate in human perception, like preferences for closure, symmetry, etc. (74). Likewise, work in image processing and computer vision has attempted to define these principles mathematically or computationally, for instance, as a set of rules (75).

However, in more recent computational models, Gestalt principles are seen as emergent properties that reflect, rather than determine, perceptions of structure in an agent's visual environment. For example, early approaches to image inpainting—that is, reconstructing a missing/degraded part of an image—used rule-like principles to determine the structure of missing content, while later approaches use machine learning to capture structural regularities from data and apply them to new images (76). This seems reasonable as a model of Gestalt phenomena in human cognition; it is because of our years of experience with the world around us that we see Fig. 5, *Left* as partially occluded/degraded views of whole objects.

Image inpainting represents a fascinating area of imagery-based abilities for artificial agents (54), which we used in our model of Gestalt-type problem solving on the Raven's test (42), as described earlier. Other work in computer vision and machine learning studies the extent to which neural networks not explicitly designed to model Gestalt effects might exhibit such effects as emergent phenomena (77–81).

Learning a Problem-Solving Strategy

Relatively little research in AI has proposed methods for automatically generating problem-solving procedures for intelligence tests, despite the extensive research on manually constructed solution methods or methods that rely on a large number of examples (20). How does a person obtain an effective problem-solving strategy for a task they have never seen, on the fly and often without explicit feedback? Some human research suggests that children learn to solve a widening range of problems through two primary processes of 1) “strategy discovery,” that is, discovering new strategies for certain problems or tasks, and 2) “strategy generalization,” that is, adapting strategies they already know for other problems or tasks (82, 83).



Fig. 5. Images eliciting Gestalt “completion” phenomena. *Left* contains only scattered line segments, but we inescapably see a circle and rectangle. *Right* contains one whole key and one broken key, but we see two whole keys with occlusion.

Some AI research on strategy discovery can be found in the area of inductive programming or program synthesis; that is, given a number of input–output pairs, constraints, or other partial specifications of a task, together with a set of available operations, the system induces a “program” or series of operations that produces the desired behaviors (84). In other words, “Inductive programming can be seen as a very special subdomain of machine learning where the hypothesis space consists of classes of computer programs” (85). Inductive programming has been applied to some intelligence test-like tasks, such as number series problems (86), and to simple visual tasks like learning visual concepts (87, 88). However, more research is needed to expand these methods to tackle more complex and diverse sets of tasks. For example, given the imagery-based strategies described above, a challenge for imagery-based program induction would be to derive these strategies automatically from a small set of example Raven's problems.

AI research has often investigated strategy generalization through the lens of integrating planning with analogy. Case-based planning looks at how plans stored in memory are retrieved at the appropriate juncture, modified, and applied to solve a new problem (89). The majority of this work has focused on agents that use propositional knowledge representations, and very little (if any) has applied these methods to address intelligence tests.

Research on strategy selection and adaptation would be enormously informative for studying not just how people approach a new type of intelligence test but also interproblem learning on intelligence tests, that is, learning from one problem (even without feedback) and using this knowledge to inform the solution of the next problem. In humans, one fascinating study gave each of two groups of children a different set of Raven's-like problems to start with, and then the same final set of problems that had ambiguous answers (53). Depending on which set of starting problems they received, the children predictably gravitated toward one of two profiles of performance on the final problems. Modeling these phenomena remains an open challenge for AI research.

Learning the Problem Definition

Even with intelligent agents that generate their own problem-solving strategies or programs, the problem definition—that is, the problem template and goal—is still provided by the human system designer. Interactive task learning is an area of AI research that investigates how “an agent actively tries to learn the actual definition of a task through natural interaction with a human instructor, not just how to perform a task better” (90). Research in interactive task learning generally involves designing agents or robots that learn from both verbal and nonverbal information, that is, instructions along with examples or situated experiences (91, 92).

Such multimodal inputs are used all of the time in human learning, including on intelligence tests: Most tests combine verbal (spoken or written) instructions with simple example problems to teach the test-taker the point of each new task that is presented. For example, the Raven's test typically begins with spoken instructions to select the answer choice that best fills in the matrix, together with a very simple example problem that the test administrator is supposed to show the test-taker, along with the correct answer.

Any Raven's agent must contain information about the problem definition in order to parse new problems appropriately and to follow a procedure that attains the goal. Moreover, agents should be able to modify their problem definition to accommodate slight problem variations. For example, if a new problem is presented with two empty spots in the matrix, a robust agent should be able to infer that this problem requires two corresponding answer responses.

In all extant Raven's agents, knowledge of the problem definition is manually provided by system designers. While these concepts may seem straightforward to a person, and indeed are usually trivial to program into an agent as static program elements, it is a challenging open question to consider where these concepts come from, and how they might be learned. For example, people gain extensive experience in taking multiple choice tests from a very early age, especially in modern societies, but we do not know precisely how this knowledge is represented, or the mechanisms by which it is generalized to new tasks.

The interesting subproblem of "nonverbal task learning" considers how the task definition can be learned purely through a small number of observed examples, without the use of explicit language-based information at all (93). While nonverbal mechanisms are undoubtedly at play in multimodal task learning for most people, nonverbal task learning in its pure form does also occur.

There are many clinical populations in which individuals have difficulties in using or understanding language, including acquired aphasia or developmental language disorders. Nonverbal intelligence tests are specifically designed for use with such populations, and they avoid verbal instructions altogether (94). In these tests, examiners initially show test-takers a simple example problem and its solution. Test-takers must learn the task definition (e.g., matching shapes, finding one shape in another, completing a visual pattern, etc.) by observing the example, and then use this knowledge to solve a series of more difficult test problems.

A small but intriguing set of converging research threads in AI have pinpointed the importance of nonverbal task learning. One recent study using robots looked at how abstract goals can be inferred from a small number of visual problem examples and applied to new problems, where the goal is represented in terms of a set of programs that meets it (95). Even more recently, a new Abstraction and Reasoning Corpus has been proposed for artificial agents, containing 1,000 visual tasks with distinct goals; agents must infer the goal for a given task from a few examples and then use this knowledge to solve new problems (96). Both of these tasks are similar to the Raven's test in the sense that, even though the Raven's test ostensibly only has a single goal (i.e., choose the answer that fits best), different Raven's problems can be thought of as requiring different formulations of this overarching and extremely vague goal. These examples also pose interesting questions about the extent to which problem goals might be implicitly represented within an agent's problem-solving strategy, instead of explicitly, and the pros and cons of each alternative.

Note that this discussion only considers goals that are well defined, at least in the minds of the problem creators. Intelligence tests are a rather odd social construct for this reason; in a way, the test-taker is trying to infer the intent of the test designer. How agents (or humans) represent and reason about their own goals might involve an extension of the processes described here, or they might be different modes of reasoning altogether.

Conclusion and Implications for Cognitive Science

We close by returning to the motivating questions from the Introduction. The cognitive science question is, what are the computations taking place in someone's mind when they use visual imagery?

AI research alone cannot, of course, fully answer this question, and so we presented a second, more limited question: If you have an intelligent agent that uses visual imagery-based knowledge representations and reasoning operations, then what kinds of problem solving might be possible, and how would it all work?

In this paper, we have presented a review of AI research and open lines of inquiry related to answering this question in the context of imagery-based agents that solve problems from the Raven's Progressive Matrices intelligence test. We discussed 1) why intelligence tests are such a good challenge for AI; 2) a framework for artificial problem-solving agents; 3) several imagery-based agents that solve Raven's problems; and 4) how an imagery-based agent could learn its domain knowledge, problem-solving strategies, and problem definition, instead of these components being manually designed and programmed.

More generally, whether or not imagery-based AI agents are at all similar to humans, designing, implementing, and studying such agents contributes valuable information about what is possible in terms of computation and intelligence. AI research that develops different kinds of agents is helpful for sketching out different points in the space of what is possible, and AI research that enables such agents to learn is helpful for hypothesizing how and why various computational elements of intelligence might come to be. Then, further interdisciplinary inquiries can proceed to connect findings and hypotheses derived from these lines of AI research to corresponding lines of research about what humans do.

Data Availability. There are no data underlying this work.

ACKNOWLEDGMENTS. Thanks go to the reviewers for their helpful comments. This work was funded, in part, by NSF Award 1730044.

1. T. Grandin, *Thinking in Pictures, Expanded Edition: My Life with Autism* (Vintage, 2008).
2. J. Gleick, *Genius: The Life and Science of Richard Feynman* (Vintage, 1992).
3. G. J. Feist, *The Psychology of Science and the Origins of the Scientific Mind* (Yale University Press, 2008).
4. N. J. Nersessian, *Creating Scientific Concepts* (MIT Press, 2008).
5. M. Giaquinto, *Visual Thinking in Mathematics* (Oxford University Press, 2007).
6. E. S. Ferguson, *Engineering and the Mind's Eye* (MIT Press, 1994).
7. M. Petre, A. F. Blackwell, Mental imagery in program design and visual programming. *Int. J. Hum. Comput. Stud.* **51**, 7–30 (1999).
8. D. W. Dahl, A. Chattopadhyay, G. J. Gorn, The use of visual mental imagery in new product design. *J. Mark. Res.* **36**, 18–28 (1999).
9. K. R. Wanzel, S. J. Hamstra, D. J. Anastakis, E. D. Matsumoto, M. D. Cusimano, Effect of visual-spatial ability on learning of spatially-complex surgical skills. *Lancet* **359**, 230–231 (2002).
10. J. Foer, *Moonwalking with Einstein: The Art and Science of Remembering Everything* (Penguin, 2011).
11. B. K. Bergen, *Louder than Words: The New Science of How the Mind Makes Meaning* (Basic, 2012).
12. J. S. Hutton et al., Home reading environment and brain activation in preschool children listening to stories. *Pediatrics* **136**, 466–478 (2015).
13. M. Hegarty, Mechanical reasoning by mental simulation. *Trends Cogn. Sci.* **8**, 280–285 (2004).
14. D. Van Garderen, Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. *J. Learn. Disabil.* **39**, 496–506 (2006).
15. J. Pearson, S. M. Kosslyn, The heterogeneity of mental representation: Ending the imagery debate. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10089–10092 (2015).
16. N. S. Newcombe, T. F. Shipley, "Thinking about spatial thinking: New typology, new assessments" in *Studying Visual and Spatial Reasoning for Design Creativity*, J. S. Gero, Ed. (Springer, 2015), pp. 179–192.
17. M. Knauff, E. May, Mental imagery, reasoning, and blindness. *Q. J. Exp. Psychol.* **59**, 161–177 (2006).
18. D. L. Schwartz, Physical imagery: Kinematic versus dynamic models. *Cogn. Psychol.* **38**, 433–464 (1999).
19. M. O. Belardinelli et al., An fMRI investigation on image generation in different sensory modalities: The influence of vividness. *Acta Psychol.* **132**, 190–200 (2009).
20. J. Hernández-Orallo, F. Martínez-Plumed, U. Schmid, M. Siebers, D. L. Dowe, Computer models solving intelligence test problems: Progress and implications. *Artif. Intell.* **230**, 74–107 (2016).
21. J. Raven, J. C. Raven, J. H. Court, *Manual for Raven's Progressive Matrices and Vocabulary Scales* (Harcourt Assessment, Inc., 1998).
22. T. G. Evans, "A program for the solution of geometric-analogy intelligence test questions" in *Semantic Information Processing*, M. Minsky, Ed. (MIT Press, Cambridge, MA, 1968), pp. 271–353.
23. R. P. Brown, E. A. Day, The difference isn't black and white: Stereotype threat and the race gap on Raven's advanced progressive matrices. *J. Appl. Psychol.* **91**, 979–985 (2006).

24. R. A. Brooks, Intelligence without representation. *Artif. Intell.* **47**, 139–159 (1991).
25. R. E. Snow, P. C. Kyllonen, B. Marshalek, The topography of ability and learning correlations. *Adv. Psychol. Hum. Intell.* **2**, 47–103 (1984).
26. V. Prabhakaran, J. A. Smith, J. E. Desmond, G. H. Glover, J. D. Gabrieli, Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's progressive matrices test. *Cogn. Psychol.* **33**, 43–63 (1997).
27. R. P. DeShon, D. Chan, D. A. Weissbein, Verbal overshadowing effects on Raven's advanced progressive matrices: Evidence for multidimensional performance determinants. *Intelligence* **21**, 135–155 (1995).
28. M. Dawson, I. Soulières, M. A. Gernsbacher, L. Mottron, The level and nature of autistic intelligence. *Psychol. Sci.* **18**, 657–662 (2007).
29. I. Soulières et al., Enhanced visual processing contributes to matrix reasoning in autism. *Hum. Brain Mapp.* **30**, 4082–4107 (2009).
30. J. Glasgow, D. Papadias, Computational imagery. *Cogn. Sci.* **16**, 355–394 (1992).
31. M. Kozhevnikov, S. Kosslyn, J. Shepard, Spatial versus object visualizers: A new characterization of visual cognitive style. *Mem. Cogn.* **33**, 710–726 (2005).
32. M. Kunda, Visual mental imagery: A view from artificial intelligence. *Cortex* **105**, 155–172 (2018).
33. A. Lovett, K. Forbus, Modeling visual problem solving as analogical reasoning. *Psychol. Rev.* **124**, 60–90 (2017).
34. P. Langley, The cognitive systems paradigm. *Adv. Cogn. Syst.* **1**, 3–13 (2012).
35. P. A. Carpenter, M. A. Just, P. Shell, What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychol. Rev.* **97**, 404–431 (1990).
36. D. Rasmussen, C. Eliasmith, A neural model of rule generation in inductive reasoning. *Top. Cogn. Sci.* **3**, 140–153 (2011).
37. C. Stranegård, S. Cirillo, V. Ström, An anthropomorphic method for progressive matrix problems. *Cogn. Syst. Res.* **22**, 35–46 (2013).
38. E. Hunt, "Quote the Raven? Nevermore" in *Knowledge and Cognition*, L. W. Gregg, Ed. (Lawrence Erlbaum, Oxford, United Kingdom, 1974), vol. 9, pp. 129–158.
39. M. Kunda, K. McGregor, A. K. Goel, A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. *Cogn. Syst. Res.* **22**, 47–66 (2013).
40. K. McGregor, M. Kunda, A. K. Goel, Fractals and Ravens. *Artif. Intell.* **215**, 1–23 (2014).
41. S. Shegheva, A. Goel, "The structural affinity method for solving the Raven's Progressive Matrices test for intelligence" in *Thirty-Second AAAI Conference on Artificial Intelligence* (Association for the Advancement of Artificial Intelligence, 2018), pp. 714–721.
42. T. Hua, M. Kunda, "Modeling gestalt visual reasoning on Raven's Progressive Matrices using generative image inpainting techniques" in *Annual Conference on Advances in Cognitive Systems* (Palo Alto Research Center, 2020).
43. Y. Yang, K. McGregor, M. Kunda, "Not quite any way you slice it: How different analogical constructions affect Raven's matrices performance" in *Annual Conference on Advances in Cognitive Systems* (Palo Alto Research Center, 2020).
44. D. Hoshen, M. Werman, IQ of neural networks. arXiv:1710.01692 (29 September 2017).
45. D. G. Barrett, F. Hill, A. Santoro, A. S. Morcos, T. Lillicrap, Measuring abstract reasoning in neural networks. arXiv:1807.04225 (11 July 2018).
46. F. Hill, A. Santoro, D. G. Barrett, A. S. Morcos, T. Lillicrap, Learning to make analogies by contrasting abstract relational structure. arXiv:1902.00120 (31 January 2019).
47. X. Steenbrugge, S. Leroux, T. Verbelen, B. Dhoedt, Improving generalization for abstract reasoning tasks using disentangled feature representations. arXiv:1811.04784 (12 November 2018).
48. S. van Steenkiste, F. Locatello, J. Schmidhuber, O. Bachem, Are disentangled representations helpful for abstract visual reasoning? arXiv:1905.12506 (29 May 2019).
49. C. Zhang, F. Gao, B. Jia, Y. Zhu, S. C. Zhu, "Raven: A dataset for relational and analogical visual reasoning" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, 2019), pp. 5317–5327.
50. C. E. Bethell-Fox, D. F. Lohman, R. E. Snow, Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence* **8**, 205–238 (1984).
51. M. Kunda, "Visual problem solving in autism, psychometrics, and AI: The case of the Raven's Progressive Matrices," PhD thesis, Georgia Institute of Technology, Atlanta, GA (2013).
52. M. Barnsley, L. P. Hurd, A. K. Peters, *Fractal Image Compression* (AK Peters, Boston, MA, 1992).
53. J. R. Kirby, M. J. Lawson, Effects of strategy training on progressive matrices performance. *Contemp. Educ. Psychol.* **8**, 127–140 (1983).
54. J. Yu et al., "Generative image inpainting with contextual attention" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Institute of Electrical and Electronics Engineers, 2018), pp. 5505–5514.
55. M. Kunda, "Computational mental imagery, and visual mechanisms for maintaining a goal-subgoal hierarchy" in *Proceedings of the Third Annual Conference on Advances in Cognitive Systems (ACS)*, A. Goel, M. Riedl, Eds. (Cognitive Systems Foundation, 2015), p. 4.
56. A. Newell, H. A. Simon, Computer science as empirical inquiry: Symbols and search. *Commun. ACM* **19**, 113–126 (1976).
57. A. X. Chang, M. Savva, C. D. Manning, "Learning spatial knowledge for text to 3D scene generation" in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, W. Daelemans, Eds. (Association for Computational Linguistics, 2014), pp. 2028–2038.
58. R. N. Shepard, Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychol. Rev.* **91**, 417–447 (1984).
59. R. Held, A. Hein, Movement-produced stimulation in the development of visually guided behavior. *J. Comp. Physiol. Psychol.* **56**, 872–876 (1963).
60. G. Wiedenbauer, J. Schmid, P. Jansen-Osmann, Manual training of mental rotation. *Eur. J. Cogn. Psychol.* **19**, 17–36 (2007).
61. G. Wiedenbauer, P. Jansen-Osmann, Manual training of mental rotation in children. *Learn. Instr.* **18**, 30–41 (2008).
62. B. W. Mel, "A connectionist learning model for 3-d mental rotation, zoom, and pan" in *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (Cognitive Science Society, 1986), pp. 562–571.
63. K. Seepanomwan, D. Caligiore, G. Baldassarre, A. Cangelosi, Modelling mental rotation in cognitive robots. *Adapt. Behav.* **21**, 299–312 (2013).
64. R. P. Goebel, The mathematics of mental rotations. *J. Math. Psychol.* **34**, 435–444 (1990).
65. R. Memisevic, G. E. Hinton, Learning to represent spatial transformations with factored higher-order Boltzmann machines. *Neural Comput.* **22**, 1473–1492 (2010).
66. R. Memisevic, Learning to relate images. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1829–1846 (2013).
67. C. Finn, I. Goodfellow, S. Levine, "Unsupervised learning for physical interaction through video prediction" in *Advances in Neural Information Processing Systems* D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, R. Garnett, Eds. (Neural Information Processing Systems Foundation, 2016), pp. 64–72.
68. R. Mottaghi, M. Rastegari, A. Gupta, A. Farhadi, "What happens if...? learning to predict the effect of forces in images" in *European Conference on Computer Vision*, B. Leibe, J. Matas, N. Sebe, M. Welling, Eds. (Springer, 2016), pp. 269–285.
69. N. Watters et al., "Visual interaction networks: Learning a physics simulator from video" in *Advances in Neural Information Processing Systems*, I. Guyon et al., Eds. (Neural Information Processing Systems Foundation, 2017), pp. 4539–4547.
70. A. Nair et al., "Combining self-supervised learning and imitation for vision-based rope manipulation" in *2017 IEEE International Conference on Robotics and Automation (ICRA)* (Institute of Electrical and Electronics Engineers, 2017), pp. 2146–2153.
71. K. Kotovsky, H. A. Simon, What makes some problems really hard: Explorations in the problem space of difficulty. *Cogn. Psychol.* **22**, 143–183 (1990).
72. R. Primi, Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence* **30**, 41–70 (2001).
73. J. Wagemans et al., A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol. Bull.* **138**, 1172–1217 (2012).
74. G. Kanizsa, *Organization in Vision: Essays on Gestalt Perception* (Praeger, 1979).
75. A. Desolneux, L. Moisan, J. M. Morel, *From Gestalt Theory to Image Analysis: A Probabilistic Approach* (Springer Science & Business Media, 2007), vol. 34.
76. C. B. Schönlieb, *Partial Differential Equation Methods for Image Inpainting* (Cambridge University Press, 2015).
77. M. H. Herzog, U. A. Ernst, A. E. Eitzold, C. W. Eurich, Local interactions in neural networks explain global effects in Gestalt processing and masking. *Neural Comput.* **15**, 2091–2113 (2003).
78. C. Prodhöhl, R. P. Wörtz, C. Von Der Malsburg, Learning the Gestalt rule of collinearity from object motion. *Neural Comput.* **15**, 1865–1896 (2003).
79. A. Amanatiadis, V. G. Kaburlasos, E. B. Kosmatopoulos, "Understanding deep convolutional networks through Gestalt theory" in *2018 IEEE International Conference on Imaging Systems and Techniques (IST)* (Institute of Electrical and Electronics Engineers, 2018), pp. 1–6.
80. G. Ehrensperger, S. Stabinger, A. R. Sánchez, Evaluating CNNs on the Gestalt principle of closure. arXiv:1904.00285 (30 March 2019).
81. B. Kim, E. Reif, M. Wattenberg, S. Bengio, Do neural networks show Gestalt phenomena? An exploration of the law of closure. arXiv:1903.01069 (4 March 2019).
82. D. F. Bjorklund, *Children's Strategies: Contemporary Views of Cognitive Development* (Psychology, 2013).
83. R. Siegler, E. A. Jenkins, *How Children Discover New Strategies* (Psychology, 2014).
84. S. Gulwani et al., Inductive programming meets the real world. *Commun. ACM* **58**, 90–99 (2015).
85. J. Hernández-Orallo, S. H. Muggleton, U. Schmid, B. Zorn, Approaches and applications of inductive programming (Dagstuhl Seminar 15442). *Dagstuhl Rep.* **5**, 89–111 (2016).
86. J. Hofmann, E. Kitzelmann, U. Schmid, "Applying inductive program synthesis to induction of number series a case study with igor2" in *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, C. Lutz, M. Thielscher, Eds. (Springer, 2014), pp. 25–36.
87. B. M. Lake, R. Salakhutdinov, J. B. Tenenbaum, Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).
88. K. Ellis, D. Ritchie, A. Solar-Lezama, J. Tenenbaum, "Learning to infer graphics programs from hand-drawn images" in *Advances in Neural Information Processing Systems*, S. Bengio et al., Eds. (Neural Information Processing Systems Foundation, 2018), pp. 6059–6068.
89. D. Borrajo, A. Roubířková, I. Serina, Progress in case-based planning. *ACM Comput. Surv.* **47**, 35 (2015).
90. J. E. Laird et al., Interactive task learning. *IEEE Intell. Syst.* **32**, 6–21 (2017).
91. T. R. Hinrichs, K. D. Forbus, X goes first: Teaching simple games through multimodal interaction. *Adv. Cogn. Syst.* **3**, 31–46 (2014).
92. J. Kirk, A. Miner, J. Laird, Learning task goals interactively with visual demonstrations. *Biol. Inspir. Cogn. Arc.* **18**, 1–8 (2016).
93. M. Kunda, "Nonverbal task learning" in *Proceedings of the 7th Annual Conference on Advances in Cognitive Systems*, M. T. Cox, Ed. (Cognitive Systems Foundation, 2019).
94. L. S. DeThorne, B. A. Schaefer, A guide to child nonverbal IQ measures. *Am. J. Speech Lang. Pathol.* **13**, 275–290 (2004).
95. M. Lázaro-Gredilla, D. Lin, J. S. Guntupalli, D. George, Beyond imitation: Zero-shot task transfer on robots by learning concepts as cognitive programs. *Sci. Robot.* **4**, eaav3150 (2019).
96. F. Chollet, On the measure of intelligence. arXiv:1911.01547 (5 November 2019).