In J. W. Houpt & L. M. Blaha (Eds.) (2016), Mathematical Models of Perception and Cognition: A Festschrift for James T. Townsend. Volume 1. New York: Psychology Press. (pp. 223-244).

Symmetry provides a Turing-type test for 3D vision

Zygmunt Pizlo
Department of Psychological Sciences, Purdue University

1. Introduction

The question of whether computers and robots will ever be able to think as well as we do has occupied researchers working in Artificial Intelligence for more than 50 years, ever since Alan Turing (1950) proposed his behavioral test. In this test, a computer tries to imitate a human being's use of natural language in a conversation carried on via a teletype machine back in the 1950s and via emails today. The human participant in this exchange is required to decide whether he is conversing with another human being or with a computer, and if it is hard for the human being to tell which it is, Turing, and some others, have concluded that the computer participating in this discussion is as intelligent as we humans are. This test has been criticized on a number of counts, with the best known criticism coming from Searle (1980). Searle's criticism is elaborated in his "Chinese room" thought experiment, which makes it clear that even if a computer passes the Turing test by manipulating words appropriately, one is not entitled to conclude that the computer understands what is being said. Searle's criticism is clearly valid because similar behaviors, in themselves, need not imply that similar underlying mechanisms are governing these behaviors. One should note that Searle's (1980) criticism of Turing's test is not very different from Chomsky's (1959) criticism of Skinner's behavioristic theory of language. Today, virtually everyone acknowledges that Skinner was obviously wrong, and one should find it easy to appreciate that Turing was, too.

Despite this impasse, some progress has been made towards understanding the relationship between observable behavior and the unobservable mental states that underlie perception. Signal Detection Theory (SDT) was devised precisely to allow us to infer the observer's percept without confounding it with other, non-perceptual aspects of his mental states, but signal detection measures do not go nearly far enough. They do not tell us how the percept is represented in the observer's mind. They only tell us what a subject can discriminate. For example, these methods only allow us to verify whether two subjects' abilities to discriminate between two wavelengths of light are the same, but they do not allow us to find out whether the subjects' subjective experiences of color, say pure green, are actually the same. If SDT cannot do this with two subjects, there is no way it will be able to allow us to compare a human subject's percepts with a computer's. So, an experiment using signal detection methodology, despite its obvious advantages over "classical" psychophysical methods cannot be, and never has been, able to serve as a Turing-type test in perception. The present paper describes a new way of explaining percepts, a way that will actually tell us how the percept is represented in the observer's mind, or in a computer model of it. This will be explained with examples taken from the perception of 3D shapes. It is unlikely that this approach can be extended to perceptual properties other than 3D shape, but understanding why it will be hard to do this may help us develop new, perhaps, analogous, ways of studying problem solving, language and thinking. The essence of my new approach to a valid Turing-type test for perception is that it must be based on fundamental principles used to explain phenomena in the "hard" natural sciences, such as physics, rather than being based on, and confined to, observable behavior as has been done till now. What are these fundamental physical principles and how will they be used?

2. How physicists explain natural phenomena

Natural phenomena in physics are said to always be invariant under some transformations. For example, when we test Newton's second law in two different places, say in two different labs in the same building, we expect to get the same results. We do, which means that Newton's laws are invariant, or symmetric with respect to a translation in space. Similarly, the outcome of an experiment is expected to be the same when the experiment is performed at two different times in the same laboratory, say today and tomorrow. When it is the same this means that the law of nature is also invariant, or symmetric, with respect to a translation in time. The same is true with a rotation in space. These facts can be stated succinctly, namely, *the result of a transformed experiment is the same as the transformed result of the experiment* (Rosen and Freundlich, 1978). To illustrate, take the example of a translation in time. When the outcome of an experiment, which was performed for the first time today, is examined tomorrow, that is, when its outcome was translated in time from today to tomorrow, the outcome will be the same as the outcome of an identical experiment performed tomorrow, that is, when the experiment was translated in time because it was performed tomorrow, not today. So, for a natural law N and a transformation, or mapping M, the following equivalence takes place: NM = MN. Wigner (1967) pointed out that without such invariances (symmetries) there actually would be no such thing as science. This observation is obviously not only confined to physics. It applies equally well in biology and psychology (Narens, 2002).

The symmetry of natural laws is only one of three fundamental characteristics that are used to explain phenomena in physics. The least-action principle is the second fundamental principle. This principle states that phenomena in physics can be formulated in such a way that the cause-effect relations can be shown to minimize some quantity. More generally, some quantity is stationary (the first derivative of this quantity is zero). The Greeks way back when had some intuition about this least-action principle. In modern times, it was Fermat, then Leibniz, and later Maupertuis, Euler, Lagrange and Hamilton, who provided more and more solid formulations of this principle (see Hildebrandt and Tromba, 1996, for a lucid review). Fermat's principle is one of the most commonly known. It says that the reflection and refraction of light can be explained by assuming that light "chooses" the path that minimizes the total time traveled. Another common example, often used to illustrate an application of a least-action principle, can be found in electrical circuits where Kirchhoff's laws for currents and voltages can be derived by assuming that an electrical circuit minimizes the total amount of heat generated in its resistors.

This way of thinking in physics culminated with Noether's (1918) theorems in which the conservation laws can be derived by applying the least-action principle to the symmetries (invariances) of the natural laws. The conservation laws are the third fundamental characteristic upon which our modern physics is built. For two examples, the conservation of energy results from time symmetry, and the conservation of momentum results from position symmetry. These three fundamental aspects of physics, namely, symmetry, the least-action principle and the conservation laws allow physicists to describe and explain underlying, hidden, cause-effect relations. Furthermore, they allow abstract representations of physical phenomena that are not amenable to direct observation. They provide the basis for the model of atom, the concept of black holes and the big-bang theory. This way of thinking and proceeding will be applied to the study of visual perception in this paper (for a preliminary presentation of this idea see the "Note added in proofs" in Pizlo et al., 2014). Percepts will actually be easier to deal with than atoms, black holes and big-bangs because they are not extremely small, extremely far away, or in the extremely distant past. They will, however, present one special difficulty because percepts reside in the mental, not in the physical world. A hundred years ago this was believed to be an insurmountable obstacle because mental phenomena were thought to be outside the scope of scientific study, but modern psychological and engineering communities have been willing to take mental phenomena seriously for quite a while now, at least since the modern interest in Artificial Intelligence began with Turing's (1950) seminal paper.

3. Importing the Least Action Principle into perception

The suggestion that the least-action principle operates in perception first appeared in Mach's (1886) book on perception. Mach also discussed the ability of the human visual system to detect symmetrical patterns in this book. The important role that symmetry plays in perception will be discussed in detail in the next section. Here, only the status of the least-action principle will be evaluated. Mach was actively involved in the discussion of many new trends in physics that took place at the end of the 19^{th} century. The extent and significance of his interests can be seen in his book on mechanics published in 1883. Mach was an unusually creative and prescient thinker. He actually claimed that the scientific laws are "constructed" by the mind of the scientist, rather than "discovered" and that the role of a scientific theory is to describe experimental data economically, rather than to describe, or to explain, the Laws of Nature. So, if one follows Mach's lead, one can claim that the least-action principle, commonly assumed to operate in the physical world, actually originates in the scientist's mind because the scientist's mind used the least-action principle to represent the data he had on hand as economically as possible. We probably will never know whether this claim is really correct, but what is important here is that Mach was the first to claim that the least-action principle operates in the human mind. He went on to support his claim by providing concrete examples taken from visual perception. He pointed out that when a straight-line segment is projected on the observer's retina, the observer perceives a straight-line segment, rather than a circle. Both, a straight-line segment and a circle are possible interpretations, but the simpler straight line segment is perceived. Mach went even one step further and pointed out that the simpler interpretation is also more likely, setting the stage for the modern equivalence of the simplicity and likelihood principles (e.g., Li & Vitanyi, 1994; Chater, 1996). The Gestalt Psychologists took Mach's simplicity principle very seriously and made this principle the foundation of their Revolution without ever citing Mach (Wertheimer, 1923; Koffka, 1935). Gestalt Psychologists called their simplicity principle "Prägnanz", a concept very close to what Mach meant when he talked about the economy of perceptual representation. The Gestalt Psychologists deserve a lot of credit for emphasizing simplicity to a much greater extent than Mach ever did. They also deserve a lot of credit for taking the next very important step, when, in 1920, they introduced the concept of *isomorphism* between the physiological representation of the physical stimulus in the brain and the stimulus's perceptual representation. Wolfgang Köhler, before formally joining the Gestalt movement, received his psychological training at the University of Berlin between 1907 and 1909 under the guidance of Carl Stumpf. He also received solid training in physics at the same time under the guidance of Max Planck. So, he clearly had the background needed to pick up the least-action principle where Mach left it. By 1920, Mach was dead so Köhler did not have to give Mach credit for introducing this important idea to perception. Köhler (1920) claimed that the brain acts as an electrical circuit whose minimum state represents the Prägnanz of the percept. Köhler knew that the end-state of an electrical circuit always ends with the minimal amount of heat generated by its resistors. He also knew that this least-action principle allows one to derive Kirchhoff's laws described in 1845, so for Köhler the least-action principle can act on two levels, the physiological and the mental. He was, however, not in a position to make any meaningful elaborations of the neurophysiological underpinnings of the least-action principle in the brain simply because the brain was not sufficiently understood back then. He did, unfortunately, try to do this, committing himself to the idea that the brain is a "volume conductor". This idea did not have a long life and it distracted many contemporary neuroscientists who felt that they had to test it.

Sixty-five years after Köhler published his 1920 paper, Poggio et al. (1985) published a plausible physiological model of an electrical circuit that might realize the least-action principle in the brain. These authors were able to elaborate Köhler's ideas considerably because a number of new tools were available to them, including the Theory of Inverse Problems, the Regularization method, Information Theory as well as Rissanen's (1978) elaboration of Kolmogorov complexity that did not exist in the first half of the 20^{th} century. Furthermore, much more was known about neural circuits in the brain in 1985 than was known to Köhler in 1920. Today, all computational models of vision take the form of the least-action principle (Pizlo, 2001, 2008), but not everyone working on them today is willing to make claims about the underlying physiological mechanisms as Köhler did. Despite such widespread skittishness, it seems highly likely that some version of what Köhler (1920) had in mind, and what Poggio et al. (1985) described, is actually going on when the brain produces percepts because visual percepts are produced very quickly. This can only happen when the minimum of some cost function is found by a physiological process resembling the least-action principle known to govern all physical and chemical events. The kind of iterative processes used by computers today cannot act fast enough to pull this off. Once this is appreciated, the brain/percept isomorphism first proposed by Köhler becomes both self-evident and necessary. Other cognitive functions, such as thinking and language, may be different. They may not make use of a physiological least-action principle because they tend to be much slower than perception.

4. Bringing Symmetry into theories of Perception

Mach (1886), as pointed out above, was the first to discuss the role of symmetry in vision. He used 2D mirror, rotational and translational symmetries as examples when he did this. He pointed out that mirror symmetrical configurations are the easiest to see as being self-similar when they were compared to rotational and translational symmetrical configurations. This convinced him that mirror symmetry is special for a human observer when compared to the other types of symmetry. Wertheimer (1923) made the next important contribution to the role of symmetry in visual perception when he used interwoven sine- and square-waves to illustrate what he meant by ”simplicity” with the grouping principle he called “good continuation”. The periodically repeating patterns of sine- and square-waves he used made it very easy to disambiguate the x-intersections of the two waves because of the redundancy introduced by their symmetry. This was important because Wertheimer insisted that the simplicity of a curve is not caused by its geometrical simplicity, for example, a straight line is simpler than a curved line. Its simplicity was produced by something quite different. The example he chose to illustrate this difference implies that the similarity of a pattern to itself (aka symmetry) is a better way to define the simplicity of the pattern than any geometrical measure of the pattern such as its curvature or variation of the curvature. From this point on, symmetry was tested, on and off, by a host of students of human vision, but *only* as a source of redundancy that could make it easier to store and remember symmetrical patterns. This made sense in light of Shannon’s (1948) Information Theory that had been picked up by psychologists who thought that detecting and removing redundancy was the primary task confronting the visual system (see, for example, Attneave, 1954; Barlow, 1961; van der Helm, 2000). This focus on symmetry as redundancy encouraged, actually forced, vision scientists to use stimuli that were symmetrical when they were presented to the observer’s retina. This was unfortunate, as well as unnatural, because natural symmetrical objects are 3D, not 2D, and they almost never produce 2D *symmetrical* retinal images. A 2D retinal image of a 3D symmetrical object is itself symmetrical but only for a very restricted number of viewing directions. Such viewing directions, which are called “*degenerate”*, are so unlikely to occur in our natural life that their probability can be said to be zero.

This was the role symmetry played in visual perception until 2006, when we showed that 3D mirror symmetry is used by the human visual system to recover 3D shapes from a single 2D retinal image (Pizlo, Li and Steinman, 2006). The retinal image of a 3D mirror-symmetrical object is always 2D and usually asymmetrical, so assuming that 3D objects are symmetrical is the *only* effective way for the visual system to recover the 3D shapes of the objects from 2D retinal images. Two important facts must be noted here, namely: (i) natural objects are 3D and many of them are mirror-symmetrical and (ii) we see them as 3D and mirror-symmetrical. These facts imply that a 3D mirror reflection, which defines mirror symmetry, operates as well in the physical world as it operates in the mental world. With this established, we can use the concepts and jargon borrowed from physics introduced in section 3 to elaborate this story. A 3D mirror-symmetrical object does not change under reflection. This means that the object is invariant (symmetric) under 3D reflection. The same is true with the perceptual representation of a 3D mirror-symmetrical object. Specifically, we see symmetrical objects as symmetrical, which means that a mirror reflection of the perceptual representation of a 3D mirror-symmetrical object remains mirror-symmetrical. Now, let M represent a 3D reflection of a physical object, or of the object's perceptual representation. Let N represent a psycho-physical natural law describing how the 3D physical world is transformed into its perceptual representation in the mind of the observer. What has just been said about the symmetry of objects and the symmetry of percepts can be expressed as follows: NM = MN, which means that the *percept of a 3D reflected symmetrical object is the same as a 3D reflected percept of the original object* (see Fig. 1). This statement is intuitively obvious, but it is far from trivial. Note that this statement would not be true if we did not perceive symmetrical objects as symmetrical. This claim will be illustrated by using Marr's theory of vision because his theory, as all prior theories of human vision, ignored symmetry completely. Assume that the percept is like Marr's 2.5D representation , namely the observer's percept represents *only* the visible surfaces of a 3D opaque object. So, according to Marr, when the right half of a horse's body is facing an observer, the observer can only perceive the right half of the horse's body, and after a 3D reflection of this percept, the right half becomes the *left half of the horse*. This would be the result after applying the transformation "MN" to the 3D horse, that is, MN (horse) = left half of the horse. Now, if we first reflect the 3D horse, we get the same horse, but the percept now, according to Marr, is the *right half of the horse*. This is the result of applying the transformation "NM" to the horse, that is, NM (horse) = right half of the horse. Clearly, the right half of a horse is not the same as its left half [1], so, NM ≠ MN in Marr's theory. This means that Marr's theory stands no chance, whatsoever, of becoming a law of nature, so it is neither surprising nor disappointing, that Marr's theory has largely been forgotten. Symmetry, simply put, cannot be ignored.

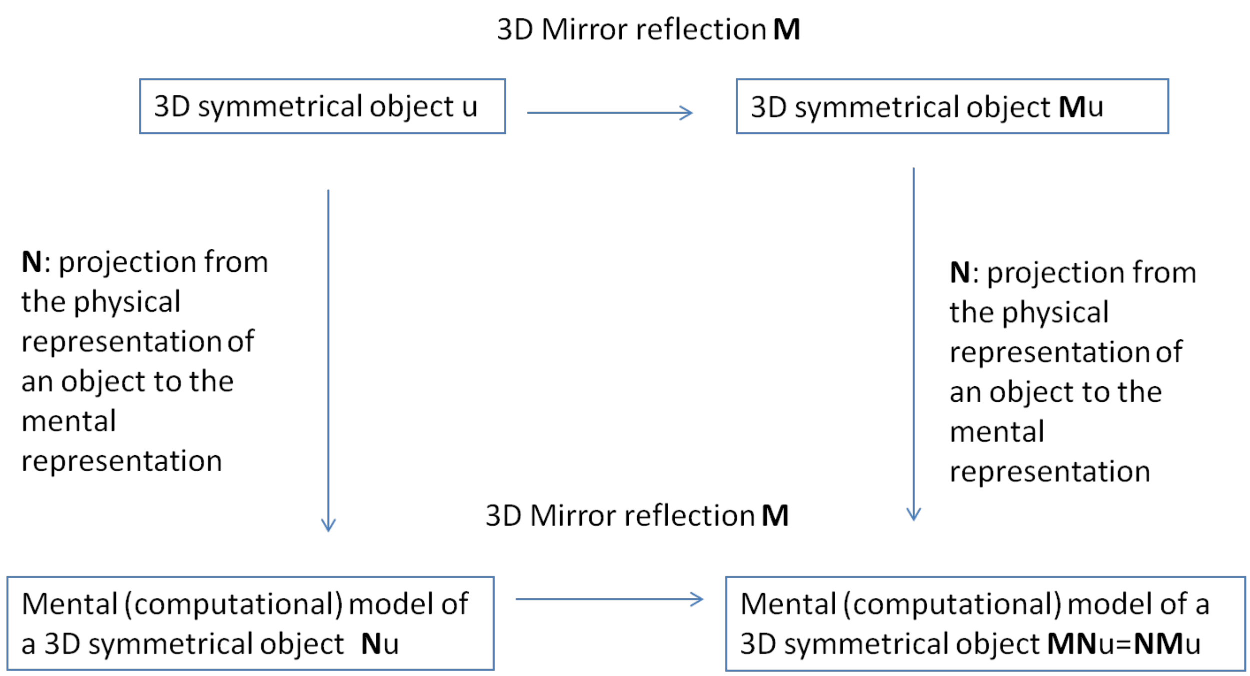

Figure 1. **N** is our natural law for mapping a 3D mirror-symmetrical object, u, to its mental representation. **M** is a 3D reflection.

[1] The right half is identical to the left half after a 3D reflection, but this additional reflection is not allowed here because it would lead to the comparison between the transformation MNM and the transformation MN, which does not characterize the symmetry of a natural law.

With this made clear, our discussion returns to the symmetry of objects and the symmetry of percepts. As long as we see symmetrical objects as symmetrical, the relation MN = NM is satisfied for the psycho-physical law describing the perception of 3D objects. This statement is very important because it is identical to the statement used in Section 2, where we described what it means for a law of nature to be symmetric (invariant) with respect to a transformation: "*the result of a transformed experiment is the same as the transformed result of the experiment"* (Fig. 1).

To summarize: in the previous section (see Section 3), we established that the natural law for the perception of 3D objects is governed by the least-action principle. In this section, we established that it is characterized by invariance (symmetry) in the presence of 3D reflections. A cost function that represents the least-action principle will be described in the next section where it will be shown that when this cost function is applied to the 2D retinal image of a 3D symmetrical object, the 3D shape is "conserved". This psycho-physical conservation is called "veridical" in shape perception. This way of deriving the conservation of shape through the application of a minimum principle to object's symmetry is analogous to the manner in which Noether's theorem established a one-to-one correspondence between the symmetries of natural phenomena and the conservation laws. The fact that *when a 3D object undergoes a transformation between the physical world and its mental representation, its 3D shape remains unchanged,* provides the basis for comparing the perceptual representations of two different human observers or the perceptual representations of a robot and a human observer.

5. Veridicality of 3D shape perception seen as a conservation law

Assume that we are given a 2D camera image of a set of N pairs of 3D points that form a mirror-symmetrical configuration. Call this set of 3D points a "3D object". Real objects, such as chairs, horses and bell peppers, are *not* sets of points. In fact, they cannot be described adequately by points. We use points, despite this here, simply because this makes it easy to explain how our cost function leads to the veridical perception of a 3D shape. In prior work, we showed how the cost functions used could recover the shapes of a wide variety of natural objects, including birds, praying mantises, spiders and jeeps (Pizlo et al., 2014). When we say that a set of 3D points is mirror-symmetrical, we mean that there is a plane that bisects the line segments connecting the individual symmetrical pairs of points. This plane is called the "plane of symmetry" and the lines connecting pairs of mirror-symmetrical points are called "symmetry lines". These symmetry lines are all parallel to each other and they are orthogonal to the plane of symmetry. In the 2D camera image, the images of the symmetry lines intersect at a single point called the "vanishing point" (Sawada et al., 2015). The line connecting the center of a perspective projection of the camera with the vanishing point on the 2D camera's image is parallel to the 3D symmetry lines (Sawada, 2011). It follows that the position of the vanishing point in the 2D camera image, uniquely determines the 3D orientation of the symmetry plane relative to the camera, but not its distance from the camera. Once we determine which pairs of points in the 2D camera image are projections of the pairs of symmetrical points in 3D, the *symmetry correspondence* in the image has been established (Sawada et al., 2014). Once symmetry correspondence is known, a single 2D camera image determines a *unique* 3D symmetrical interpretation of the 3D points up to some unknown overall size. This unknown *size* means that the 2D image could have been produced by a small object close to the camera, or by a large object far away. But the *shape* of the recovered set of 3D points is determined uniquely. By shape, here, we mean the ratios of pairwise distances and the sizes of angles. According to this conventional definition, the shape of an object does not change when the object is rigidly translated, rotated or its size is changed uniformly. Obviously, the shape of a symmetrical object will not change after a 3D reflection, either. So, by applying a *single a priori* constraint, symmetry, which *only* says that the 3D set of points is symmetrical, allows us to recover the 3D shape from a single 2D image. The recovered 3D shape will be veridical, which simply means that the recovered 3D shape is identical to the shape of the 3D set of points "out there". This straightforward geometrical fact is nothing short of remarkable because it provides a solution to what has been considered for dozens, if not hundreds of years, the most difficult problem in human vision, namely, producing 3D percepts from 2D retinal images (Pizlo, 2008). We will now use this geometrical fact to explain the relation among a symmetry *a priori* constraint, the least-action principle and veridical shape perception as it is used in 3D vision.[2]

Let the matrix $\mathbf{X}_{4x2N}$ represent our 3D mirror-symmetrical object. The columns of **X** represent individual 3D points. We have 2N columns that are N pairs of symmetrical points. The first three rows of **X** represent X, Y and Z, Euclidean coordinates of the points. The last row is a set of "1"s. In this way, our matrix **X** is defined by homogeneous coordinates. Homogeneous coordinates are useful because they allow representing a perspective projection, which is a non-linear transformation, by using matrix multiplication, which is a linear operation. Let the matrix $\mathbf{x}_{3x2N}$ represent the camera image of **X**. Again, we use homogeneous coordinates, so each column of **x** is a set of three numbers: x*, y* and w. The usual Euclidean coordinates on the image are obtained by taking ratios: x = x*/w, y = y*/w. Finally, let $\mathbf{A}_{3x4}$ represent a perspective projection that maps the 3D space to the 2D retinal or camera image. So, the relation between the 3D object **X** and its camera image **x** is expressed as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{X} \qquad (1)$$

If **A** were a square matrix whose inverse $\mathbf{A}^{-1}$ existed, the recovery of the 3D object **X** would have been trivial: it would have been produced by pre-multiplying both sides of equation (1) by $\mathbf{A}^{-1}$. But in our case, the inverse $\mathbf{A}^{-1}$ does not exist, which means that without any *a priori* constraints, for a given camera image **x,** there are infinitely many **X**s that satisfy equation (1). Now, we define two standard terms of a cost function. The first term, E_1, evaluates how well the recovered object $\hat{X}$ fits the given 2D camera (or retinal) image **x**:

$$E_1 = \sum_{i=1}^{2N} [(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2] \qquad (2)$$

where the 2D image $\hat{x}$ is produced by the recovered 3D object $\hat{X}$ using equation (1).

The second term, E_2, evaluates the degree of asymmetry of the recovered object $\hat{X}$. Let $L_{2(i-1)+1}$ and L_{2i} be the distances of the corresponding 3D points of the recovered object $\hat{X}$ from the symmetry plane. If the recovered 3D object is perfectly symmetrical, these two distances are equal. But remember that the family of possible 3D interpretations includes many objects that are not symmetrical, so our cost function will evaluate the asymmetry of each of them. The simplest way to evaluate asymmetry is as follows[3]:

[2] Many other theories of 3D shape perception have been proposed during the long history of research on this important subject. A universal characteristic of all these other theories was that they minimized the role of priors, if they included them at all. Symmetry was completely left out, so it is not surprising that these theories could not explain veridical vision. The authors of these theories felt that this is just fine because all existing empirical data showed that vision was far from veridical. To the best of the present author's knowledge all of these experiments on 3D shape perception were flawed because they used impoverished stimuli and degenerate viewing conditions. Furthermore, symmetrical objects, which are found *everywhere* in our natural environment were completely excluded in their laboratory experiments. The rest is history.

[3] Equation (3) is a necessary condition for mirror symmetry. The additional constraint that is needed for a sufficient condition is that the line segment connecting the two symmetrical points is the normal of the symmetry plane.

$$E_2 = \sum_{i=1}^{N} (L_{2(i-1)+1} - L_{2i})^2 \tag{3}$$

Note that if instead of discrete points, we are dealing with curves, the summations in equations (2) and (3) are substituted by integrals. The cost function E is a weighted sum of the two terms:

$$E = E_1 + \lambda E_2 \tag{4}$$

The 3D object $\tilde{X}$ that minimizes E is the best 3D recovery. Note that the parameter λ in equation (4) controls the relative importance of the symmetry *a priori* constraint and the 2D image. Typically, when the 2D image is unreliable, λ is set to be large. Otherwise, λ is small. Also note that if X is perfectly mirror-symmetrical, the two terms E_1 and E_2 are invariant to a 3D reflection of the object, simply because the object, itself, is invariant to 3D reflection. This is analogous to the properties of a Lagrangian that is used in least-action principles as employed in physics.

The cost function E closely resembles what physicists call “action”, which is an integral of a Lagrangian. In this way, finding the object that minimizes E is completely analogous to how the least-action principle operates in physics. If there is no noise in the 2D image and if the 3D object **X** in front of the camera, is perfectly symmetrical, minimizing E will produce a 3D recovery whose shape will be *identical* to the shape of **X**. This is guaranteed by the theorem described at the beginning of this section. If this recovery is done by an observer’s visual system, rather than by a camera, the observer’s percept of the 3D shape will be called “veridical”. This is what we mean by our psycho-physical shape conservation law. It is produced by applying our least-action principle to a 3D object that is mirror-symmetrical.

To summarize, when an observer looks at a 3D object, his mental representation cannot be a copy of this object simply because a mental representation does not have physical characteristics like mass, albedo or stiffness. But, we have been able to show that the mental representation of the 3D shape of the object is a copy of the *actual* 3D shape of this object. Its shape did not change during the transition between the physical and the mental representations. The object’s shape was conserved. It follows that the perceptual representations in the mind of two people looking at the same 3D shape will be identical. Also, if a robot is given appropriate visual equipment, namely two cameras that acquire 2D images of a 3D scene, and if the robot uses a shape recovery algorithm based on our least-action principle, the robot’s perceptual representation of 3D shapes will be the same as ours.

We can conclude that the perception of shape is a Law of Nature in the same way that the Laws of Physics are Laws of Nature because shape perception is governed by the same three fundamental characteristics: symmetry, a least-action principle and conservation. This is not to say that percepts are physical phenomena. They obviously are not, but we can say that percepts have the same scientific status as the physical phenomena believed by everyone to be real. This means that there is no reason to assume any fundamental difference between the shape perception of humans and robots with visual systems designed to emulate them. Shape conservation can provide a Turing-type test for determining whether a robot can see as well as we do. The next section describes how such tests were performed with our robot, called “Čapek”.

6. Empirical tests verifying that Čapek sees as we do

In the same year that Köhler was describing his psycho-physical least-action principle, Karel Čapek, a Czech writer, coined the term “robot” in his 1920 play “*Rosumovi Univerzální Roboti*”, which translates into English as ”Intelligent Universal Robots”. We named our first robot Čapek to honor this writer.

Čapek arrived in our lab in the Spring of 2010. By then, we had already worked out our algorithms for the 3D recovery of shapes based on the cost function described in the previous section, so the time was right to start comparing Čapek's 3D shape perception to ours. The time had come to run our Turing-type tests. At this point the reader will do well to look at animations illustrating several examples of these recoveries (go to http://shapebook.psych.purdue.edu/1.2/ and http://shapebook.psych.purdue.edu/2.4/demo_skeleton1.html. When you do this, you will find that the 3D shapes recovered by Čapek are identical to the shapes you see. We have published formal comparisons of Čapek's and human performance in several papers. For example, Li et al. (2011) compared the results of 4 human subjects with Čapek's (our computational model) when they recovered 3D *shapes* from 2D images, monocularly and binocularly. We have also compared human and our robot's performance in the recovery of 3D *scenes.* In these experiments, several pieces of children's furniture were placed 2m - 5m away from our observers, humans and our robot. The ground truth about the positions and sizes of objects was established by using a *PhaseSpace* optical system with accuracy better than 2 cm, and precision 2 mm to locate the furniture within the experimental chamber. Čapek's and our human observers' recoveries of 3D scenes was compared to this ground truth, as well as to each other. Figure 2 is an example of one of Čapek's recoveries.

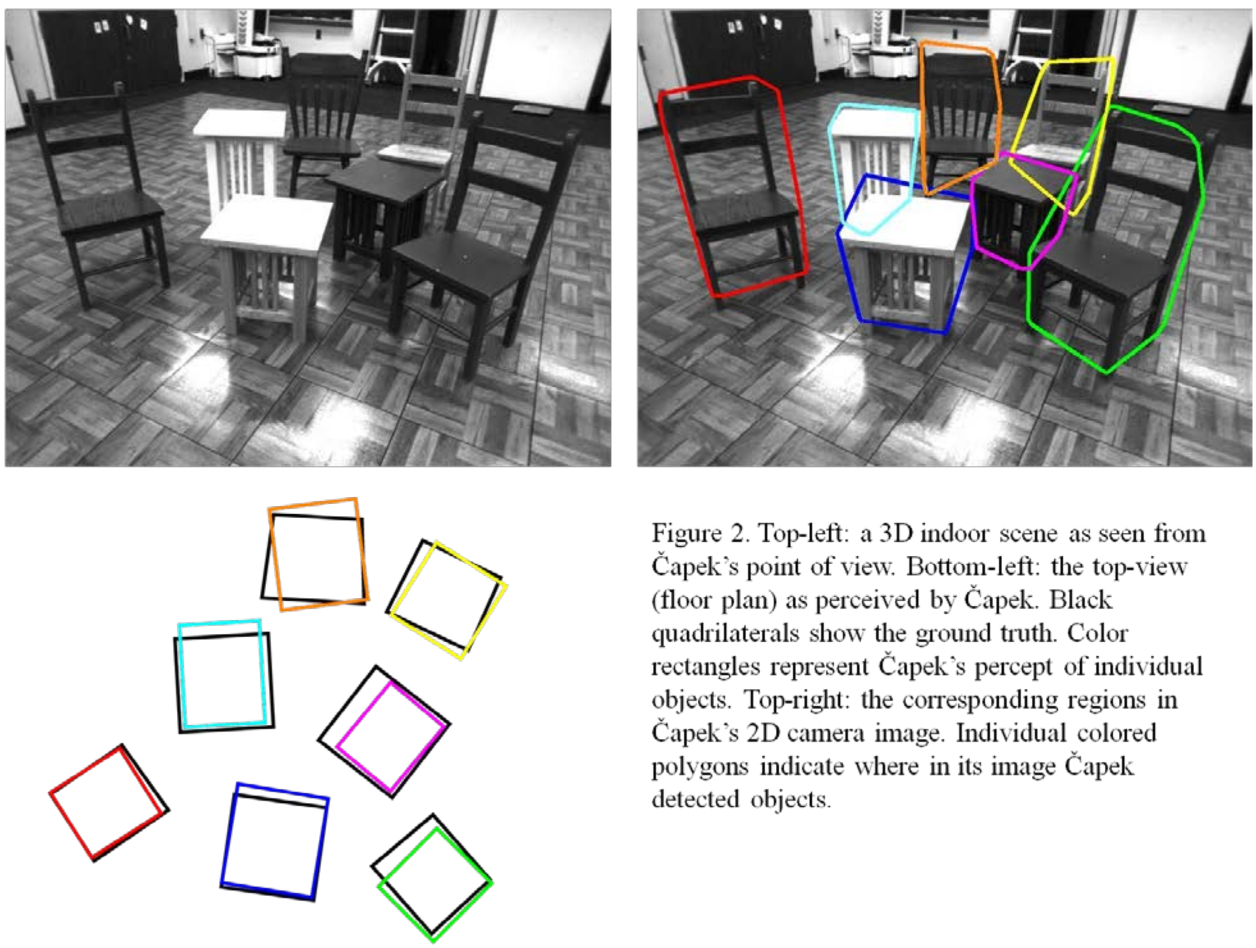

Figure 2. Top-left: a 3D indoor scene as seen from Čapek's point of view. Bottom-left: the top-view (floor plan) as perceived by Čapek. Black quadrilaterals show the ground truth. Color rectangles represent Čapek's percept of individual objects. Top-right: the corresponding regions in Čapek's 2D camera image. Individual colored polygons indicate where in its image Čapek detected objects.

The precision of Čapek's recoveries of the pairwise distances of the objects was between 7% and 10%. The average precision of each of the 3 human subjects, who were tested with similar scenes, was 7%, 8%

and 10%.[4] All 3 of the human's results fell within Čapek's range, which is a very good fit, indeed (Kwon et al., submitted). One simply could not know whether you were looking at what Čapek was seeing, or what one of our human subjects was seeing, by looking at the data they produced. The accuracy (absence of systematic errors) of both the human subjects and of Čapek was also tested in a task in which both were required to construct a right isosceles triangle by positioning 3 pieces of furniture within the chamber. Both accomplished this task with very small systematic errors. Their accuracies as geometers was also good. The angles of the right isosceles triangles they constructed were good enough to satisfy Euclid himself (the average right angles constructed by the 3 subjects were, 87°, 89.8° and 91.6°. The average right angle constructed by Čapek was 88.5°. These results show two important things; first, both the human's and the robot's 3D shape and 3D space perception were virtually identical, and second, both the human's and Čapek's performance was veridical: both saw the physical world as it really is. So, we have not only developed a Turing-type test of 3D vision, which clearly showed that both humans and a machine vision system can see the world the same way and accurately, but we also justified the methodology underlying our success by using a formalism taken from physics. Simply put, we *explained* why our test actually worked. The extent to which this test can be generalized to other properties of visual percepts will be discussed in the concluding section. Do not expect to see very many visual applications because the property called "shape" in visual perception is *unique* because of its complexity (Pizlo, 2008, Pizlo et al., 2014). Shape by being both complex and symmetrical is the *only* communication channel that provides us with accurate information about our external world. We, as humans, would probably not even be here as we are if shape did not permit us to see our world veridically.

7. Generality and implications of our test

Our derivation of the veridicality of 3D shape and 3D space perception as a form of a conservation law was based on two critical facts, namely, the 3D objects are symmetrical and the visual system recovers their shapes by applying an *a priori* simplicity constraint to the sensory data provided by a 2D retinal image. It follows that the approach described in this paper: (i) cannot be generalized to percepts for which symmetry cannot be defined, and (ii) cannot be used if the visual system does not apply a least-action (simplicity) principle. These two criteria exclude a large fraction of possible visual percepts, perhaps all but a very few. Consider the difficulty inherent in trying to use our approach for the perception of the hues of light, say green vs. blue vs. red vs. yellow. These perceptions (appearances) of hues rest entirely on sensory coding. There is no reason to believe that the visual system uses anything like a least-action principle to determine how a hue will be perceived. Now consider symmetry, the *sine qua non*, for shape. Defining the symmetry for any given physical stimulus and the percept associated with it require that the stimulus and the percept are sufficiently complex. For example, it must be possible to define two halves of the percept, and then verify whether they are similar. Obviously, with hues this will not work. A hue is represented by a single point on the boundary of the chromaticity diagram. There is no way to split this kind of point into two. Shape, unlike hue, has all the required properties; it is very complex both geometrically and perceptually. In the limit, one needs an infinite number of dimensions to describe the shape of an object. In the real world, it has been shown to have 400 or more dimensions (Pizlo, 2008). This is very different from color, size, lightness or speed. These perceptual characteristics are at most three-dimensional. So far, we have been able to extend our theory of the veridical perception of 3D shapes to the veridical perception of 3D scenes. We were able to do this because 3D shapes reside in a 3D space, and simply by assuming gravity and a common ground plane, the symmetry of 3D shape can be used to "calibrate" the Euclidean 3D space. This seems to be about as far as we can go in vision. The application of symmetry, a least-action principle and conservation may actually end right here. We know of nothing at this time outside of 3D shape and 3D space perception that can benefit from the principled approach described in this paper, an approach that is based on how explanation is done today in the "hard sciences". Note that when we say that we cannot do more than what we have done with shape and space

[4] The human subjects viewed binocularly and Čapek was fitted with a pair of cameras.

so far, we are actually agreeing with David Marr's definition of the *goal of vision science*, as stated in the introductory part of his book (Marr, 1982, p. 36). Here, Marr claimed that explaining the perception of shape and space is the central task for vision science. Other visual percepts, such as color, texture or motion are only secondary. So, we agree wholeheartedly with Marr's stated goal, but we disagree completely with how he tried to achieve it (see Section 4, above).

Having reached what seemed to be an insoluble impasse in visual science we were encouraged to ask whether there are phenomena in cognition, outside vision, that has a sufficient degree of complexity to allow symmetry to operate? There are if we accept Shepard's (1981) claim that all cognitive functions have spatial representations because they were formed, during our long evolutionary adaptation, on top of spatial representations already existing in vision. Consider three unambiguous examples. Here, the Traveling Salesman Problem (TSP) stands out. This problem requires a human being to find the shortest possible tour among a number of cities. TSP is called "intractable" because finding the shortest tour often requires prohibitively long solution times. Fifteen years ago we showed that human subjects produce near-optimal tours very quickly (Graham et al., 2000). We know that these solutions are so good (fast or nearly optimal) because the subjects used vision to solve the TSP. We also know that the brain uses a self-similar pyramid structure in its visual system to represent the problem on many levels of scale and resolution (Pizlo et al., 2006; Pizlo & Stefanov, 2013). One can say that the visual system can impose a self-similar structure on the problem, a structure that the problem, itself, did not have. Once this is done, the visual system uses self-similar (symmetrical) operations to solve the problem. Pyramids, which are hierarchical representations, are becoming widely used across many different types of problem solving, as well as in theories of human and machine learning. This is not surprising because of how important clustering, or chunking, is in cognitive operations and how common hierarchical relations are found throughout nature. Taxonomy, biological classification, is probably the best example.

The second example that will be discussed here deals with how problems are solved in physics. Good physics students, are said to be good, because they solve problems well, but how they do this is not understood. Understanding a problem stated verbally and then solving it, is difficult enough to have resisted, so far, many attempts to write computer algorithms that can solve these problems as well as good students do. The absence of a working theory of how to solve physics problems should not be surprising because there are no educational approaches that tell us how to teach new physics students how to solve problems. Few seem to be concerned with this educational issue perhaps because everyone knows that solving a new problem requires creative thinking (aka insight), so it is expected to be very difficult, and no one is surprised when this expectation is met. But: (i) once we know that physical phenomena are governed by symmetry (invariance), a least-action principle and conservation laws, and (ii) that the human mind makes use of these fundamental characteristics, it may be possible to teach students how to look for symmetries within the verbal statement of a physics problem and then infer the corresponding conservation and least-action principle. Take a simple example: "the 12 edges of the cube each contain a 1Ω resistor, and the task is to calculate what the equivalent resistance is between two opposing corners (see Figure 3)." This problem is not easy if one uses straight circuit analysis, but it becomes almost trivial, if we recognize and make use of the symmetry of the cube's configuration. The gist of this solution is as follows: (for more details see Rosen, 2008, pp.114-119; or the following site: http://www.rfcafe.com/miscellany/factoids/kirts-cogitations-256.htm). The cube is rotationally symmetrical around the line connecting two opposite corners. Rotating the cube by 120° around this line maps the cube onto itself. The same is true with a rotation by 240°. This means that the current "I" entering the corner of the cube marked by A, must be distributed equally between the three output branches (I/3 in each branch). The same must be true with the three currents flowing into corner B (I/3 in each branch). The remaining 6 branches form three pairs, each pair having input current I/3. Again, because of symmetry of the cube, the current in each of these 6 branches is half of I/3, which is I/6. Using Ohm's law, the voltage across each resistor is the resistance (here 1Ω) times the current through this resistor. Taking any of the three resistors connecting A and B, gives us total voltage 1/3·I + 1/6·I + 1/3·I = 5/6·I, between the corners A and B. The equivalent resistance between A and B is the total voltage (here 5/6·I) divided by the current entering the corner marked by A (here I). So, the equivalent resistance is 5/6Ω. We think that the symmetry of natural phenomena, together with the least-action principle and conservation laws, can actually provide a common denominator for solving *all* physics problems, including word problems. If this works out, it might even lead to the application of our approach to more general understanding of much less restricted kinds of verbal communication.

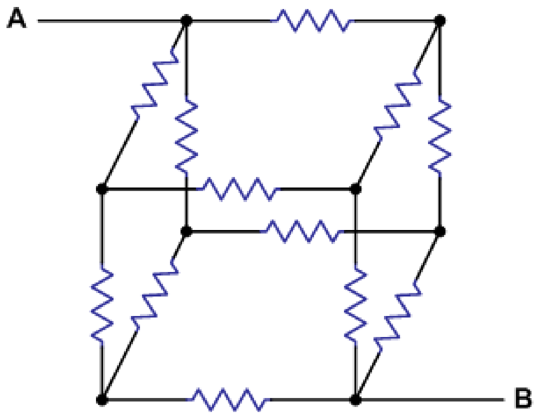

Figure 3. Each edge of the cube is a 1Ω resistor. Find the equivalent resistance between corners A and B.

Here, it must be pointed out that *noticing* the 3D rotational symmetry of the cube consisting of resistors (just above) actually *changed the representation of the problem*. We started with 12 resistors, some of which are in series, while others are in parallel, forming 6 loops, each loop sharing a resistor with another loop. After 3D symmetry was "imposed" on this set of resistors, the entire cube was split differently by using subsets of resistors that represented the *rotational* symmetry of the cube's configuration. This made solving the problem easy. The fact that changing the representation of a problem can make a seemingly insoluble problem easy is not news. The Gestalt psychologists (*viz.*, Duncker, 1945; Wertheimer, 1945) made a big deal about this 70 years ago by pointing out that "Productive Thinking", which required insight, could do just this. In one of their examples, they asked the subject to construct 4 identical triangles, using only 6 wooden matches. This problem cannot be solved on a flat surface. Everyone starts by trying to do it this way and fails. The solution is actually simple once the problem is *visualized* in a 3D representation. Changing the representation from 2D to 3D was all that had to be done to solve what seemed, in 2D, to be a impossible problem. One might be tempted to believe that it is not a complete coincidence that the 4 triangles in this 3D configuration form a regular tetrahedron, one of the Platonic solids that are characterized by multiple symmetries. It is important to note that the representation of a problem can be changed when the type of symmetry already present in the problem is changed, rather than when symmetry is imposed on a problem that did not have any symmetry. The latter is the kind of problem just described above. We changed the type of symmetry with good effect recently when we explained how the visual system detects closed contours (Kwon et al., submitted). The simplest closed contour is a circle, so it is not surprising that all prior approaches to the detection of closed contours started by trying to find circle-like curves on the retina. This could not work because the visual system analyzes contours in V1 in the cortex not on the retina. The retina only detects center-surround organizations. It turns out that a circle on the retina is transformed into a straight line in V1, so a rotational symmetry is transformed into translational symmetry when information is moved up in the visual system. This change of representation greatly simplifies the problem of detecting closed contours because detecting a smooth open line in V1 can be done by finding the least-cost path. This discovery shows that the kind of symmetry used in visual processing is important when the least-action principle is applied.

The third example considers how our approach, which is based on the way phenomena are explained in the hard sciences, now including 3D vision, might be extended into less restricted forms of linguistic communication. The 3 texts just below show how environments (3D scenes) were described in a hundred year period during which the modern novel underwent a host of changes. These three authors contributed a lot to changing our novel's style but the way they described scenes changed very little as they did this. Each text consists of 128 words. It was selected haphazardly from sections in each book with only one constraint, namely, it did not contain any dialogue.

Flaubert *Madame Bovary 1856*

"It was a substantial-looking farm. In the stables, over the top of the open doors, one could see great cart-horses quietly feeding from new racks. Right along the outbuildings extended a large dunghill, from which manure liquid oozed, while amidst fowls and turkeys, five or six peacocks, a luxury in Chauchois farmyards, were foraging on the top of it. The sheepfold was long, the barn high, with walls smooth as your hand. Under the cart-shed were two large carts and four ploughs, with their whips, shafts and harnesses complete, whose fleeces of blue wool were getting soiled by the fine dust that fell from the granaries. The courtyard sloped upwards, planted with trees set out symmetrically, and the chattering noise of a flock of geese was heard near the pond."

Proust *Swan's Way 1913*

"At one side of her bed stood a big yellow chest-of-drawers of lemon-wood, and a table which served at once as pharmacy and as high altar, on which, beneath a statue of Our Lady and a bottle of Vichy-Célestins, might be found her service-books and her medical prescriptions, everything that she needed for the performance, in bed, of her duties to soul and body, to keep the proper times for pepsin and for vespers. On the other side her bed was bounded by the window: she had the street beneath her eyes, and would read in it from morning to night to divert the tedium of her life, like a Persian prince, the daily but immemorial chronicles of Combray, which she would discuss in detail afterwards with Françoise."

Hemingway *Old Man of the Sea 1952*

"They walked up the road together to the old man's shack and went in through its open door. The old man leaned the mast with its wrapped sail against the wall and the boy put the box and the other gear beside it. The mast was nearly as long as the one room of the shack. The shack was made of the tough bud shields of the royal palm which are called guano and in it there was a bed, a table, one chair, and a place on the dirt floor to cook with charcoal. On the brown walls of the flattened, overlapping leaves of the sturdy fibered guano there was a picture in color of the Sacred Heart of Jesus and another of the Virgin of Cobre."

The 3D scenes described *verbally* in these three passages should allow all literate readers to *visualize* what each of the authors had in mind when these texts were written down. Why? It seems most likely that the author saw a 3D scene, rather like the one described, represented it visually in his mind, and then depicted it using words. The author chose his words to make it possible for the reader to recover the 3D scene for himself. The same could be done by an artist, who produced a 2D painting, or a drawing, of a 3D scene. We all would expect that the end results would be very similar. There is a large difference in these two types of communication because the painting would be a 2-dimensional image, whereas written, or spoken, sentences would have only a single dimension – each new word follows the previous one. There is another important difference. The linguistic depiction, unlike scenes represented by images, relies heavily on the reader's familiarity with the names of objects, such as "barn", "bed", "table", or "horse", the meanings of prepositions, such as "on", "under" or "between" as well as the meaning of verbs. With images, the viewer may or may not be familiar with some objects represented, but would still be able to recover their 3D shapes and would also be able to recover their 3D spatial relations. So, how is familiarity exploited in linguistic communication? Familiarity with the objects and with their names conveys information to the reader about their symmetry. Note that horses, and almost all other animals, are mirror-symmetrical with the planes of symmetry almost always parallel to the direction of gravity. Beds, barns and tables usually have rectangular bases. This means that they have at least two planes of symmetry, both vertical. Rooms are almost universally rectangular with 3 planes of symmetry. Walls are impenetrable, so a bed is never placed in two rooms at the same time. Objects are also impenetrable, which means that they can be next to each other, possibly touching or on top of one another and certainly touching because of the operation of gravity. A door is almost always rectangular and it swings around a vertical axis. The bottom of the door is usually next to the floor, while the bottom of a window is not. All objects are three-dimensional, and all, except for birds and airplanes, reside on the common horizontal ground because of gravity. Objects that are included in a linguistic description are always also included in the reader's mental representation. Not much more is included. By way of illustration, an upright piano might have been standing in the corner of the bedroom in one of the stories, or just outside the barn in the other, but if the piano was not included in the linguistic description, the reader will assume that the piano was not actually in the scene described. A reader simply does not add or multiply objects, or other things in his representation without a good reason. Instead, he uses a version of *Occam's Razor*, that will be called a *simplicity principle* if he happens to be a Gestalt Psychologist or one of their disciples. So, linguistic communication is facilitated by capitalizing on several of the natural regularities that are often prominent within our natural physical environments, namely: (i) the symmetries of familiar objects and the relationships among them, (ii) their three-dimensionality, (iii) their impenetrability, (iv) the direction of gravity and (v) the orientation of the horizontal ground, which is usually orthogonal to the direction of gravity. These regularities are governed by an overarching simplicity principle. Noticing these regularities allows us to recover 3D scenes through a slow, and often painfully conscious effort, that can characterize reading, or listening to linguistic descriptions. The visual system has a much easier job and it can produce the same result by recovering 3D shapes and 3D scenes from a couple of 2D retinal images. It can do this automatically, almost instantaneously, and without any effort. This is obviously not the entire story because linguistic communication can do much more than convey information about 3D shapes and scenes. At this point we suspect that most of our readers recognize that when language is used to describe, and to manipulate concepts outside of the geometry of 3D space, the concept called "symmetry" will be generalized to the concept called "analogy". Analogy refers to the similarity, or identity, of one concept to another. It is used in this way in all aspects of cognition, including decision-making, memory, creative thinking, emotion, explanation, and problem solving, as well as in the perception of properties other than 3D shape and scenes. Instantiations of analogy include exemplification, comparison, metaphor, simile, allegory and parable. Analogy is the key to understanding proverbs and the idioms used in ordinary language, but it is also used in science and philosophy when concepts such as association, comparison and correspondence are employed. What is being said here is that new insights into language communication and its role in cognition may be obtained through formalizing "analogy" by using the tools that have been developed for working with symmetry. Once this is done, we may be able to extend what we accomplished with 3D shapes and 3D scenes, where we explained how human beings form their veridical visual representations to other aspects of cognition, particularly to language, where we hope to be able to show how human beings form their veridical representations of somebody else's thoughts. This will set the stage for generalizing our Turing-type test for shapes and scenes to a Turing-type test for language and thinking.

The author thanks Dr. Robert M. Steinman for suggestions that influenced this paper as well as Dr. Acacio de Barros for very useful discussions and comments. The author is currently supported by a grant from the National Eye Institute (1R01EY024666-01). The theoretical and empirical work of the author and his collaborators, referred to in this paper, was supported by NSF, AFOSR, DOD and DOE.

References

Attneave, F. (1954) Some informational aspects of visual perception. *Psychological Review*, 61, 183-193.

Barlow, H.B. (1961) Possible principles underlying the transformations of sensory messages. In: W.A. Rosenblith (Ed.), Sensory Communication (pp. 217-234). NY: Wiley.

Chater, N. (1996) Reconciling simplicity and likelihood principles in perceptual organization. Psychological Review, 103, 566-581.

Chomsky, N. (1959) A Review of B. F. Skinner's Verbal Behavior. Language, 35, 26-58.

Duncker, K. (1945). On problem solving. Psychological Monographs. 58, No. 270.

Graham, S.M., Joshi, A. and Pizlo, Z. (2000) The traveling salesman problem: a hierarchical model. Memory & Cognition 28, 1191-1204.

Hildebrandt S. and Tromba, A. (1996) The Parsimonious Universe. Shape and form in the natural world. New York: Springer.

Koffka, K. (1935) Principles of Gestalt Psychology. New York: Harcourt, Brace.

Köhler, W. (1920/1938) Physical Gestalten. In: W.D. Ellis (Ed.) A source book of Gestalt psychology. (pp. 17-54). NY: Routlege & Kegan.

Kwon, T.K., Agrawal, K. Li, Y. and Pizlo, Z. (submitted) Spatially-global integration of closed, fragmented contours by means of shortest-path in a log-polar representation.

Li, Y., Sawada, T., Shi, Y., Kwon, T., & Pizlo, Z. (2011). A Bayesian model of binocular perception of 3D mirror symmetric polyhedra. Journal of Vision, 11(4), 1–20.

Li, M. & Vitanyi, P.M. (1994) An introduction to Kolmogorov complexity and its applications. NY: Springer.

Mach (1883/1919) The Science of Mechanics. A critical and historical account of its development. Chicago, IL: Open Court.

Mach, E. ((1886/1959) The Analysis of Sensations. New York: Dover.

Marr, D. (1982) Vision. New York: W.H. Freeman.

Narens, L. (2002) Theories of Meaningfulness. Mahwah, NJ: Erlbaum.

Noether, E. (1918). Invariante Variationsprobleme. Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen, Math-phys. Klasse 1918: 235–257.

Pizlo, Z. (2001) Perception viewed as an inverse problem. *Vision Research*, 41, 3145-3161.

Pizlo, Z. (2008) 3D shape: its unique place in visual perception. Cambridge, MA: MIT Press.

Pizlo, Z., Li, Y. & Steinman, R.M. (2006) A new paradigm for 3D shape perception. European Conference on Visual Perception. St. Petersburg, Russia. August, 2006. Abstract published in *Perception*, 35, Supplement, p. 182.

Pizlo, Z., Stefanov, E., Saalweachter, J., Li, Z., Haxhimusa, Y. & Kropatsch, W.G. (2006) Traveling Salesman Problem: a Foveating Pyramid Model. Journal of Problem Solving 1, 83-101.

Pizlo, Z. & Stefanov, E. (2013) Solving large problems with a small working memory. Journal of Problem Solving 6(1), 34-43.

Pizlo, Z., Li, Y., Sawada, T. & Steinman, R. M. (2014) Making a machine that sees like us. NY: Oxford University Press.

Poggio, T., Torre, V. & Koch, C. (1985) Computational vision and regularization theory. Nature 317, 314-319.

Rissanen, J. (1978). A universal prior for integers and estimation by minimum description length. Annals of Statistics, 11, 416–431.

Rosen, J. (2008) Symmetry Rules. Berlin: Springer.

Rosen, J., & Freundlich, Y. (1978). Symmetry and conservation. Am. J. Phys. 46(10).

Sawada, T., Li, Y. & Pizlo, Z. (2011) Any pair of 2D curves is consistent with a 3D symmetric interpretation. Symmetry 3, 365-388.

Sawada, T., Li, Yunfeng & Pizlo, Z. (2014) Detecting 3-D mirror symmetry in a 2-D camera image for 3-D shape recovery. Proceedings of IEEE 102, 1588-1606.

Sawada, T., Li, Y. & Pizlo, Z. (2015) Shape perception. In: Busemeyer, J.R., Townsend, J.T., Wang, Z.J. & Eidels, A. (Eds.), Oxford Handbook of Computational and Mathematical Psychology, NY: Oxford.

Searle, J. (1980) Minds, Brains and Programs. *Behavioral and Brain Sciences* **3** (3): 417–457.

Shannon, C. E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, 27, 623-656, 379-423.

Shepard, R.N. (1981) Psychophysical complementarity. In: M. Kubovy & J.R. Pomerantz (Eds.), (pp. 279-341). Perceptual Organization, Hillsdale, NJ: Erlbaum.

Turing, A.M. (1950) Computing Machinery and Intelligence, *Mind* 59, 433–460.

Van der Helm, P.A. (2000) Simplicity versus likelihood in visual perception: from surprisals to precisals. Psychological Bulletin, 126, 770-800.

Wertheimer, M. (1923/1958) Principles of perceptual organization. In: D.C.Beardslee & M.Wertheimer (Eds.) Readings in Perception, pp. 115-135. NY: D. van Nostrand.

Wertheimer, M. (1945). Productive Thinking. New York: Harper & Brothers.

Wigner, E.P. (1967) Symmetries and Reflections. Scientific Essays. Bloomington, IN: Indiana University Press. (pp. 28-37).