

## *Problem Solving*

### *Definition of the Main Concepts*

#### 1.1 Gestalt Influence

It is almost universally acknowledged that problem solving is one of the most, if not the most fundamental cognitive ability. We solve problems, both small and large, easy and difficult, all of the time. Problem solving includes planning your way home, planning a tour of several cities or countries for a vacation, playing chess, solving physics and math problems, as well as proving math theorems. It also includes creative problem solving, such as formulating scientific theories. The field called computer science (CS) has always concentrated on problem solving. A subfield of CS called artificial intelligence (AI) started, in the middle of the last century, by asking whether a computer could solve problems.

Now, consider the toy problem of constructing four equilateral triangles by using six identical matchsticks. This problem appears, at first, impossible because after you construct the first triangle by using three matchsticks and use the next two matchsticks to produce the second triangle simply by sharing one edge with the first triangle, you are left with only one matchstick to construct two more triangles. The solution becomes obvious once you realize that this problem should be solved in a three-dimensional (3D) space, rather than on the two-dimensional (2D) flat table on which you surely decided to work when you began to solve this problem. It is also important to note that the 3D solution (a regular tetrahedron) is highly symmetrical. This is not a coincidence because the task was to construct four identical triangles. As a result, the tetrahedron can be transformed (mapped) to itself by applying multiple 3D rotations and reflections. The problem of constructing four equilateral triangles using six matchsticks is usually categorized as a member of the class called “insight problems.” Insight problems were used by the gestalt psychologists in the beginning of the twentieth century when they brought the attention of psychologists to problem solving as a mental function. These gestalt psychologists were convinced that insight problems are special because they are usually solved by changing the mental representation of the problem, rather than by learning and experience. The gestalt psychologists claimed that *everyone* who solves this four triangles problem begins with the 2D representation, and that the problem will not be solved until the problem solver changes their representation to 3D. Recall that the “gestalt revolution” was launched as a reaction to the *empiristic* approach in psychology, which claimed that our mental abilities were the result of learning by accumulating sensory experience starting at birth. John Locke’s name is usually mentioned here as a clear example of empiristic thinking because he claimed that the mind of a newborn baby is a *tabula rasa*

(a blank slate) upon which one's life experiences will be written. The gestalt psychologists were *nativists* because they held that the mind of a newborn baby already has some innate knowledge of the external world, and that this knowledge includes some abstract characteristics such as the concept of causality, the three-dimensionality of our physical world, some basic concepts contained in Euclidean geometry, such as the straight line and symmetry, as well as the concept of motion. Note that this is by no means an exhaustive list of the innate concepts that were postulated by the gestalt psychologists. For our purposes, the most important assumptions the gestalt psychologists made were that innate intuitions exist and that learning accumulated during one's lifetime is not the only source of knowledge about our external world. This nativistic view is no longer considered to be either extreme or exotic as it was a century ago. Thanks to the constant progress made in genetic science since Watson and Crick published their model of the structure of DNA in *Nature* in 1953 after "borrowing" some of Rosalind Franklin's data, it is reasonable to assume nowadays that considerable learning occurred during the course of our evolution and that this evolutionary experience is now coded in our genes.

Consider three examples used by Hochberg (1978) to illustrate nativism in visual perception, the second area, besides problem solving, where the gestalt psychologists provided fundamental and long-lasting contributions. Thorndike (1899) showed that a newly hatched chick would jump off a low stand, but it refused to jump from a tall one, demonstrating that depth perception, an essential aspect of 3D vision, is innate in chickens. In fact, it also suggests that the chick not only perceives depth, but also has some intuitions about both gravity and mass. We humans are smarter than chickens, so it is reasonable to assume that we have at least as many useful innate intuitions as chicks do. This was demonstrated by Gibson and Walk (1960) who showed that a toddler systematically refused to crawl off a deep cliff. The cliff was actually covered by a strong sheet of glass that eliminated all possibility of the toddler falling off the cliff if it could not perceive depth and crawled onto the deep side. Also note that the toddler's perception of depth was aided by covering the table and the region around and below it with a highly structured checkerboard pattern. The toddler showed that he trusted his 3D vision by avoiding the "visually deep" region beyond the table's edge. When the cliff was shallow, the toddler had no problem crawling over it. In the third example, Michael Wertheimer (1961) showed that a newborn infant turned her head in order to look in the direction from which a sound was coming, demonstrating that she was born with concepts of auditory, visual, and motor spaces, and that all of these three spaces were already coordinated when she was born.

Before we introduce some new concepts used to define problem solving, consider another insight problem, in which a solution becomes obvious when you change the original representation. Assume that at 9 AM you started walking uphill on a mountain trail. It took you several hours to get to the top of the mountain. To make things more complex, your speed of walking was not constant. You stayed overnight at a hotel on the top of the mountain and the next morning, at 9 AM, you started walking down the hill using exactly the same trail. Walking downhill was easier and faster. But you wondered whether there was a place on the trail that you passed today at exactly the same time as yesterday when you started by walking uphill. For most people it is hard to wrap their head around this problem. Some people are inclined to say yes, others think that there was no such place. Regardless of the proposed answer, it is difficult to come up with an

intuitive justification of the answer chosen. The only way to make the answer obvious to everyone is to provide a different representation of the same problem. Specifically, you should consider two people walking on the same trail on the same day in opposite directions, one from the bottom of the hill and the other from the top. If they started at the same time (or more generally, if one person started before the other finished), they had to meet each other somewhere on the trail and the place where they met is the place you passed at the same time on both days. It does not matter for solving the problem where they met. In fact, there is not enough information to compute where they met. The important thing for solving this problem is to realize that they had to meet because they walked in opposite directions on the same trail. In this case, instead of thinking about one person walking up and then down the trail on two consecutive days, one should think about two people, one walking up and the other down, on the same day on the same trail. So, again, changing the representation of the problem does the trick.

We will discuss insight problem solving in detail in Chapter 6, but at this point, I owe the reader a more constructive and more rational criterion that allows for classifying a problem as an insight problem. The observation that one's mental representation needs to change before an insight problem is solved is interesting, but before the mental representation actually changes, we have no criterion for knowing whether such a change will actually ever happen. Furthermore, even if the representation does change, we can never see or directly measure someone else's mental representations. This means that the "changing representation" criterion will remain both elusive and subjective until we know more about how it works.

## 1.2 Insight Problems: The Status of the "Aha!" Criterion

Gestalt psychologists emphasized the role of changing the representation of a problem, but they did not have a theory of how any mental representation, correct or incorrect, is produced in the first place. In the remainder of this chapter, we will develop the argument that a mental representation is the *sine qua non* of problem solving, that is a *necessary condition*, without which problem solving cannot occur. This elaboration allows us to include a large class of problems that are not insight problems in our discussion. Furthermore, it will build a bridge to AI research, a specialty that has concentrated its efforts on non-insight problems.

If someone has not been influenced by the gestalt psychologists, including their conjecture about the significance or importance of changing the representation in insight problem solving, they are likely to use an alternative criterion for deciding whether the problem is an insight problem. It is called the "aha!" experience. If you exclaim "aha!" at the moment the solution suddenly occurs to you, you had an *insight*. It is not difficult to see that this criterion is as catchy as it is useless. Its claim to fame seems to be based exclusively on being an English version of the Greek "Eureka" ("I found it") that Archimedes shouted when he discovered a method for measuring the volume of an irregularly shaped object in his bathtub. The problem Archimedes was trying to solve was verifying whether a crown was made of pure gold, or whether some gold was replaced by silver. Silver is less dense than gold, so a mixture of silver and gold would have a slightly greater volume than a piece of pure gold of the same weight. Archimedes solved this

problem by examining the invariance of a balance with the crown suspended from one arm of the balance and a piece of pure gold suspended from the other arm, when this system was translated from one physical medium (air) to another (water). If the balance is maintained after the two arms are simultaneously submerged in the water, one can conclude that the goldsmith did not cheat by reducing the purity of the gold. Could Archimedes have solved his problem by simply collecting and comparing the volumes of the displaced water for the crown and for the piece of pure gold? Historians claim that in Archimedes' day, there were no methods for precise measurement of the volume of water. He surely could and did verify the invariance of his balance, as described here. It took physicists more than 20 centuries to fully appreciate the fundamental importance of invariance. The idea here is that if you get a bright idea and say "aha!," your experience was similar to Archimedes's, even if your discovery is not as creative as his. Everyone in cognitive psychology seems to agree that insight problems are special, so the question becomes whether we can offer less elusive criteria than either the "aha!" experience or the change in mental representation. Here is my insight about insight: most problems that people classify as insight are difficult to solve, but their solutions, once guessed, are easy to explain and verify. This is, in fact, an informal definition of one of the most important concepts in computational complexity theory, namely NP-completeness, a concept that was introduced into computer science 50 years ago (see Garey & Johnson, 1979). This observation shows that the gestalt psychologists were actually ahead of their time when they insisted on the special role of insight problems in problem solving. Is every NP-complete problem an insight problem? Probably not. But NP-completeness offers a fresh look at the concept of insight and it provides an *objective criterion* for the classification of problems. This idea will be elaborated later in this book, when invariance and symmetry are discussed in greater detail.

Before we move to problems that are not insight problems, consider one more illustration of how an insight problem is solved. This problem comes from Polya's (1945) classic book titled *How to Solve It*. The task is to inscribe a square into an arbitrary triangle. By "inscribe," we mean that two corners of the square are located on one of the three sides of the triangle, and the other two corners touch the two remaining sides of the triangle, as shown in Figure 1.1a. The problem boils down to deciding the size of the square that will fit into the particular triangle. If the square is too small, or too large, it will not satisfy this requirement (Figure 1.1b). But, if we start "inscribing" arbitrary squares, either too small or too large, as is shown on Figure 1.1b, we actually have the solution at hand. Do you see it? The top-right corners of all these squares are collinear. If the top-left corner of the square with an arbitrary size touches the left side of the triangle (as it does in Figure 1.1b), the line connecting the bottom-left vertex of the triangle with the top-right corner of the arbitrary square intersects the right side of the triangle at the point that is at the top-right corner of the inscribed square. This solves our problem (Figure 1.1c). It is easy to see that this is an application of the intercept theorem attributed to Thales (c.624–c.548 BCE), which is based on the similarity of triangles. It is this similarity of triangles that leads to the invariant ratios that are needed here. According to some historical sources, Thales used this theorem to estimate the height of Cheops's Pyramid by using its shadow. Aha! Note this use of Thales's insight in a new context. This is insight, too.

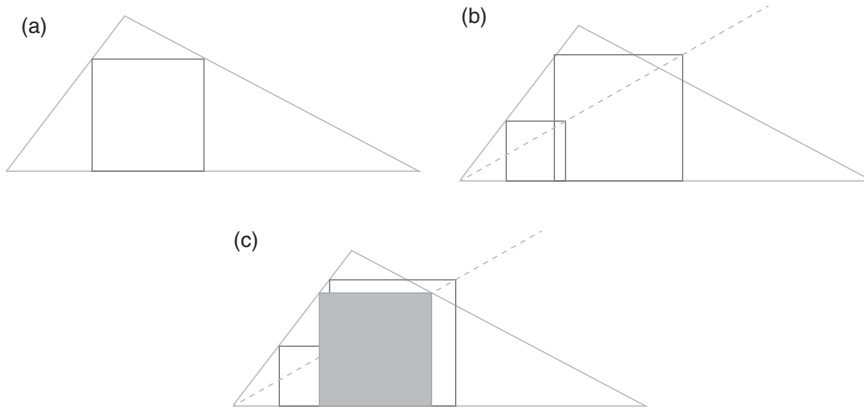


Figure 1.1 Inscribing a square into a triangle (from Polya, 1945). (a) inscribed square. (b) squares whose sizes are not correct. (c) the dashed line intersects the right side of the triangle at the point that is the top-right corner of the inscribed square.

### 1.3 Search Problems

This book will also discuss search problems, which are different from insight problems. Search problems did not attract the attention of the gestalt psychologists, but they became the center of interest of the two fathers of AI, Newell and Simon, when they started to write computer algorithms to solve logic problems near the middle of the twentieth century. Search problems have remained the focus of much AI research ever since. Take the 15-puzzle problem as an example (see Figure 1.2). This puzzle is a one-person game whose history is at least as long as the history of gestalt psychology. This game is, in fact, a member of another class of problems in computational complexity theory, namely a class called NP-hard. Informally, NP-hard problems are (i) difficult to solve, and (ii) it is difficult to verify whether a solution is actually a solution. So, what is this game about? Think about this game as a physical board with 15 movable tiles and one empty space. Legal moves are those that move a tile horizontally or vertically to an empty space.<sup>1</sup>

Figure 1.2 shows a possible start state and a goal state. The task is to rearrange the tiles from a start state to the goal state with the fewest number of moves. The number of different states in the 15-puzzle is one half of the number of permutations of 16 elements. Recall that the number of permutations of  $N$  elements is represented by the symbol  $N!$  ( $N$  factorial) and it is a product of all natural numbers between 1 and  $N$ :  $N! = N \cdot (N-1) \cdot (N-2) \cdot \dots \cdot 2 \cdot 1$ . It is easy to see why permutations represent different states of the 15-puzzle. But why divide  $16!$  by 2? The reason is that legal moves can produce only one half of all possible permutations. The other half of the permutations are not accessible. In other words, the 15-puzzle consists of two disjoint sets of states and one cannot jump from one set to another by making legal moves. This fact can be proved mathematically by considering the invariance aspect of permutations (you will hear a great deal about

<sup>1</sup> Here is an example of a virtual 15-puzzle: [www.artbylogic.com/puzzles/numSlider/numberShuffle.htm?rows=4&cols=4&sq=1](http://www.artbylogic.com/puzzles/numSlider/numberShuffle.htm?rows=4&cols=4&sq=1).

5	1	3	4
2		7	8
9	6	10	12
13	14	11	15

(a)

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

(b)

Figure 1.2 A start state (a) and a goal state (b) (from Pizlo & Li, 2005).

invariants and symmetries in this book). The existence of these two disjoint sets was used 100 years ago to popularize this puzzle. The start state was produced by swapping (physically) tiles 14 and 15. A big monetary prize was offered for a solution that would produce the goal state shown in Figure 1.2b by any sequence of legal moves. We now know that this is impossible. In fact, swapping any pair of tiles physically brings the 15-puzzle to the other half of the permutations. All of this will be explained in Chapter 5. At this point, the 15-puzzle is important for us because it is a good way to illustrate the fact that problem solving can be thought of as a *goal-directed activity*. Indeed, the very definition of how the 15-puzzle is played includes the concept of the start state and the goal state, and the task (problem) is to get to the goal.

The early stages of AI research did use the 15-puzzle, as well as other sizes of the same type of puzzle, namely, the 5-, 8-, and 24-puzzle, as examples for formulating theories of problem solving. The number of states in these puzzles is  $(N+1)!/2$ . Here  $N$  is the number of tiles and  $(N+1)$  is the size of the board ( $2 \times 3$ ,  $3 \times 3$ , and  $5 \times 5$  in the examples just mentioned). The fact that these puzzles are members of the class of NP-hard problems means that finding the shortest number of moves to the goal state may require a brute-force search through most or all of the states. This kind of search is impractical because  $N!$  is a large number and it grows very quickly with  $N$ . For the 15-puzzle, the number of states that can be produced by legal moves is  $16!/2 \approx 10^{13}$ . This is 100 times more than the number of neurons in your brain. Another way to illustrate how big this number is, is to realize that if you started at the time of the Big Bang, that is, about 14 billion years ago, and kept producing 2 states a day, you would have just finished looking through all of the states in the 15-puzzle. And, if you produced half a billion states per second and started at the Big Bang, you would have just finished looking through all states of a 24-puzzle. Do people solving such problems actually examine a large or a small fraction of all possible states, and if they examine only a small fraction, how small is this fraction? You will surely be surprised by the answers when you get them.

#### 1.4 The Scientific Status of Goal-Directed Behavior

Interestingly, but not terribly surprisingly, all examples of *problem solving*, not only the 15-puzzle, can be viewed as a “goal-directed activity” (e.g., Newell & Simon, 1972; Anderson, 1980; Russell & Norvig, 2018). Planning a tour that visits several countries in Europe, following the shortest path to reach a goal in a maze, playing a game of chess, proving the Pythagorean theorem in geometry, designing an experiment to test a new theory in the natural sciences, formulating a new theory in science, or deciding about

one's career path are all examples that make it clear that problem solving is a goal-directed behavior. Here, I will explain the nature of a goal-directed behavior by emphasizing how it is different from behavior as it was conceived in the now outdated stimulus–response approach, introduced and favored by the Watson/Hull behaviorist tradition and rejected by Tolman and Lashley working in the cognitive tradition.

The stimulus–response tradition, labeled by its originators *behaviorism*, assumed that all behavioral actions of an animal are a direct consequence of the physical stimuli impinging on the animal. This view of behavior might have looked scientific back in the day because it did not violate the prevailing view of our physical world, according to which an *effect* always *follows* a *cause*. When one moving billiard ball hits another, the resulting movement of the second ball is the *effect* that follows the impact, the *cause*, produced by the movement of the first ball. It would obviously be counterintuitive to think that the movement of the *first* ball, before the impact, was caused by the movement of the *second* ball after the impact, as if the first ball initiated its movement *in order* to make the second ball move. A future event cannot be the cause of an event occurring now. This has been the accepted view everywhere in science during the modern era that started in the seventeenth century. Recall that ancient Greek philosophers such as Aristotle 2,400 years ago, did allow a reversed order of a cause and its effect. According to Aristotle, when you drop a stone, it falls down because it *wants* to be as close as possible to its natural place, which for a heavy object like a stone is the center of the Earth. So, being close to the center of the Earth in the next few seconds or minutes is actually what *causes* the movement of the stone, now. This kind of explanation, which is called *teleological*, has been discredited in science for a long time, but recently we were faced with a need to revive it, somehow, in order to fit goal-directed behavior into the realm of modern cognitive science. How can this be done? Consider the following everyday life example. I turn on my coffee maker now *in order to* drink coffee five minutes later. Did my drinking coffee five minutes in the future cause me to turn on the coffee maker, now? A contemporary physicist would say that this is impossible. And s/he would be right. So, what is going on?

Howard Warren (1916) provided the first satisfying conceptual explanation of causality in the nature of goal-directed (purposive) behavior. He said that “A human act is said to be purposive when it is preceded by an idea representing the situation which the act itself brings about” (8). The essence of what Warren is saying is that although a future goal cannot cause (control) the present action, a *model* or *representation* of the future goal can control the present action. If the model is accurate (the coffee maker works), the goal will most likely be achieved. But, if the model is not accurate (the coffee maker is broken), the goal will not be attained. There is nothing mysterious here; no laws of physics are contradicted, but there cannot be any goal-directed (purposive, intelligent) behavior without accurate mental representations of the environment. This observation was the cornerstone needed to launch what we now call the Cognitive Revolution (Miller, Pribram & Galanter, 1960; Neisser, 1967).

Mental representations of quite a few animals were studied in the twentieth century, including, rats, dogs, monkeys, and chimpanzees, as well as humans. As pointed out earlier, the gestalt psychologists assigned a central role to mental representations when they called attention to the fact that solving an insight problem requires changing its



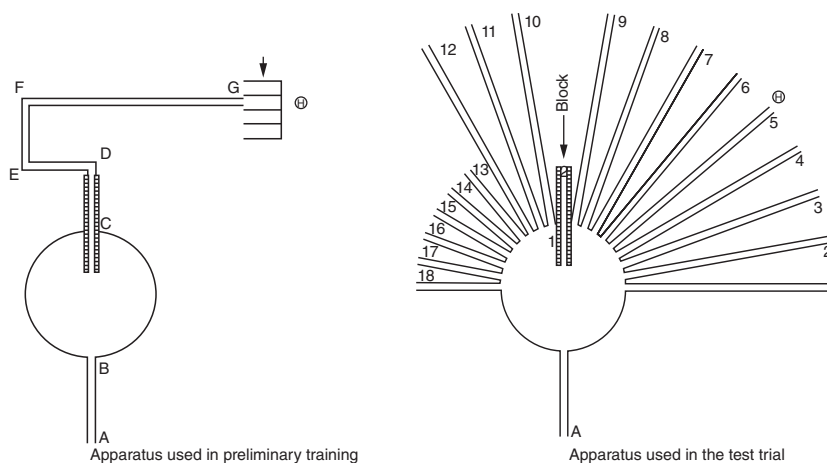


Figure 1.3 Mazes used by Tolman (1948).

representation. But now we can see that the concept of mental representation is even more fundamental than the gestalt psychologists claimed because mental representation is a necessary condition for any goal-directed action, including solving problems that are not insight problems. Without mental representations, goal-directed actions would remain outside of modern science, and even more importantly, we humans would be unable to plan and carry on goal-directed actions if we did not have mental representations. Without goal-directed actions, we humans could not be “intelligent.” Edward Tolman, who worked in the first half of the twentieth century, was one of the first to use the concept of goal-directed (purposive) behavior in his theories and experiments. Look at Figure 1.3 taken from his 1948 paper. The rat was trained in the simple maze shown on the left. The entrance is marked as A and the goal (food) is on top-right rendered with an H within a circle. After the training was completed, the rat was presented with the maze shown on the right, whose entrance was identical to the entrance of the training maze, but the rest of the maze was changed. Faced with a blocked alley that used to go to the goal, the rat came back to the circular chamber, and almost immediately ran along the alley marked as 5 which led directly to the position where the food had been located during the training trials. The rat chose an available shortcut, when the familiar path was blocked. This choice could not be a result of training. It was the result of the rat creating and using an accurate spatial mental map of the maze.

Dogs and chimpanzees also can use spatial maps when they go around an obstacle. Wolfgang Köhler used the configuration in which a dog or a chimp stands on one side of a transparent fence, and food is placed on the other side of the fence. The animal quickly realizes that the fence cannot be penetrated, looks around and runs around the fence. This behavior is not trivial because the animal must, at first, face away, putting the food out of sight as it turns and runs away from the food. But once we assume that the animal has a spatial map of its environment, this behavior seems natural. The chicken, whose “intelligence” I praised when I discussed its innate depth perception earlier in this chapter, fails this “obstacle test.”



Humans can obviously pass this kind of “obstacle test,” and they do it in exactly the same way, by using a mental representation of their environment. This mental representation consists of both physical and geometrical characteristics. Geometry is represented by Euclidean lines and points on a plane, and physics is represented by the assumption that the fence is impenetrable. Physics and geometry were also used in planning and executing our solution to the 15-puzzle. All of our physical behaviors are based on geometrical/physical representations.

To summarize this discussion of goal-directed actions and mental representations: what I am saying is that mental events, namely mental representations of future goals and future actions, are essential for explaining goal-directed behavior, specifically, it is my plan to drink coffee, not drinking the coffee itself, that causes me to turn on the coffee maker. This view of goal-directed behavior brings mental events into the forefront of natural science because this view implies that ***mental events, in the form of abstract representations and plans, can cause physical events such as behavioral actions.*** Note that this claim is almost never made explicit or emphasized in cognitive science despite the fact that without this claim, there would be no reason to talk about a Cognitive Revolution because cognitive science would not be bringing anything new to the natural sciences. Note that I am not proposing what Gilbert Ryle criticized in his 1949 book as a Cartesian “ghost in a machine.” Abstract representations and plans never exist without physiological “hardware” in the form of neuronal circuits in a biological brain or without physical hardware in the form of electronic circuits in a computer. But keep in mind that despite the fact that hardware is necessary for abstract representations to be formed and used, hardware *alone* is not sufficient to actually explain these representations.

Consider the following example. When I tell you to move tile “1” from its current place, e.g., the state in Figure 1.2a, to the top-left corner, you can execute this plan fairly easily. Also, if I tell you to move tile “1” from its current place to the top-right corner, you can execute this plan, as well. The movements of your fingers can be traced back to the activity of neurons in the motor cortex in your brain that sent the motor signals to your fingers, and the activity of these motor neurons was, in turn, caused by the activity of the neurons in your visual cortex when you looked at the 15-puzzle. In fact, this would have to be a recursive (repetitive) sequence of neural events in the visual and motor cortices because there would have to be several moves of individual tiles performed under visual control, and the whole sequence would stop after tile “1” ends up in the goal position. I could also say that the entire sequence of activities of visual and motor neurons would stop after a signal from some other neurons caused a pause of all the motor acts. The actual sequence of motor acts, including the command to stop, would be controlled by the neural signals in your brain that represent my verbal charge for you to move “1” to the top-left (or the top-right) corner. But my verbal charge can also be viewed as an acoustic wave generated by my vocal chords that was caused by neuronal firing in the “language” part of my brain. This causal chain of events (neurons in my brain, acoustic wave, neurons in your brain and then movements of your fingers) looks like a reasonable way to describe what happened during solution of the 15-puzzle, except that there is no way to use this description as an explanation of what happened. The only sensible way to explain what happened is to say that you were told to move tile “1” to the top-left (or right) corner of the puzzle. That’s all that’s needed here. Neurons and acoustic waves is

the wrong *level of analysis*, despite the fact that all of these events were involved.<sup>2</sup> The real, albeit abstract action, happened on a higher level that included a representation of the physical environment (the 15-puzzle), as well as the goals and the plans required to achieve them. This makes sense considering the fact that you were actually interacting with the physical tiles that you moved mechanically from one place to another. In other words, the goal-directed behavior was defined by the 15-puzzle, not by the neurons in your brain: when you were making progress with moving tile 1 to its intended position, it was tile 1 itself that was moving closer to the goal position. Saying that it was the neuronal response representing tile 1 that was becoming more similar to the neuronal response representing the goal position, seems, at best, a very awkward “translation” of what really happened in the physical environment. The *abstract representation* of the geometrical and physical characteristics of the game, and of its rules, as well as of your actions, is the *only* “common denominator” in the system that includes both us and the physical array of tiles. After all, the *only* criterion for deciding whether your goal-directed action was successful is the situation in the physical world, not in your brain. Finally, note that the brain is quite complex: it has billions of neurons and even more connections among its neurons. We may never be able to figure out which neurons and which connections correspond to a particular mental and physical event. So, our explanation, based on mental representations of the physical and geometrical environment, is simpler (more economical). Finally, note that this explanation satisfies the long-prized scientific principle called Occam’s razor.

The observation that mental events – such as representations of goals, representations of the environment, reasoning about the goals, goal-directed (purposive) actions, information processing, memory, learning, inferences, concept formation, thinking (including creative thinking), and language – cannot be reduced to the laws of physics (although they do not violate these laws), led to a revolution in psychology. We call this the “Cognitive Revolution.” It also marked the beginning of artificial intelligence (AI), which is the quest for emulating the human mind with a machine. More generally, AI has the goal of creating thinking machines. Finally, note that the main push behind AI actually came from the attempt to make mechanistic versions of teleological systems (Rosenblueth et al., 1943).

It was computer scientists and engineers such as Wiener (1948), Shannon (1948), and Turing (1950) whose theories and technological inventions in the areas of information theory, cybernetics, and artificial intelligence, demonstrated to the students of cognitive science how they can put their ideas of goal-directed behavior on solid grounds. Wiener’s (1948) theory of control systems was critical in this context. He pointed out that control systems can behave in a goal-directed way, providing a technical explanation of Warren’s (1916) concept of purposive behavior. Control systems of the sort described by Wiener

<sup>2</sup> This description of two levels of analysis, one representing the physiological and the other representing the mental level, is completely analogous to the double-aspect view in the mind/body problem, which is a characteristic of a neutral monism. One of the best illustrations of the double-aspect view comes from William James (1904: 480) who compared (i) a *physical* description of a painting, like Raphael’s *The School of Athens*, where the description characterizes the chemical composition of the paint at each point in the fresco, to (ii) a *mental* description, in which one would simply say that the painting depicts a number of students, all involved in conversations. Both of these two descriptions are valid, but one cannot be easily translated into the other.

had been used by engineers for centuries (see Bennett, 1979, 1993 for an historical review). The main three elements of a control system are the sensors, which measure the current state of the system and its environment, the desired set-point (representation of a goal) for the state of the system, and the negative feedback loop that uses the difference (error) between the current and desired states to correct the current state. The control of temperature in a house is one of the simplest examples of an engineering control system. Maintaining body temperature, blood sugar level, and blood CO<sub>2</sub> level are examples of biological control systems that serve to achieve and maintain *homeostasis* (Bernard, 1878; Cannon, 1932). Engineering and biological control systems emphasize the nature of the negative feedback and the stability of the control. In a goal-directed behavior, such as problem solving, it is the representation (model) of the environment, produced from sensory data, that takes center stage. The feedback correction and the stability of control are also important, but it is the representation itself that seems to be absolutely fundamental in problem solving.

### 1.5 Forming Mental Representations

How are mental representations of the environment formed? Mental representations are produced from sensory data, but the data are never sufficient, so mental representations are always based on *inferences*. Consider the well-understood case of visual representations. Our physical world is three-dimensional (3D), but the visual data, which are available on the retina of the human eye, are two-dimensional (2D). But, despite the fact that the retinal image is 2D, our visual representations of the environment are 3D. Furthermore, we actually see things “veridically,” which simply means that we see 3D objects and 3D scenes the way they are “out there.” Using technical jargon, the problem of recovering a 3D visual representation from 2D retinal data is *ill-posed*. This means that there are always infinitely many 3D interpretations that are consistent with any 2D retinal image. The only way to choose a unique and veridical interpretation is to impose a priori constraints on the family of possible interpretations (Pizlo, 2001). Combining visual data and constraints leads to solving a *constrained optimization problem*, that is, finding a 3D interpretation that is as close to the visual data as possible, and, at the same time, as close to the a priori constraints as possible. The constraints could be either innate (hardwired into the brain and present at birth), or learned after birth. The *symmetry* of objects is the most fundamental visual constraint (Pizlo, 2008). Although we do not have direct empirical evidence, it seems pretty certain that the symmetry constraint is innate. This is why all humans see 3D objects and scenes in the same way. Note that symmetry is a mathematical concept that does not have to be learned. Symmetry refers to the intrinsic geometrical characteristics of an object. In the case of mirror-symmetry, which is the most common type of symmetry in nature, the left half of an object is identical or similar to its right half. One does not need to see many examples of mirror-symmetrical objects in order to form the concept of mirror-symmetry. Mirror-symmetry can be defined mathematically. Symmetry is also fundamental in physics – physicists claim that symmetry existed in the universe starting right after the Big Bang, which happened about 14 billion years ago. It follows that when life emerged on Earth about 1 billion years ago, symmetry had already been present in the environment for millennia. Symmetry

surrounded the first life forms as the first animals evolved. In fact, the bodies of most animals have always been mirror-symmetrical because of the animal's locomotion. Clearly, there was more than enough time for the evolutionary process to acquire and make good use of the concept of symmetry. Note that not only animal bodies are symmetrical – their brains are symmetrical, too: the two hemispheres are anatomically very similar. So, if the genetic code determining the anatomy of our body and our brain has symmetry built into it, why shouldn't a symmetry constraint be built into the part of the genetic code that determines visual mechanisms (algorithms)? The issue of nature vs. nurture has always been fascinating. It will surely remain a topic for debate for a long time. From time to time in this book, we will consider claims about the innateness of proposed inferential mechanisms, but this endeavor will always be of only secondary importance. It is the mental inferences, themselves, not their origin, that are of primary importance when we discuss problem solving.

Now that we know how the human mind produces veridical visual representations of the environment, we can emulate this ability in robots (Pizlo et al., 2014). This is an ongoing project but some preliminary results show that a robot can produce the same visual representations we humans do. This sets the stage for building bridges between cognitive science and AI. Once a 3D visual representation of the environment is available to a human observer, he can plan visual navigation in, and interact with, the surrounding environment. For example, an observer can choose the shortest path to a goal such as the exit of a room. If the room is empty, a straight line is the shortest path. In the presence of obstacles, the shortest path will be a curve that goes around the obstacles. It is important to realize that the problem of reaching a goal in a 3D environment requires solving an optimization problem in the presence of constraints. It is an optimization problem because there are multiple paths from the start to the goal, and only one of them can be the best. So, solving the shortest path problem resembles the visual inferences discussed in the previous paragraph. In both cases, (i) producing a visual representation of the environment, and (ii) planning the shortest path to the goal in this environment, the human mind solves the constrained optimization problem. So, solving a problem is like making a second-order inference: one inference is built from another inference. In technical parlance, two constrained optimization problems are solved, one after the other. We will see throughout this book that many cases of what we call problem solving can be viewed as tasks involving inferences, and we will often deal with a sequence of two or more inferences.

## 1.6 Problems to Solve

- 1 When you stand in front of a mirror, and you move your right hand, your reflected copy moves her left hand. How does the mirror know to reflect left and right, but not top and bottom? (The answer is in Chapter 2.)
- 2 There are three rooms. One of them contains an expensive item (e.g., a car), whereas the other two rooms contain objects with very low value. You don't know which room contains which item. You are asked to point to one of the rooms. After that, the host, who knows which room contains the car, opens the door of one of the other

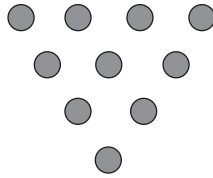


Figure 1.4 Equilateral triangle formed by closely packed pennies.

- rooms without the car. Now you are told to make your final pick: either you can stay with your original choice, or you can switch to the other room. Should you switch in order to increase your chances of getting the car? (After Burkholder, 2012.)
- 3 You start at a point and walk one mile south. Then you turn and walk one mile west. Finally, you turn and walk one mile North. Is it possible that you ended up at your starting point?
  - 4 Glove Selection: There are 20 gloves in a drawer: 5 pairs of black gloves, 3 pairs of brown, and 2 pairs of gray. You select the gloves in the dark and can check them only after a selection has been made. What is the smallest number of gloves you need to select to guarantee getting the following: (a) at least one matching pair; (b) at least one matching pair of each color? (From Levitin & Levitin, 2011, with permission from Oxford Publishing Limited.)
  - 5 Ferrying Soldiers: A detachment of 25 soldiers must cross a wide and deep river with no bridge in sight. They notice two 12-year-old boys playing in a rowboat by the shore. The boat is so tiny, however, that it can only hold two boys or one soldier. How can the soldiers get across the river and leave the boys in joint possession of the boat? How many times does the boat pass from shore to shore in your algorithm? (From Levitin & Levitin, 2011, with permission from Oxford Publishing Limited.)
  - 6 Inverting a Coin Triangle: Consider an equilateral triangle formed by closely packed pennies or other identical coins like the one shown in Figure 1.4. (The centers of the coins are assumed to be at the points of the equilateral triangular lattice.) Flip the triangle upside down in the minimum number of moves if on each move you can slide one coin at a time to its new position. (From Levitin & Levitin, 2011, with permission from Oxford Publishing Limited.)
  - 7 Sorting 5 in 7: There are five items of different weights and a two-pan balance scale with no weights. Order the items in increasing order of their weights, making no more than seven weighings. (From Levitin & Levitin, 2011, with permission from Oxford Publishing Limited.)