MBS TECHNICAL PAPER 18-06

Unifying Physics and Psychophysics on the Basis of Symmetry, Least-Action ≈ Simplicity Principle, and Conservation Laws ≈ Veridicality

Zygmunt Pizlo, Department of Cognitive Sciences, University of California-Irvine, Irvine, CA 92697.

Abstract.
Psychophysics refers to the branch of Experimental Psychology that deals with the study of Sensation and Perception. A consensus has grown up among experts in Psychophysics during the last hundred years that the human being's percepts are inferences that are based on a minimum, or simplicity principle, which is applied to the currently available sensory data. These educated guesses play the critical role in establishing *veridical* perceptual representations of the 3D environment, where by “veridical” we mean that the percept agrees with what is “out there.” These veridical representations cannot be achieved without making use of symmetries, much like those known in Physics, where they are essential for characterizing our physical world and deriving the conservation laws. But, unlike in Physics, the important role that symmetry plays in Psychophysics has only been demonstrated and explained within the last ten years. Symmetries represent regularities in our physical world. These symmetries also serve as the source of the redundancies that are inherent in 3D objects and that make vision possible. The main goal of this paper is to show that the similarity between the mathematical formalisms used in Physics and in Psychophysics is not coincidental, and that exploring this similarity can benefit the sciences called Perception and Cognition. This paper includes a brief tutorial about symmetry groups and their relationship to transformation groups as well as to their invariants. It was included to make this material available to readers who are not familiar with these topics.

Keywords: Least-Action, Simplicity Principle, Conservation Laws, Veridicality, Inverse problems, Priors.

The issue of the relationship between Physics and other natural sciences is not raised very often simply because there is such a profound difference in the methods and results between the hard science of Physics and softer sciences such as Biology and Psychology. In the most recent history known to the author, this issue was raised by physicists who were probably emboldened by winning the Nobel Prize. Ernest Rutherford (1871-1937), for example, once said that “all science is either Physics or stamp collecting.” This rather pejorative view of every science other than Physics, was shared by his younger colleague, Richard Feynman (1918-1988), who in a 1974 commencement speech, claimed that “Psychology is a pseudoscience, in the sense that psychologists merely imitate the behavior of the scientists. They collect data, they present talks at conferences, put forth hypotheses, test them and publish papers, but they are not discovering any Laws of Nature and they do not understand the phenomena under study” (Pizlo et al., 2014, p.221).

A less dismissive voice can be found in Feynman’s contemporary, Eugene Wigner (1902-1995), who said that:

> “As far as the physical sciences are concerned, the role of invariance principles does not seem to be near exhaustion. … Hence, invariance principles, giving a structure to the laws of nature, can be expected to act as guides also in the future and to help us to refine and unify our knowledge of the inanimate world. One is less inclined to optimism if one considers the question whether the physical sciences will remain separate and distinct from the biological sciences and, in particular, the sciences of the mind. There are many signs which portend that a more profound understanding of the phenomena of observation and cognition, together with an appreciation of the limits of our ability to understand, is a not too distant future step. At any rate, it should be the next decisive step toward a more integrated understanding of the world. … I confess that I have no conception what the structure of this more integrated science may be and it would be surprising if it continued to contain a hierarchy similar to the one described before, in which invariance and symmetry principles have

> definite places. ... This state of affairs should be remedied by a closer integration of the now separate disciplines" (Wigner, 1967, pp. 36-37).

The present paper discusses the concerns that Wigner raised half a century ago. Specifically, it explores how the *fundamental* relationship established in Physics by Noether (1918), among (i) symmetries, (ii) a least-action principle and (iii) the conservation laws can be used in Psychology to derive *veridical perceptions* by applying *a minimum principle* to *symmetries* present *in the physical world*.

Classical and Modern Psychophysics

Fechnerian Chain of Events

Psychophysics was formally established as the study of Perception by Fechner in his seminal book, *"Elements of Psychophysics"*, published more than 150 years ago (Fechner, 1860/1966). According to Fechner, a percept is the end-result of a causal chain of events that starts with (i) a physical object (called a distal stimulus), followed by (ii) a distribution of physical energy on the surface of the receptors (called a proximal stimulus) that is (iii) transduced, transforming the physical energy of the proximal stimulus into the bioelectrical energy used by the nervous system, which is (iv) processed by the brain to (v) produce a percept. Fechner recognized that the percept, as a mental event, is *qualitatively* different from the physical event that caused the percept, but he was unable to shed light on the relationship between the two, other than proposing the use of physical intervals to measure perceptual intervals. This suggestion was based on the assumption that the percept is a *function*, in the mathematical sense of this word, of the physical stimulus. This proposal allowed Fechner to measure *absolute* and *difference* thresholds. It also made it possible to speculate about methods that might be used to measure perceived magnitudes such as brightness, loudness or beauty.

The Fechnerian paradigm has dominated the field of Perception for over a hundred years despite the fact that the Gestalt view of Perception, which became important near the beginning of the 20th C., did not fit into the Fechnerian formalism. For Fechner, Perception boiled down to feature detection. For the Gestalt Psychologists a percept was essentially an "educated guess" that was based on the incomplete sensory data available at the moment. We now know that the Gestalt Psychologists' approach to Perception was on the right track in no small measure because they realized that both physical and perceptual reality is three-dimensional (3D) despite the fact that the retinal image has only two-dimensions (2D). It follows that the visual system must *infer* the missing depth information. This inference is done by combining a simplicity principle with the 2D retinal image. Without a simplicity principle, there will always be more than one possible perceptual interpretation, usually infinitely many interpretations.

These considerations violate the Fechnerian claim that a percept is a *function* of the stimulus because a mathematical function cannot have more than one output for any given input. The Necker cube shown in Figure 1 is a well-known example of this fact. This line drawing of a wire cube leads to two different percepts. They correspond to the two different depth-orders of the vertices, edges and faces. So, in this stimulus, one input produces two different perceptual outputs. This is why we call such stimuli ambiguous, or bi-stable. One could try to defend the Fechnerian view by pointing out that you never see both of these cubes at the same time, but this defense is a distraction because it does not address the main weakness of the Fechnerian paradigm.

Now, look at Figure 2a which is a circle. This looks like the top of an ordinary water or wine glass. The top of the glass is a perfect circle (or nearly so) because it is a continuous curve (note that we are ignoring the fact that the glass is composed of discrete atoms). When you look at a circle, you see a circle. There seems to be nothing surprising here, but is there? Figure 2b shows how the physical input to the visual system is actually likely to look. It will look like a set of discrete samples produced by a continuous circle simply because of the discrete nature of the visual receptors (cones). Figure 2c shows a more realistic input to our visual system because the receptors in our eye are not arranged regularly on our retina. Once one realizes that a set of discrete points is the real input to one's visual system, it is surprising that we perceive a continuous line, and the fact that we all perceive a *perfect* circle is even more surprising.

It is obvious that the visual system interpolates across the spaces between the receptors, but there are many ways to do this interpolation. One could use a linear or a curvilinear interpolation. We all know from our introductory statistics class that there are always many different regression lines that can approximate a set of data points. The problem being discussed here is a version of a regression problem, slightly more advanced than the kind you studied in college. The regression problems we all solved in our stats classes called for fitting a low-degree polynomial function to the data points. Note that here, the visual system uses a closed curve that is a vector function. Furthermore, this curve does not need to have an analytical representation. It turns out that the visual system produces the simplest closed curve that can be used as an approximation of the discrete input to the visual system. If by *simplest* one means *smoothest*, the result in both Figures 2b and 2c, will be a circle.

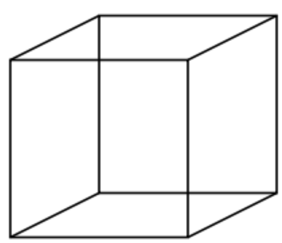
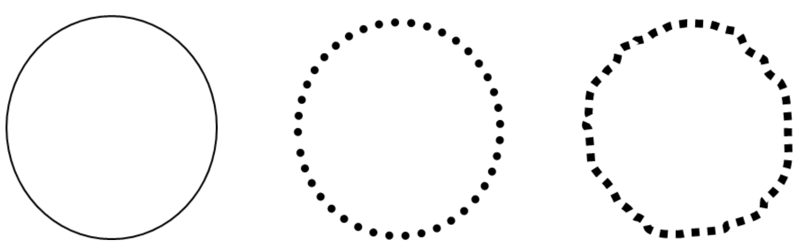


Figure 1. Necker cube that produces two percepts, corresponding to the two depth-orders.



(a) (b) (c)

Figure 2. A circular distal stimulus (a) and its discrete versions (b) and (c) that resemble the input to a visual system sampling with discrete receptors (after Pizlo, 2008). Note that for illustrative purposes, the discrete samples (dots) in (b) and (c) were made very much sparser than the receptors at the center of the retina. This is the reason why (b) and (c) here look so different from (a).

Perception Viewed as an Inverse Problem

We have known for over 30 years that vision is an Inverse Problem (Poggio et al., 1985; Pizlo, 2001). The Inverse Problems paradigm requires breaking the chain of events proposed by Fechner into two parts. The first part is called the "forward problem": it corresponds to forming a retinal (or camera) image of the physical stimulus. Let **V** be a 3D object represented by a 3xN matrix whose columns are Cartesian coordinates of N points of the object. Similarly, let **v** be a 2D camera image represented by a 2xN matrix whose columns are coordinates of points on the image. In the simplest case of an orthographic projection **A** we have:

$$\mathbf{v} = \mathbf{AV} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{V} \qquad (1)$$

In this projection, the x and y coordinates of **V** and **v** are identical, and the z coordinate of **V** is ignored. In a perspective projection, which represents the actual image formation in the eye or camera, the vectors **V** and **v** have dimensions 4xN and 3xN, respectively, and they contain homogeneous, rather than Cartesian coordinates. Matrix **A**, representing the camera geometry, has dimensions 3x4 (Pizlo, 2008).

The word "problem" in the expression "forward problem" is not the best word to use because no problem actually needs to be solved here. We already know how the retinal image is formed, namely the light reflected from the object is brought into focus on the retina by the optical system of the eye (the cornea and the crystalline lens). A computer graphics person working on an animated movie would have to solve a *problem* in geometrical optics when he writes a computer program that computes synthetic images of objects and scenes, but in real life, this forward problem is "solved" by the natural laws of optics, specifically by the reflection and refraction of light.

The second problem that comes up when Fechner's chain of events is broken into two parts is called an "inverse problem". This problem boils down to *inferring* the geometrical and physical characteristics of a distal 3D stimulus from its proximal stimulus, its 2D retinal image. Inverse Problems in vision are ill-posed, which means that there are always infinitely many possible interpretations of the retinal image. This is obvious once one realizes that matrix **A** in equation (1) is rectangular. This implies that it does not have a unique inverse. Put simply, the depth information has been lost in the forward problem and it cannot be computed from the 2D image.

The only known way to solve an Inverse Problem is to impose constraints on the family of possible interpretations. The visual system must know something *a priori* about what the observer is looking at within his visual field. This was formalized by Tikhonov in the early 1960s (Tikhonov & Arsenin, 1977) in his Regularization Theory. According to this theory, the unique, ideally-correct, interpretation **V′** is produced by minimizing a *cost function* E, which consists of two terms:

$$E(\mathbf{V}') = \|\mathbf{A}\mathbf{V}'-\mathbf{v}\| + \lambda\|S(\mathbf{V}')\| \qquad (2)$$

In our example of 3D vision, the first term evaluates how different the retinal projection **AV′** of the perceptual interpretation is from the retinal image **v** that was used to produce the percept. The second term evaluates how different the perceptual interpretation **V′** is from the simplicity constraints that were imposed on the family of possible percepts. In the example shown in Figure 1, the 3D *symmetry* of the percept is the *simplicity constraint* and the function S(**V′**) evaluates the degree of asymmetry of **V′**. Note well that by maximizing the 3D symmetry leads to a unique interpretation, namely a cube, when the 2D image of a cube is used. The parameter λ in equation (2) represents the relative importance of the sensory data vs. the simplicity constraint. If the sensory data are very reliable, λ should be small. If the sensory data are not reliable (noisy), λ should be large. In the example shown in Figure 2, the *smoothness* of the perceived curve is the *simplicity constraint*. Smoothness can be defined in more than one way – in my example, it can be defined as a constant curvature. A circle is a curve with a constant curvature. It is also a *maximally-symmetrical curve*.

Bayesian inference is a probabilistic version of the Regularization method:

$$p(\mathbf{V}'|\mathbf{v}) = c\cdot p(\mathbf{v}|\mathbf{V}')\cdot p(\mathbf{V}') \qquad (3)$$

where c is a normalizing constant. The simplicity constraints become the Bayesian "prior" p(**V′**), and the data are incorporated in the "likelihood function" p(**v**|**V′**). The Bayesian "posterior" p(**V′**|**v**) is used to decide which is the best interpretation. Typically, the maximum of the posterior is used (MAP estimator), but other properties of the posterior probability distribution can be used, as well. Note that there is no parameter λ in the Bayesian formulation. This parameter has been incorporated in equation (3) as the ratio of the variances of the likelihood and the prior.

The bottom line is that combining *a priori* constraints (aka priors) with data is the only known way of making inferences in Perception and Cognition. From here on, the expressions “solving an Inverse Problem” and “making inferences” will be treated as synonymous. All inferences are based on the available data (sensory or cognitive), but these inferences always go beyond the data. This is the essence of inferences. The person making inferences wants to know the true state of affairs “out there”, but he has to do this on the basis of *incomplete* data. Data, by their very nature, are always incomplete. This is as true about a 2D retinal image produced by a 3D object, as it is about the words and sentences that convey one person's thoughts to another. These are only two examples taken from the enormous range of cognitive inferences that can be dealt with once this approach is adopted. This range includes decision making, problem solving, motor control, concept formation and categorization, social cognition, as well as formulating scientific theories. *A priori* constraints, which represent a person’s knowledge of a particular domain, are essential for making correct (veridical) inferences. Many of us used to think that Perception and Cognition rely on the combination of available cues: the more cues, the better the inference. But we know better now. Now, we recognize that *a priori* constraints are at least as important as the data. Combining data with *a priori* constraints always takes the form of an *optimization problem* that uses a Regularization or Bayesian method. This means that *constrained optimization*, a very well developed family of methods in Applied Mathematics, is the formal tool best suited for making and explaining inferences. It is not coincidental that the least-action principle used in Physics is also a *constrained optimization*. This term simply describes how Nature works. In Physics (and Chemistry) the least-action principle produces new invariants, called "Conservations", from symmetries in the Natural Laws. In Cognition, minimizing a cost function produces new knowledge that is the result of an inference from the available data combined with the simplicity constraints that are inherent in the cognitive problem.

To summarize this section, the New Psychophysics differs from the Fechnerian (conventional) Psychophysics by recognizing (i) that the inverse problem of perceptual interpretation is always ill-posed, and (ii) that combining the sensory data with *a priori* constraints is the only way to produce a unique, veridical percept. By “veridical” I mean that the percept agrees with what is “out there.”

Veridical 3D Mental Representations

It should be obvious that simplicity constraints will only work if the physical objects around us are as simple as our definition of simplicity, so the fact that we used 3D and 2D symmetry in the examples shown in Figures 1 and 2 was not coincidental. Symmetry is an effective *a priori* constraint in our natural world because most natural objects are symmetrical, or nearly so. The bodies of all animals are mirror-symmetrical because of the way the animals move. There are only a very few exceptions, such as the fiddler crab, whose body is only approximately symmetrical, but birds, mammals, fish, and insects are all mirror-symmetrical. A jellyfish is characterized by a rotational-symmetry, but note that its rotational-symmetry implies multiple mirror-symmetries. In a mirror-symmetrical object, the two halves are identical; one is a mirror-reflection of the other. Plants are symmetrical because of the way they grow. Petals of flowers form a rotational-symmetry, which means that one petal can be produced from another petal by rotating it around the center of the flower by 360˚/n, where n is the number of petals. One can also rotate the entire flower by 360˚/n instead of rotating an individual petal. This kind of rotation will produce the same flower. The symmetry of most flowers resembles the symmetry of the jellyfish: such flowers are also mirror-symmetrical. An example of a rotationally-symmetrical flower, which is not mirror-symmetrical, is shown in Figure 3. Now consider that leaves of most trees are mirror-symmetrical and all leaves on a given tree are identical. This is another form of symmetry – any leaf can be produced by any other leaf simply by imposing a rigid motion in 3D space and size scaling. Finally, man-made objects are symmetrical because of the function they serve. A chair is usually mirror-symmetrical because of how we sit on it. Screwdrivers are rotationally-symmetrical because of the way we drive screws in or out. A car and an airplane are like animals. They are symmetrical because of the way they move. A completely asymmetrical object would be dysfunctional. If the application of a simplicity constraint to retinal images produces veridical percepts, as it always does with natural, symmetrical objects, the resulting percepts will be good models of our physical reality. Again, by veridical we simply mean that we see things the way they are "out there." What are such visual models good for?


Figure 3. This flower is rotationally-symmetrical, but it is not mirror-symmetrical.

We have known at least since 1948 that the only reliable way to produce purposive (goal-directed) behavior is to use models of the environment. This was the essential motivation of a new field linking engineering and biology called "Cybernetics" (Wiener, 1948; Ashby, 1956). Most of our actions are goal-directed in the sense that we choose actions *now* in order to achieve some *future* goal. I turn on my coffee-maker *in order to* drink coffee 5 minutes from now. On the surface, this might look like Aristotelian *teleology*, in which a future event (drinking coffee) can affect (control) a present event (turning on the coffee-maker), but the Laws of Physics tell us that the future cannot affect the present. So how can Cyberneticians pull this off? The answer is not complicated. In Cybernetics, it is the *model (prediction) of the future*, not the future, itself, that is used to *control (select) the actions now*. If this model is good, our current actions will make it likely that we will get to the desired goal. When the model is not as good, the future goal may not be achieved. There is nothing mysterious here. But without a good model (representation) of the environment, purposive behavior would be impossible.[1] It follows that the importance and ubiquity of purposive behaviors provides a *rational* justification for accepting the claim that we humans, strive to form veridical representations of our environment, the environment that includes 3D objects, 3D scenes, as well as other humans and their purposive actions. But note that there is also growing *empirical* evidence demonstrating that our mental representations of objects, scenes, as well as of other humans, and their mental representations, are indeed veridical, or close to veridical (see Pizlo et al., 2014, for a review of experiments on veridicality in the perception of 3D shapes and 3D scenes). From here on, this paper will focus on explaining the kind of mathematical and computational formalisms needed to form veridical representations, without worrying about whether such representations actually exist. They must.

The remainder of this paper builds on the new paradigm for Psychophysics that has been described so far. It explains in some detail, the nature of: (i) symmetries that are the primary source of the *a priori* constraints, (ii) the simplicity principle and (iii) visual representations. These three entities, which figure

[1] Ernest Nagel (1953) seemed to have missed this point completely when he was trying to figure out the nature of purposive actions. He was aware of Wiener's (1948) Cybernetics, but he overlooked the critical importance of models and representations (Nagel, 1953, p. 198). Warren (1916) seemed to have been one of the first, if not the first, to explain the role of mental models in goal directed actions. Tolman's (1932) work on mental maps and purposive behavior provided empirical evidence of how this works in rats. Finally, it was Conant & Ashby (1970) who emphasized and formalized the role of representations and goals in cybernetics.

prominently in modern Psychophysics, will be compared to the three foundations of modern Physics, namely, symmetry, the least-action principle and the conservation laws. We will begin with a brief tutorial about symmetry because symmetry provides the essence for all this and we cannot discuss the new approach in any depth if we do not share some terminology and some concepts.

Symmetry in Mathematics

Symmetry is one of the most fundamental characteristics in both Physics and Mathematics. Feynman et al. (1963) credited Hermann Weyl for providing us with a good, intuitive definition of symmetry, namely "a thing is symmetrical if there is something we can do to it so that after we have done it, it looks the same as it did before." A human face is mirror-symmetrical because the face in a mirror looks the same as the face in front of the mirror. This is, of course, only approximately true because no human face is *perfectly* symmetrical. Only, after we sculpted a perfectly-symmetrical face, would my example be accurate.

Groups of Transformations and Symmetry Groups

A symmetry group is defined by an equivalence class (invariance) of a transformation group (Rosen, 2008). So, what is a "transformation group?" A set of transformations is a group if: (i) the transformations are closed (a composition of two transformations from the set is in the set), (ii) the identity transformation, which does nothing, is in the set, (iii) for each transformation in the set its inverse, which undoes the transformation, is also in the set, and finally, (iv) the transformations are associative, which means that the result of a composition ABC of 3 transformations, A, B and C does not depend on whether we first compose A and B and the result is composed with C ((AB)C), vs. whether A is composed with a composition of B and C (A(BC)). If the order of transformations does not matter (AB=BA), the group is called commutative.

Consider, as an example, a group of rigid motions. The meaning of the technical term "rigid motion" is the same as it is in colloquial English, namely, moving a chair around without breaking it or distorting is an example of what is meant by a rigid motion. For simplicity of presentation, 2D rigid motion will be discussed now. The group of 2D rigid motions consists of 2D translations and 2D rotations around the origin of the coordinate system. Let (x,y) be Cartesian coordinates of a point before rigid motion and (x_1,y_1) the coordinates after rigid motion:

$$\begin{aligned} x_1 &= x\cdot\cos\alpha + y\cdot\sin\alpha + t_x \\ y_1 &= -x\cdot\sin\alpha + y\cdot\cos\alpha + t_y \end{aligned} \tag{4}$$

where α is the angle of rotation and t_x, t_y are translations along the x and y axis, respectively.
It is easy to see that 2D rigid motions satisfy the group axioms. Note that this group is not commutative: translations do not commute with rotations. How do we define a symmetry group in the group of rigid motions? Note that rigid motions do not change Euclidean distances between pairs of points (Euclidean distance is the conventional distance defined as $[(x_2-x_1)^2+(y_2-y_1)^2]^{1/2}$). This means that a Euclidean distance is an invariant of the rigid motion group. We can say that two figures (objects) are identical if all pairwise distances within one figure (object) are identical to the corresponding distances within the other figure (object). When this is the case, one figure (object) can be produced from the other by a rigid motion. So, identical figures (objects) constitute equivalence classes. It follows that rigid motions form a symmetry group with a Euclidean distance as an invariant. We call it a symmetry group because all objects that are identical can be transformed into one another by a rigid motion.

Figure 4. Wallpaper with translational symmetry.

Some readers may be surprised by what was said in the previous section. They know why a human face or an animal body is considered symmetrical, but they do not think of a rigid motion as a form of symmetry. But it is. A simple, well known example, is wallpaper symmetry. If an arbitrary 2D figure or pattern is *repeated* with fixed intervals along both the x and y directions, and we assume that the wallpaper extends to infinity, translation of the wallpaper by any multiple of these intervals, produces the same wallpaper in the sense that if the translated wallpaper is superimposed on the wallpaper before the translation, these two patterns will coincide (see Figure 4). It is important to point out what makes a wallpaper an intuitively obvious example of a translational symmetry in the way that the rigid motion of an arbitrary object does not. The difference is caused by the *redundancy* contained within wallpaper. The very fact that a pattern is repeated along x and y dimensions makes the wallpaper redundant in the same way that a mirror-symmetrical object is redundant because the left and right halves of an object are identical. The definition of symmetry in mathematics does not require redundancy. It only requires that an equivalence relation is added to the concept of a group. Redundancy, however, becomes essential in perception. In fact, redundant patterns such as mirror-symmetrical animal bodies, rotationally-symmetrical flowers and snowflakes, as well as translationally-symmetrical wallpapers, are standard examples of how symmetry is used in vision and art. One of the main contributions of this paper is the discovery that the *redundancy aspect of symmetry is not only useful for defining visual symmetry, but that it also becomes a fundamental characteristic without which vision, as we know it, would be impossible*.

We will now return to discussing symmetry in mathematics, where redundancy is not an essential element. As explained above, the rigid motion of figures and objects is an example of a symmetry group that includes translations and rotations. Now, it is easy to see how we can form a more general symmetry group by defining symmetry as a transformation that leaves only some essential characteristics of an object intact (invariant), even if the object, itself, is changed. We can add a uniform size scaling to the group of rigid motions, resulting in what is called in geometry a "similarity" group:

$$x_1 = s\cdot(x\cdot\cos\alpha + y\cdot\sin\alpha + t_x)$$
$$y_1 = s\cdot(-x\cdot\sin\alpha + y\cdot\cos\alpha + t_y) \quad (5)$$

where s is the scaling factor.

In a similarity transformation, the object changes because its size and pairwise distances change. What remains invariant are the angles defined by triples of points. Figures and objects that have identical angles are said to have identical shapes, so shape is the invariant in a similarity group. For example, a small and a large square have the same shape: one square can be transformed into the other by a rigid motion and size scaling. Once an invariant is identified, this invariant becomes the essential characteristic that is used to establish the equivalence class and the unessential characteristics can be ignored. So, whenever there is a transformation group and its invariants are identified, we can talk about symmetry and we can make use of the rich formalism of symmetry groups. This is how Felix Klein understood and classified geometries in his 1872 Erlanger Program (Klein, 1939). This is also how physicists understand the symmetries of the Natural Laws. This includes the Lorentz group used in Special Relativity, the Lie groups used in the Standard Model of particle physics, as well as all recent attempts to unify forces, such as the Grand Unified Theory and the Theory of Everything. Symmetry groups have played a central role in algebra, starting with the seminal work of Evariste Galois that was published posthumously in 1846. Some areas of chemistry are also built around symmetries of molecules and crystals.

The symmetry group represented by the similarity equivalence relation, just introduced, can also be used to define the redundancy of objects. An ordinary tree is an example of redundancy generated by a similarity group in the same way that a wallpaper was an example of redundancy generated by a rigid motion group. Leaves on every tree have identical shapes, but some leaves are smaller and others are larger. We can transform one leaf into another by using a 3D rigid motion plus size scaling. This means that every tree is characterized by redundant information related to the similarity of its leaves. Generalized cones (GC), introduced by Binford (1971) and used by Biederman (1987), provide another example, in which a similarity group is used to describe the redundancy contained within an individual object. Look at Figure 5. This object was constructed by starting with a square at the bottom and then taking a copy of this square, translating it, rotating it a little in 3D, changing its size slightly and pasting the result. This was repeated 20 times in Figure 5 to produce this 3D generalized cone. This is an example of a 3D discrete similarity group because all 20 planar cross-sections of this object have identical shapes: they all are related to each other by a 3D *similarity* transformation: rigid motion plus size scaling. Real objects, like the limbs of animals, their torsos, as well as fruits and vegetables, are continuous rather than discrete. We can describe these objects by using a continuous similarity group in which the 3D rigid motions and size changes of cross-sections are infinitesimally small. The symmetry group that characterizes these objects leads to redundancy because all infinitely-many cross-sections have similar shapes.

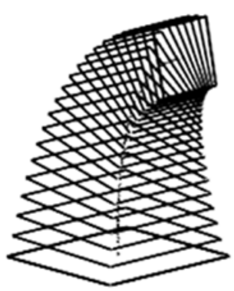


Figure 5. A generalized cone produced by applying a similarity transformation to one cross-section of the object (after Pizlo et al., 2014).

There are even more general transformation groups in geometry, namely affine, projective and topology. These groups are more general than rigid motion or similarity and they are characterized by their invariants (see Pizlo, 2008, for a discussion of this). This means that they are also symmetry groups. Note, however, that we do not have many, if any, real-life examples in which these more general symmetry groups actually generate redundancies in objects.

We will now return to the mirror-symmetry group in order to provide an even better explanation of the contrast between symmetry that generates redundancy and symmetry that does not. Note that the difference between these two aspects of symmetry has never been emphasized, perhaps even been completely ignored, in the past. The mirror-symmetry group is the simplest group because it contains a single transformation (other than identity), specifically, mirror-reflection, and this transformation is its own inverse. Any object, including asymmetrical objects, such as the chair in Figure 6a, can be mirror-reflected. The mirror-reflected object will be different from the original object in the sense that the former cannot be transformed by a rigid motion into the latter. In other words, the two versions of an asymmetrical object cannot occupy the same region in 3D space. With the chair shown in Figure 6a, the left part of the back is taller, when "left" is defined by your actually sitting on the chair. Mirror-reflection is a *symmetry group* regardless whether the objects are mirror-symmetrical or not mirror-symmetrical. The equivalence relation used in this case is equivalence under a reflection transformation, which means that an object and its reflection are in the same equivalence class. Put simply, this is how we group objects in the case of a mirror-symmetry group.

But, now we take objects that are identical to their mirror-reflections, i.e., mirror-symmetrical objects, such as the chair in Figure 6b. Here, we can use the same mirror-symmetry group as described in the previous paragraph, but now our symmetry does generate redundancy in the objects, specifically, in any mirror-symmetrical object, its left half is identical to its right half.

To summarize, we showed that the three elementary symmetry groups, *viz.*, mirror-reflection, rigid motion and similarity, can be used in cases where redundancy is generated, as well as in cases where redundancy is not generated. The mathematics used in these two cases is identical, but it is the redundancy aspect of symmetry that will prove to be essential in the New Psychophysics. We will review how symmetry is used in Physics before we discuss how symmetry is used in Psychophysics. This is important because Physics can provide us with a complete formalism that goes beyond the concept of symmetry.

a b

Figure 6. (a) an object that is not mirror-symmetrical; more precisely, not perfectly mirror-symmetrical. (b) a mirror-symmetrical object.

Symmetry in Physics and its implications

Symmetry in Nature as Regularity

If a student of Physics performs an experiment in his own lab, perhaps testing Newton's Second Law of Motion (F=ma), and subsequently repeats this experiment in somebody else's lab (say one floor up in the same building), he expects to obtain the same result. This subsequent confirmation in a different location is called the *symmetry* of this law. Symmetry here means that this law is invariant under translation in space. This is completely analogous to object invariance (symmetry) in the presence of rigid motions in space: if we measure pairwise distances between points of a rigid object in our lab and measure the corresponding distances after the object has been moved one floor up in our building, the distances will match because they are invariant. We now know that the Laws of Physics are symmetrical under translation in space. The same is true with translation in time, and with rotation in space. These three symmetries are the simplest and the most often mentioned symmetries of the Natural Laws (Feynman et al., 1963). But, note that not all geometrical transformations lead to invariance in the Natural

Laws. The Laws of Physics are not symmetrical under uniform scaling of space. This fact was first established by Galileo in 1638. Any attempt to build a copy of a building, ship or airplane by increasing its size uniformly by a factor of 100, or even only 10, would not succeed because the larger replicas would probably collapse under their own weight (it has been suggested that Galileo was inspired to discuss size invariance by considering whether Dante's inferno (hell) could be realized physically (Peterson, 2002)).

The description of how symmetry is used in Physics, provided in the previous paragraph, shows that symmetries refer to regularities in nature. Each Natural Law applies to infinitely many events, as long as these events differ with respect to non-essential characteristics, such as their position in space or time. If every event in nature were governed by its own law, and if there were no relationships among the laws applying to different events, it would make very little sense to talk about Natural Laws at all. If this were the case, Natural Science would not exist (Wigner, 1964). It is precisely these regularities across space, time and scenarios that allow the scientist to make predictions, and, furthermore, to make them by using only a small number of rules. The observation that science without symmetry and regularity would not exist is as true in Physics as it is in all of the other Natural Sciences, including Biology and Psychology. Note, however, that until now these sciences have neither used the mathematical formalism of symmetry nor explored the formal implications of symmetry for the analysis of their phenomena. This is not to say that Psychologists were not aware of the concept of symmetry. They were, starting with Mach's (1906) book. It is important to know this so a brief overview of how symmetry has been used by perceptionists will be provided later in this paper. The point being made here is that neither Psychologists nor Biologists treated symmetry as *the* foundational concept of their sciences, as Physicists have. Interestingly, Mathematical Psychology, a specialty within Cognitive Psychology, is a notable exception. When Mathematical Psychology was formed as a new specialty half a century ago, the primary motivation for doing this was to use symmetry in Cognitive Psychology as physicists used symmetry in Physics (see Narens, 2002, Chapter 1, for an excellent review of this endeavor). But, in keeping with the zeitgeist prevalent in the 1950s and 1960s, symmetry and invariance in Mathematical Psychology was applied primarily to psychological measurement. Note that this is the kind of measurement that Fechner had in mind when he worked on the perception of loudness and esthetical impressions. This kind of research was extended by S.S. Stevens in the middle of the last century and substantially elaborated by Luce, Falmagne, Krantz, Narens, Suppes and Tversky, since then. Perhaps it was this focus on measurement in the study of one-dimensional sensory stimuli that made it hard for symmetry to spread into the rest of Cognitive Psychology. Note that symmetry and invariance have continued to play an important role in Mathematical Psychology as evidenced by the work of Narens (2007) and Falmagne (Falmagne and & Doble, 2015) on meaningfulness of scientific laws, and more recently, by the work of Vigo (2015) on categorization.

The role of symmetry in Physics has been growing throughout the last 100 years. It started with the breakthroughs represented by Einstein's Special and General Relativity theories. Both can be attributed to treating the symmetry of nature as *the most important principle*. Since then, everything accomplished in Physics has revolved around symmetry, and the presence or the absence of symmetry has been used as a criterion for deciding whether a new theory is true even before any experimental data were collected (Rosen & Freundlich, 1978).

Conservation Laws

About the same time Einstein was formulating his General Relativity theory, Emmy Noether (1918) established a surprising link between the symmetries of Natural Laws and the Conservation Laws (Tavel, 1971). Conservation Laws, such as conservation of momentum, angular momentum and energy, had been known to physicists more or less explicitly at least since Galileo and Newton. But, before Noether, it was never entirely clear how they are related to other concepts in Physics. Before we proceed, it will be instructive to remind ourselves what those basic conservations are. Let's start with the conservation of momentum (specifically, linear momentum), which is the product of mass and velocity (remember, velocity is represented by its magnitude and direction). When two moving objects, say cars, collide, many characteristics of the two-car configuration change. Typically, the cars are wrecked as a result of the collision and they may come to full-stop or bounce-off each other. Despite the fact that many characteristics change, there is at least one that does not change, namely, the total momentum, which is the sum of the individual momentums. So, momentum is an invariant of the collision; it is *conserved*. When the collision is elastic, as in the case of billiard balls, the kinetic energy, in addition to momentum, is conserved, as well. Finally, when a particle undergoes circular motion, the particle's angular momentum, which is the product of the particle's mass, linear speed and radius of the circle, is conserved, too. So, when the radius decreases, its speed increases. We all are familiar with examples of the conservation of the angular momentum, such as a pirouette of an ice skater. Note that the case with an ice skater is more complex than with a particle because in the general case, the computation of the angular momentum uses the moment of inertia, rather than mass, but note that the general idea stays the same. All conservations in physics refer to characteristics that remain invariant, or conserved, during the physical event.

Noether (1918) showed that for every symmetry in nature there is a corresponding conservation law. So, the conservation of momentum results from the symmetry of the laws of motion in the presence of translations in space (the position of the origin of the spatial coordinate system is irrelevant for the Laws of Motion), the conservation of angular momentum results from the symmetry (invariance) of the Laws of Motion to the choice of the orientation of the coordinate system, and the conservation of energy results from the symmetry of the Laws of Motion in the presence of translations in time. Specifically, Noether showed how the Conservation Laws can be derived by applying a least-action principle to symmetries in nature. It follows that the least-action principle, which has been known since the 17th century, establishes the relationship between symmetries and conservations. Noether proved her theorem using the framework of Classical Physics, but her results were quickly generalized to Quantum Physics. It is the nature of this least-action principle that will allow us to compare the Laws of Physics with the Laws of Psychophysics. A brief description of how this principle works in Physics will be provided next.

Least-Action Principle

The least-action principle received its first formal treatment in the 17th and 18th centuries; much earlier than the formal treatment of symmetry and the Conservation Laws in Physics. These did not receive a formal treatment until the end of the 19th and the beginning of the 20th century. So, historically, the least-action principle was *studied* in isolation, separate from the symmetry of natural phenomena, despite the fact that the least-action principle *never* operates in isolation, separate from the symmetries of natural phenomena because there are always symmetries "out there." It is not a coincidence that the least-action principle was also studied in isolation in the other natural sciences. This fact will become important when we make the transition from Physics to Psychophysics later in this paper.

An intuition about the presence and nature of a least-action principle goes back to the ancient Greeks. Evidence of this can be found in Plato, Euclid, and Ptolemy's discussion of the reflection and refraction of light. The same idea shows up a millennium later in Alhazen's treatise on vision (1083), but it did not take on an important role until the middle of the 17th century when Fermat gave us the first modern formulation of the refraction of light using the principle of least-time. Despite the fact that the law of refraction had been described by Snell in 1621 and independently by Descartes in 1637, Fermat took it on, because Snell's and Descartes's formulations applied only to the simplest possible scenario. Snell and Descartes had only related the angles of incidence and refraction, when the light passes through the boundary between two optical media, to the indices of refraction of these media. Fermat was not happy with this formulation because it did not specify the entire path that the light took in a more complicated environment.

Fermat was after a more general principle. He wanted to be able to predict the path the light takes when it goes through a complex optical medium where the index of refraction changes continuously. For example, the path the light takes coming from the sun at sunset, when the light goes through the atmosphere, which has a continuously changing density, thin at the top and dense at the bottom (Feynman et al., 1963, Chapter 26, Figure 26.7). Snell's law was useless to Fermat because without differential calculus (which was not available to Fermat), it could not be applied to continuously changing density of the optical medium. Fermat accomplished his goal by conjecturing that the light chooses the path, from infinitely many paths, that minimizes the time of travel.

A more general version of Fermat's principle says that light chooses the path for which time is stationary, in which infinitesimally small variations of the path do not change the overall time. One needs to use the tools of calculus to estimate this path. Calculus was not available until the second half of the 17^{th} century, so, as already mentioned, Fermat did not have access to these tools. In fact, the least-action principle requires the calculus of variations not differential calculus. In differential calculus, one seeks the value of a variable number x for which a function y=f(x) is minimal or maximal. The necessary condition is that the first derivative dy/dx=0. But in a least-action principle, the argument of a function is a variable path or a trajectory, not a variable number. A path is a function itself, so variational calculus deals with functions whose arguments are variable functions, themselves. Such functions are called functionals. They take the form of integrals, so a least-action principle can be formulated by using integrals. The formal development of the calculus of variations did not start until the end of the 17^{th} century, which meant that despite the fact that Fermat could formulate a least-action principle, he could not apply it to cases that were more complex than those Snell and Descartes had studied. Fermat's intuition about the least-action principle, however, was right on the money, and it did not take long before others showed that the principle of least-time adequately describes the propagation of light. Furthermore, it was shown that a least-action principle generalizes to many other types of phenomena in Physics.

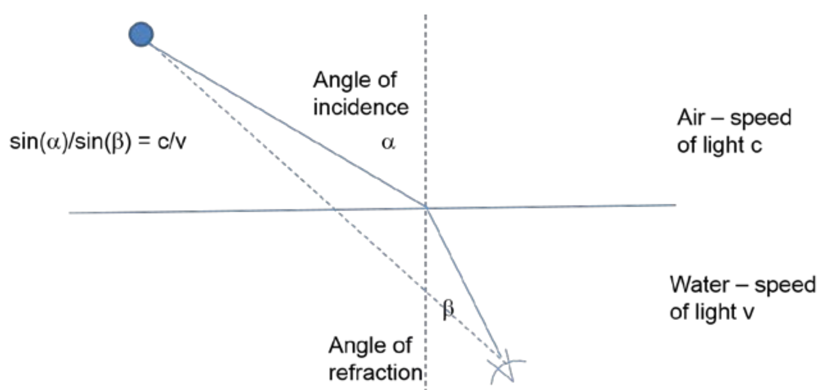


Figure 7. Solid line represents the path of least-action (minimal time). This path satisfies Snell's law of refraction. Dashed line is the shortest path. This path requires more time.

It should be intuitively clear how Fermat's least-action principle is able to account for Snell's law of refraction. A straight-line path from the source of light located in the air will *not* minimize the amount of time it takes to reach the observer's eye which is located in the water. Light will actually spend more time in the sparser medium and it will spend less time in the denser medium, resulting in the minimum total time (see Figure 7).

Fermat's principle was generalized to classical mechanics by Maupertuis and Euler in 1744, elaborated by Lagrange in 1788 and by Hamilton in 1834. When a least-action principle is applied to a moving object, it has been shown that the object will choose a path that minimizes action, where action is defined as an integral of a Lagrangian with respect to time, and where the Lagrangian is defined as the difference between kinetic and potential energy. The least-action principle was subsequently generalized to all other branches of Physics, where minimizing action leads to general solutions and gives rise to better insights into the underlying phenomena. “Action” outside of mechanics is defined differently, depending on the phenomena under study. For example, in electrical circuits, the current distributes itself so that the total amount of heat generated in resistors is minimal, and in soap bubbles, it is the total surface tension that is being minimized. Similar formulations of the least-action principle characterize chemical reactions.

The least-action principle in Physics was studied for its own sake until the beginning of the 20th century, when Noether established the relationship among the symmetries in nature, least-action and conservations. Before Noether’s paper, least-action was viewed merely as the most economical description of nature (Mach, 1893/1919, p. 490). Since Noether published her paper, least-action is no longer viewed merely as a convenient tool in formalizing the Laws of Physics. Instead, it is considered to be the essential property of nature, itself, in the same way that symmetry is the essential property of nature.

Symmetry in Sensory Systems

Here, we will try to integrate the information reviewed so far, and show how Physics can help establish the New Psychophysics on firm scientific ground. Note that we already know that veridical perceptions can be produced by solving an inverse problem, in which the symmetry of objects provides the key *a priori* constraint. In the second half of this paper, I will show that perceptual and cognitive inferences are not made by lucky coincidences that pop up in the results of a few experiments. Instead, I will show that perceptual and cognitive inferences proceed in the same way that the Natural Laws in Physics do, as explained by Noether (1918). This treatment also starts with symmetry, but here its dual nature is emphasized, namely, it contrasts symmetry that generates redundancy with symmetry that does not.

An image formation, which is a perspective projection from the 3D space to the 2D retinal image, has a number of unpleasant properties. It is a many-to-one mapping (see equation 1), which eliminates a large number of mathematical tools that are typically used in the analysis of geometrical transformations and symmetries. There is no meaningful way to define a composition of two image formations simply because an image formation in natural vision is always a 3D to 2D transformation. After the first 3D to 2D transformation is applied, the result is a 2D image, not a 3D space, so it is impossible to apply a second 3D to 2D transformation. Without a composition of transformations, there is no way to talk about the closure and associativity axioms of transformation groups. An identity transformation does not exist and an inverse transformation is not unique. Without any of the four axioms of a group, an image formation is not a part of any conventional geometry (Pizlo, 1994; Pizlo & Rosenfeld, 1992; Pizlo et al., 1997a, b). Specifically, there are no general case invariants in a 2D image of a 3D space (Burns et al., 1990). By “general case invariants” we mean invariants, such as distances or angles, that can be computed for any set of N points. This means that in the general case, symmetries characterizing natural events “out there” will be completely eliminated when a retinal image is formed. This fact imposes serious limitations on whether and how the Laws of Physics, the only real Laws in Science, can be generalized to the Laws of Perception. The hierarchy characterizing the science called Physics, described by Wigner (1967), which begins with the symmetries of the natural laws, goes through the least action principle, and culminates with the Conservation Laws, cannot be applied to the 3D to 2D projection.

Consider a few examples. The speed of movement measured in the retinal image is affected by a translation in physical space, and objects that are farther away move with a lower speed when the speed is measured on the retina. So, movement on the retina is not invariant (symmetric) in the presence of translations in 3D space. As a result, the Law of Conservation of Momentum is also not satisfied in the retinal image. This can be seen clearly when the object is moving with a constant speed towards the observer along a straight line. A constant speed means zero acceleration and zero force acting on the object. There is motion on the retina when the trajectory of the movement does not include the observer's eye, but the retinal size of the object and its retinal image speed increase as the object gets closer to the observer. When speed increases, the momentum, increases as well, violating the law of conservation of momentum. Furthermore, despite the fact that no forces act on the object, the object accelerates in the retinal image, violating the law of conservation of energy. This is a complete mess from a physicist's point of view. When the object is on a collision course with the observer, the center of mass of the object is now stationary in the retinal image. Its speed is zero, so the law of conservation of angular momentum will be violated because changing the orientation of the trajectory relative to the observer changes the object's speed from positive to zero.

So, no symmetry of Natural Laws and no Conservation Laws are available to operate on a 2D retinal image. The least-action principle fails to apply, as well. The 3D paths "out there" are projected to 2D curves in the retinal image, so the shape of the 2D retinal path is almost always different from the shape of the 3D path. There are no invariants of this transformation. It follows that if a least-action principle corresponds to the 3D curve, it will not correspond to its 2D retinal image. Take the simplest example of Fermat's principle when this principle is applied to the reflection of light, where the angle of reflection is equal to the angle of incidence. Angles are not invariant in a perspective projection, which means that Fermat's principle fails in a retinal image.

With all of these complications, it is rather surprising how much effort neuroscientists and machine learning communities have expended on trying to understand vision by limiting themselves to 2D representations of 3D scenes, or even worse, to no representations, at all. *Without restoring veridical (or nearly veridical) 3D representations of the 3D natural environment, it will be hard, if possible at all, to formulate the Laws of Vision, when by "laws" we mean what Physicists mean*. Such laws cannot exist, period, unless one is able to bring in redundancies generated by symmetries in the 3D environment. *Without 3D representations and without the underlying symmetries, vision science, both in psychology and in engineering, will always be a visual art, not a real science.*

The only way for the sensory systems to get access to the *invariants*, called symmetries that operate in the natural environment is to restore the one-to-one mapping between the 3D world and our 2D retinal images. This can be accomplished by using the *redundancies* present in the symmetry of objects. So, the visual system uses the redundancy aspect of symmetry present in the 3D environment to get access to the invariance aspect of symmetry of this environment. Once the one-to-one mapping is restored, the visual system can apply a least-action principle (called by perceptionists a simplicity principle), and derive conservations. Biological systems equipped with sensors bring-in new least-actions and new conservations. The mathematical formalism, here, is essentially the same as the formalism used in Physics. However, these biological least-actions are new because they combine *data* with *symmetries*. When a least-action principle operates in Physics, it does not make use of any data. It is applied directly to symmetries of the Natural Laws. Data *are* used in Cognitive least-actions because these least-actions are cognitive inferences based on *information*. The concept of information exists in Physics where it is related to entropy, but least-actions in Physics do not make inferences. So, a*n inference is a new, Cognitive concept that does not exist in Physics*. These new cognitive least-actions reside in the physiology and chemistry of the brain and the new conservations are the veridical mental representations of the outside world.

Why do I draw an analogy between the Conservation Laws in Physics and veridical mental representations in Psychology? Here is why: when I draw a square on a piece of paper, and I ask you to look at it, you will tell me that you see a square. One may object to this, by pointing out that the fact that you used the word "square" to label your percept does not mean that your perceptual representation of a square is actually a square. But if you tell me that you saw a quadrilateral with 4 axes of mirror symmetry, I know that you saw a square. The symmetry inherent in shapes is sufficient to compare the shapes of physical objects to the perceived shapes and verify that they are actually identical. So, as we all know, at least starting with Fechner (1860), mental events and physical events are qualitatively different, but when the objects undergo the transformation from the physical world to the mental world, the shapes of objects do not change. They are invariant. We perceive these shapes veridically. The suggestion that veridical shape perception corresponds to physical conservation is based on the mathematical formalism: both types of invariants are the result of applying a least-action principle to symmetries existing in the physical world. The rest of this section goes through details of restoring the one-to-one mapping between the 3D world and the 2D retinal image, and through a brief treatment of how the simplicity principle is used to establish veridical 3D shape perception.

Restoring One-to-One Mapping

I begin this section by reminding the reader about a broader context that becomes relevant at this point. Our natural environment and the objects within it are 3D, but our retinal image, which provides the input to the visual system, is 2D. The percept that results is 3D despite the fact that its visual input is 2D. A reader might note that this statement of the problem is not completely accurate because we always look at things with both eyes. So, the actual input to the visual system is a pair of 2D images, not a single image. Once there is a pair of images, the binocular system could use one of these 2D images along with binocular disparity to produce the needed 3D sensory data. This binocular case will be discussed again a little later, but at this point I will only remind the reader that binocular disparity is an unreliable source of visual information, and that it quickly becomes useless when the viewing distance is larger than a few meters. So, one should note that monocular 3D vision seems to be more important in natural viewing than binocular 3D vision. Furthermore, monocular 3D vision is important because it makes it clear why and how symmetry is used in vision.

So, how is the one-to-one mapping actually restored in Perception? Consider a set of points in a 3D space in a mirror-symmetrical configuration and a single 2D perspective image of these points. It turns out that *the mapping between the 3D points in a mirror-symmetrical configuration and a single 2D image is one-to-one*. This means that the group structure that permeates the physical universe, which seemed to have been destroyed by the infinite-to-one mapping when the 2D retinal image of a 3D scene is formed, can be restored when the observer looks at a mirror-symmetrical object.

I will illustrate what is meant here by showing how a pair of 3D points, forming a mirror-symmetrical configuration with respect to a symmetry plane $\boldsymbol{\pi}$, can be computed from a 2D perspective image of these points. We begin with the concept of the vanishing point $\mathbf{v_p}$ that is located in a 2D perspective image and that represents the 3D orientation of $\boldsymbol{\pi}$. The 3D orientation of $\boldsymbol{\pi}$ is uniquely determined by the 3D direction of a line $\mathbf{L}$ normal to that plane (see Figure 8). Let us choose $\mathbf{L}$ in such a way that it contains the center $\mathbf{C}$ of a perspective projection of the camera. This line will be denoted as $\mathbf{L_c}$. Now, the intersection of $\mathbf{L_c}$ with the image plane is the vanishing point $\mathbf{v_p}$ representing the 3D orientation of the symmetry plane $\boldsymbol{\pi}$. It follows that if we know $\mathbf{v_p}$, we also know the 3D orientation of $\boldsymbol{\pi}$ (this is true except for a degenerate case, where $\boldsymbol{\pi}$ contains $\mathbf{C}$). If we have a set of pairs of 3D points symmetrical with respect to a common plane $\boldsymbol{\pi}$, the 2D perspective images of lines connecting the individual pairs of 3D symmetrical points intersect at $\mathbf{v_p}$ because all of the 3D lines are, by definition of mirror-symmetry, parallel to the normal $\mathbf{L}$ of $\boldsymbol{\pi}$. Note that this property is the *only* invariant of a perspective projection of a mirror-symmetrical object. This invariant is not a general-case invariant because it cannot be applied to 3D points that are not mirror-symmetrical. This invariant is generated by the redundancy aspect of 3D mirror-symmetry and it can be used to estimate $\mathbf{v_p}$ (see Michaux et al., 2016).

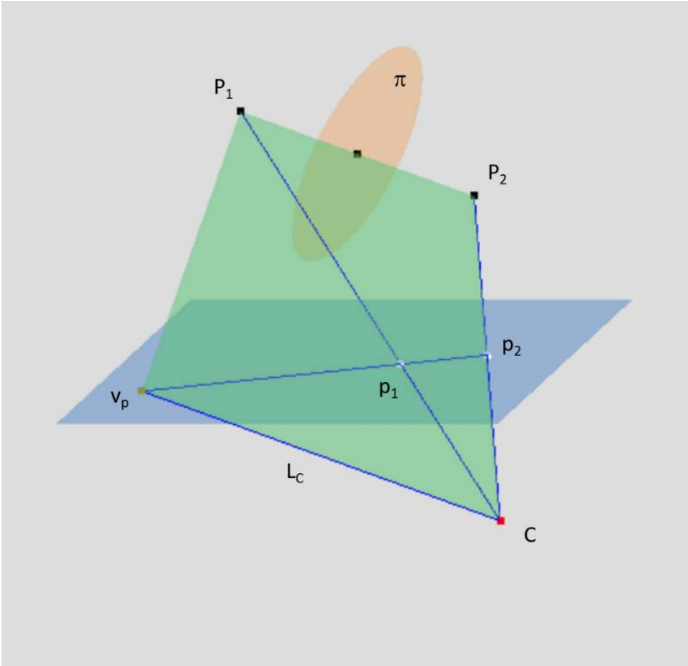


Figure 8. The vanishing point $\mathbf{v_p}$, the center of perspective projection **C**, the 3D points $\mathbf{P_1}$ and $\mathbf{P_2}$, as well as their images $\mathbf{p_1}$ and $\mathbf{p_2}$, are all coplanar (they reside on the same plane marked in green). The blue parallelogram represents the image plane on which the vanishing point $\mathbf{v_p}$ and the two image points $\mathbf{p_1}$ and $\mathbf{p_2}$ reside; these three points are collinear. The symmetry plane $\boldsymbol{\pi}$, represented by an orange ellipse, bisects the line segment connecting the 3D points $\mathbf{P_1}$ and $\mathbf{P_2}$. The line $\mathbf{L_C}$ connecting the center of perspective projection **C** and the vanishing point $\mathbf{v_p}$ is orthogonal to the symmetry plane $\boldsymbol{\pi}$.

Once $\mathbf{v_p}$ is known, we recover the 3D points as follows. Take one pair of 2D corresponding points, $\mathbf{p_1}$ and $\mathbf{p_2}$, at a time. These two points are images of points $\mathbf{P_1}$ and $\mathbf{P_2}$ that are mirror symmetrical with respect to $\boldsymbol{\pi}$. Let the angle formed by points $\mathbf{p_1}$, **C**, and $\mathbf{v_p}$ be ψ, and the angle formed by $\mathbf{p_2}$, **C**, and $\mathbf{v_p}$ be φ. Then, the distances $||\mathbf{P_1}||$ of $\mathbf{P_1}$ from **C** and $||\mathbf{P_2}||$ of $\mathbf{P_2}$ from **C** are computed according to the following equations (see Michaux et al., 2017 for details of the derivation):

$$||\mathbf{P_1}|| = 2d \cdot \sin(\varphi)/\sin(\varphi+\psi)$$
$$||\mathbf{P_2}|| = 2d \cdot \sin(\psi)/\sin(\varphi+\psi) \quad (6)$$

where d is the distance of **C** from $\boldsymbol{\pi}$. Note that the 3D coordinates of $\mathbf{P_1}$ and $\mathbf{P_2}$ can be computed from the distances $||\mathbf{P_1}||$ and $||\mathbf{P_2}||$ by using parametric equations of the projecting lines $\mathbf{Cp_1}$ and $\mathbf{Cp_2}$. The distance d in equations 6 is a free parameter representing an overall size scaling of the 3D recovered object. Specifically, if the plane of symmetry is farther from the camera, the recovered 3D object is proportionally larger. This means that equations (6) allow for a unique recovery of the 3D mirror-symmetrical shape, but not of its size. It follows that the one-to-one mapping mentioned earlier is restored between the 3D mirror-symmetrical *shape* of an object and its 2D perspective image. The size of the object can be recovered by using other cues and priors, such as the horizontal ground on which objects usually reside (Pizlo et al., 2014). If one is interested in shape, but not size, d can be set to ½ simplifying equations (6).

Redundancy is also present in other naturally-occurring symmetries. Consider generalized cones (GC) that can be described by applying a similarity group to the cross-sections of a GC. As long as the cross-sections can be identified, GCs can be recovered uniquely, which means that this symmetry restores a one-to-one mapping between the 3D shape and its single 2D image (Pizlo et al., 2014). The same is true when a 3D shape is recovered from 3D motion (Wallach & O'Connell, 1953; Ullman, 1979; Hoffman & Bennett, 1986; Philip, 1996). When an observer views a rigid object undergoing 3D motion, its 3D shape can be recovered from 3 orthographic or 2 perspective 2D views. Again, the 3D recovery does not include the size of the object, which means that the 3D recovery is related to its shape: rigid motion of an object plus size scaling (i.e., a 3D similarity group). Note that what was often called a "rigidity constraint", is equivalent to a symmetry group defined by the shape equivalence class. This means that the visual system is actually using the redundancy inherent in this group. Bringing in symmetry and substituting it for a rigidity constraint does not really change anything in the mathematics of the 3D recovery from motion, but it establishes a universal mechanism for veridical 3D vision by means of symmetry. Exactly the same symmetry group can be used when the observer views two or more identical objects in a 3D scene. Identical, or nearly identical, objects are encountered when we look at several people. The same is true when we look at a dinner table with several identical chairs, or a tree, whose leaves always have identical shapes. Symmetry can also be used to explain how a one-to-one mapping is restored in binocular vision. Binocular vision can be viewed as representing a symmetry group because the observer receives two views of the same object from a different viewing direction. So, once binocular correspondence is solved, the observer has redundant information about the object. Finally, objects characterized by rotational-symmetry, such as flowers and cylinders of revolution, also provide redundancy and allow for a unique 3D recovery.

Several examples of symmetry in 3D vision were illustrated in the previous paragraph. In all of these cases, symmetry was the source of redundancy that is sufficient, mathematically, for a unique recovery of 3D shape. This means that in all of these cases, the one-to-one mapping between the 3D shape and the 2D retinal images could be restored. This is how our visual system gets access to the 3D space in which invariances reside and in which the Natural Laws of Physics operate. So, when the visual system begins by establishing a one-to-one mapping between the sensory data and the outside world, Psychophysics, which is the science of Perception, has a real chance of resembling Physics, until now considered to be the only *real* science. Establishing a one-to-one mapping between the 3D world and the available 2D sensory data is the first step of visual processing. The second step is applying a least-action, or simplicity principle, to solve the computationally ill-posed and ill-conditioned inverse problem of making a 3D inference. An inverse problem is said to be ill-conditioned if the solution is unstable in the presence of noise in the sensory data. Note that the visual system must select a cost function and an *a priori* constraint that guarantee not only uniqueness, but the stability of perceptual interpretation, as well. I will briefly review some of the prior results on perceptual symmetry before discussing how the perceptual least-actions and perceptual conservations are connected to symmetries.

Symmetry in the Prior Psychological Research

Attneave (1954) explored the use of Information Theory, redundancy and invariance in perception to define "simplicity" as it was conceived by the Gestalt Psychologists. Redundancy is related naturally to symmetry because symmetry refers to the self-similarity of an object or a pattern. The more self-similarity, the greater the redundancy. Attneave proposed "that a major function of the perceptual machinery is to strip away some of the redundancy of stimulation, to describe or encode incoming information in a form more economical than that in which it impinges on the receptors" (Attneave, 1954, p. 189). Note that this means that Attneave was well aware that a simplicity principle was operating in perception and that he also knew how to express it mathematically. What Attneave seems to have missed is the critical role that *a priori* constraints have in perception. This omission is evident because he emphasized economical "description" rather than economical "inference". In the New Psychophysics advocated here, symmetry, acting as an *a priori* constraint, *adds* information to the incoming 2D retinal stimulus. This explains how the 3D percept is produced from a 2D image. So, in my view, when the image is formed on the retina, the visual system detects symmetries within it and uses them to produce a veridical 3D percept. But, according to Attneave's approach, symmetries (redundancies), once detected, are removed from the 2D retinal stimulus. Attneave, by missing the role of *a priori* constraints, had no way of recognizing that perception is based on an inference, or of formulating a theory of veridical 3D vision.

Leeuwenberg and his associates (Leeuwenberg, 1971; Leeuwenberg and van der Helm, 2013) used Attneave's approach, but they tried to substitute their Structural Information Theory for Shannon's Information Theory (Shannon, 1948) that had been used by Attneave. In Structural Information Theory, a few basic symmetries, such as rotation and translation, are applied to elements of a figure or a pattern. Their elements were dots, line segments or corners. The replications of these elements captured the redundancies in the stimulus. Leeuwenberg et al., like Attneave, restricted themselves to economical descriptions, rather than inferences, and their stimuli rarely went beyond 2D toy examples and when they did use 3D objects, they never tried to recover them from 2D images. Put simply, their work suffered from the same shortcomings Attneave's did.

Garner (1970) went a bit further than Attneave. He, like Attneave, recognized the importance of invariants, which suggested to him that groups and transformations should also be used. Garner, unlike Attneave, came very close to recognizing that symmetry is the *central concept* in perception. Garner, unlike Leeuwenberg, considered the symmetries of the entire figure, rather than the symmetries of the individual elements composing the figure. In other words, Garner did not try to compress the figure. He tried, instead, to determine which figure had better Gestalt properties. Unfortunately, he, like all those before him, failed to bring in the concept of *a priori* constraints and did not study the recovery of 3D shapes from 2D images.

Cassirer (1944), before anyone else, played an important role when he emphasized the role of groups and invariants in theories of Perception. It was his paper that attracted the attention of J.J. Gibson (1950), and subsequently of James Cutting (1986) to projective invariants as a possible explanation of shape constancy. Recall that shape constancy refers to the fact that the perceived shape is constant despite changes in the shape of the retinal image caused by changes in the 3D viewing direction. The main problem with Cassirer's approach, as I explained in Pizlo (2008), is that Cassirer assumed that the invariants present in the retinal image were sufficient to do the job. Simply put, he underestimated the complexity of the problem. Shape constancy, a perceptual accomplishment closely related to shape veridicality, is an invariant that is derived from symmetries, in the same way that conservations in physics are derived from symmetries, so, shape constancy is the result of a perceptual inference. It is not the result of the detection or measurement of invariants actually *in* the retinal image, the kind of "direct perception" idea that dominated Gibson's theories.

Chater (1996) understood that perception is an inference and that it is the result obtained by solving an inverse problem by combining *a priori* constraints with sensory data. He did not go on to propose a new theory of perception, but he did provide powerful arguments for the claim that there are no fundamental differences between the deterministic simplicity principle represented by regularization cost functions and probabilistic inferences represented by the Bayesian formalism. So, despite the long-standing controversy between simplicity and likelihood principles, it may be impossible to tell them apart on empirical grounds. Note that Chater managed to express with a formal argument what Mach (1906) suggested near the beginning of the 20th Century.

Next, there is a body of work by Michael Kubovy and his associates that revolved around groups, transformations and symmetries (Kubovy & Wagemans, 1995; Gepshtein & Kubovy, 2000; Strother & Kubovy, 2003). This work focused almost entirely on 2D perceptual organization, essentially providing an economical description of 2D patterns, rather than performing a veridical 3D inference.

Finally, I will compare Roger Shepard's approach to mine because, as pointed out by an anonymous reviewer, Shepard's views are precursors of the views described in this paper. I will focus on Shepard's (1994) paper as most relevant. Shepard did advocate using "universal regularities in the world," "mathematical elegance and generality of theories of that world," and the "principles that reflect quite abstract features of the world, based as much in geometry, probability, and group theory as in specific, physical facts about concrete, material objects." (p. 26). But, despite the overall similarity in our philosophy, there are also very clear, as well as critical, differences that will be highlighted below.

1. Shepard was apparently unaware of Noether's (1918) seminal theorem that established the *fundamental* relationship between (i) symmetries, (ii) a least-action principle and (iii) conservations in Physics. This relationship is at the center of my approach. I derive perceptual veridicality, which is a *cognitive conservation*, by applying *a least-action principle* to *symmetries in the physical world*. Shepard does not do this. Shepard, like all Physicists in the 19th century, is aware of the symmetry/invariance, the least-action principle and the conservation laws. He knows that all three are important, but he does not know how they are related, or perhaps even does not know that all three are related to each other.
2. My approach, which is rooted in Noether's (1918) theorem, proposes a way to connect the *mental world* to the *physical world*. Shepard does not propose this. The operation of a *least-action principle*, which resides in the neural machinery of the brain, leads, in my approach, to the *mental* representation of our *physical* world. This kind of psycho-physical connection, namely, a connection between the mental and physical worlds, is entirely absent from Shepard's framework. In his framework, a least-action principle is only responsible for deciding how *one mental representation* is related to *another mental representation*, be it in mental rotation, in color constancy or in his law of generalization. His theory does not explain the perceptual or cognitive *recovery* of objects or events "out there", the recovery that is the main focus of my theory.
3. The mapping between the physical world and the sensory data is many-to-one. This is clear in the case of visual perception, which deals with the projection from the 3D world to the 2D retina. This projection is *not* a group, so it leaves nothing invariant on the surface of receptors, unless the physical stimulus has *redundancy*. Redundancy is present in objects and events when the objects and events are characterized by symmetry, which is a self-similarity. Shepard does talk about the symmetry of objects and he also defines symmetry as self-similarity. But, note well that symmetry, as a form of redundancy, is *not e*ssential in his theory because he does not need to restore any one-to-one mapping. In my theory, symmetry as a form of redundancy, is *absolutely essential* for the human mind to perform a veridical recovery, that is, to produce veridical 3D percepts from 2D sensory data. This recovery is done by combining *a priori* constraints (aka priors) with the available sensory data. This can be modeled by Regularization Theory or by Bayesian Inference. This kind of recovery resembles the least-action principle used in Physics, but it is more sophisticated because this veridical recovery is a perceptual or cognitive *inference* that is based on *sensory data*, two concepts that do not exist in Physics. Shepard does use priors and Bayes to infer distances in perceptual space. This allows him to derive his exponential Law of Generalization, but note that his priors do not play a very important role in his derivation (see his p. 24). So, Shepard needs priors, but he hopes that they do *not* influence the result of his generalization. This is in stark contrast with my 3D recovery where symmetry and compactness priors are absolutely essential: Without priors, there is no 3D percept and there is no inference.

Least-actions and Conservations in Perception

The idea that visual perception operates in ways analogous to the way that a least-action principle operates in Physics originated with Mach (1906). Specifically, according to Mach, a least-action, or "economy principle", as he called it, surely operated in the human mind. Whether it actually operated in physical Nature, as most physicists of that time assumed, was less clear to him. Mach, being a positivist, allowed for the possibility that the least-action principle is the most economical way to describe physical events, even if these events are governed by as yet unknown laws. Köhler (1920), one of the three founders of Gestalt Psychology, was convinced that the Laws of Physics, specifically, the tendency of physical systems to move towards the state of minimal energy, could correspond to, perhaps even account for, a perceptual tendency to move towards simple interpretations (Ellis, 1938, pp. 17-54). Köhler (1920) cited Mach's (1893) book on mechanics, where the least-action principle is discussed and explained. It is not clear, however, whether Köhler was actually importing the least-action principle into perception because he never used, not even once, the term "Prinzip der kleinsten Wirkung" or "Prinzip der geringsten Wirkung", in his 263 page book.[2]

During the Cognitive Revolution, Hochberg & McAlister (1953) started to use the concept of a minimum or simplicity principle, but they did not link it to Physics. They used concepts taken from Shannon's Information Theory to justify their simplicity principle. It seems likely that Foster (1975, 1978) was the first to use a mathematical formalism of a least-action principle in Physics to explain inferences in visual perception. A decade after Foster's papers were published, the computer vision community started using *cost functions* that recognize the relationship between Inverse Problems in Applied Mathematics and inference problems within visual perception (Poggio et al., 1985). Poggio et al. did see the connection to the least-action principle in Physics, and they postulated that the brain could actually implement this least-action principle in the form of neural circuits (Poggio & Koch, 1985). Not long after this development, the human vision community started using cost functions, as well (Knill & Richards, 1996; Pizlo, 2001). In all of these cases, the cost functions and least-actions were used in isolation, without any connection to symmetry or conservations (but see Weiss, 1997 and Huh & Sejnowski, 2016). These activities resemble how least-actions were used by physicists in 18th and 19th centuries.

The first *explicit* claim about the fundamental relationship between symmetries and cost-functions in solving Inverse Problems in vision was published by the present author and his associates 10 years ago (Pizlo, Li and Steinman, 2006; Pizlo, 2008; Pizlo et al., 2014). Once we began to work within the framework provided by Inverse Problems Theory, we realized that symmetry is the perfect *a priori* constraint. It looked perfect because it was abstract, and it could be used with almost all natural objects, both familiar and unfamiliar. Actually, it can be applied to *all* objects once we exclude clouds in the sky, rocks in the mountains, and crumpled pieces of paper. Existing psychophysical results helped connect symmetry with the cost functions that can be used to solve the Inverse Problem of recovering a 3D shape from one of its 2D retinal images. A decade earlier, we had established that shape constancy is close to perfect with symmetrical objects and completely absent with completely asymmetrical objects (Pizlo & Stevenson, 1999; Chan et al., 2006). Others had already tried to use symmetry for the recovery of 3D shape. This started in 1981 with Kanade, but he, as well as others, who followed his lead, assumed that symmetry was only a useful "trick" that could be added to an already existing large bag of tricks employed in efforts to recover 3D shape. Our contribution was pointing out that *symmetry, especially the redundancy aspect of it, should be counted among the first principles of vision* (Li et al., 2013). The other two "first principles" described in this paper are "a least-action or simplicity principle," and "conservation, or constancy/veridicality laws."

Connecting these two concepts, namely symmetry and cost functions, *i.e.*, least-actions, within Psychophysics played the same role it had played when the symmetry and least-action had been connected a century earlier in Physics. The credibility of our approach, which was based on cost functions and the use of the Inverse Problems Theory in human vision, was greatly strengthened when we brought in the important role that symmetry played in the recovery of 3D shape. By 2006, the vision community was losing interest in cost functions and Inverse Problems, and was moving away from 3D shape

[2] The author is grateful to Michal Bach, who undertook to examine Köhler's (1920) book, and found that the substance of this book is represented well in Ellis's (1938) 37-page summary.

perception, arguably, the most important aspect of vision. The concepts called perceptual constancy and veridicality were becoming dirty words in the vision research literature, even taboo for some vision people. A good example of the attitude of our community towards veridicality can be seen in commentaries to a target paper authored by Hoffman, Singh & Prakash (2015). All of those who commented on this target paper, except for the present author, were absolutely convinced that veridicality in vision is a concept that makes no sense, whatsoever, and that only very naïve philosophers who were completely detached from scientific reality, would be willing to say anything about veridicality out loud, not to mention saying anything favorable. To be fair, this kind of scientific despair had received overwhelming justification from a century of psychophysical experiments, in which unnatural stimuli with no trace of symmetry had been used in all of the research on shape. Once symmetry was allowed into the lab, veridicality had to tag along. Note that outside of the laboratory, where symmetrical objects are the rule with no exceptions, there are no problems with veridicality: When you look at a symmetrical object such as a bird or a box or a fox, you see them as symmetrical. If this is not veridicality, what is? Demonstrating that veridicality can be achieved by using symmetry as an *a priori* constraint in the Regularization Method, opened the door for exploring what else can be accomplished in Psychophysics by making use of the duo called symmetry and least-actions.

Consider mirror-symmetrical shapes because this type of symmetry seems to be the most important for human beings. By using mirror-symmetry as an *a priori* constraint, the visual system recovers the 3D shapes of objects, where by shape we mean invariants of 3D rigid motion plus overall size scaling (Li, Pizlo & Steinman, 2009; Li et al., 2011; Sawada, 2010; Pizlo et al., 2014). So, starting with reflection as a symmetry group, the visual system arrives at a larger, similarity group. This fact conforms to what is known in Physics as the Curie or Symmetry Principle. It says that the symmetry group can never get smaller as events develop in time (Curie, 1894). This means that the symmetry group of the *cause* is a subgroup of the symmetry group of the *effect.* This is surely the case in the recovery of a 3D shape. The recovery begins with the 3D mirror-symmetry group, which is the smallest symmetry group, and ends with the percept of a 3D shape, which is characterized by a 3D similarity group that includes mirror-reflection, rigid motion and size scaling. This observation serves to strengthen my argument about the fundamental similarity between the Laws of Physics and Laws of Psychophysics.

The recovery of a 3D shape is done by applying four *a priori* constraints, namely, 3D mirror symmetry, 3D compactness, planarity of closed contours and minimum surface area (Li, Pizlo & Steinman, 2009; Li et al., 2011; Sawada, 2010). The symmetry constraint in our models was implemented by minimizing the sum of absolute differences between the corresponding angles. In a perfectly mirror-symmetrical shape, this sum is zero. The compactness constraint was implemented by maximizing the ratio V^2/S^3, where V and S are volume and surface area of the 3D reconstructed object. The minimum surface area constraint was used to bias the maximal compactness constraint. Minimizing the surface area makes the object thinner, and less compact. We found that this kind of biased compactness provides the best fits in our psychophysical results. Finally, maximum planarity of closed contours constraint was implemented in our models by favoring 3D curves that are better approximated by planes. These four constraints were included in a cost function with weights that were estimated by fitting the model's reconstructions to the reconstructions of human subjects. Our most recent models work with real camera images. They also explicitly measure the difference between the 2D image produced by the 3D recovered shape and the 2D image that was used to perform the recovery (Michaux et al., 2016, 2017).

Minimizing a cost function is the conventional way of solving an Inverse Problem within the framework of Regularization Theory (Tikhonov & Arsenin, 1977). The cost function, which allows one to make a visual inference, plays the role of the least-action principle in Physics. Least-actions in Physics are always applied to the symmetries of Natural Laws. Without symmetries, which are *regularities* in Nature, least-actions in Physics would never produce anything meaningful. The importance of symmetry is even more pronounced in perceptual inferences, where solving an Inverse Problem amounts to the veridical recovery of what is “out there” on the basis of incomplete data. Recall that without *redundancies* conveyed by symmetries, the sensory data available at the proximal stimulus do not have the one-to-one mapping with the objects (distal stimuli) “out there.” So, as I already pointed out earlier in this paper, *symmetry plays a dual role in Perception*, namely, symmetry serves as a regularity in the Laws of Nature, and symmetry serves as the source of redundancy in the sensory data. The first role is sufficient in Physics, but not in Cognitive Psychology.

Interestingly and surprisingly, the concept of symmetry is actually not mentioned in Regularization theory or in Bayesian inference. Both methods for making inferences use *a priori* constraints (priors), but so far, no theory of how these constraints should be selected has been developed. To date, these constraints have always been selected in an *ad hoc* manner. Sometimes these constraints were formulated trivially as implicit regularities present in all previously seen examples. The only criterion used so far has been whether the constraints worked. A comparison of how these constraints are chosen for use in Inverse Problems in Applied Mathematics and for least-actions in Physics, suggests that the *constraints used for solving Inverse Problems should always represent symmetries inherent in the problem being solved*. It seems likely that all *a priori* constraints that were effective when used to solve Inverse Problems in the past did represent symmetries, but this possibility has never, to my knowledge, been discussed either within Regularization Theory or Bayesian Inference. This should be done because the explicit use of symmetry in Inverse Problem Theory is likely to bring this theory to a new level where it might help us to include it in a Unified Natural Science.

Once 3D shapes are recovered, they become new invariants in Psychophysics, *invariants* that represent perceptual *conservations*. These conservations are different from the conservations known in Physics. For example, recovered 3D shapes are scale invariant, so scale invariance operates in Perception but recall that it does not operate in Physics. In any case, perceptual representations of the outside world are conservations that are derived by applying perceptual least-actions that resemble least-actions in Physics. It turns out that Psychophysics is based on the same hierarchy that Wigner used to describe Physics, but Psychophysical least-actions are different from physical least-actions because the former combine symmetries with sensory data. There is no concept like sensory data in Physics. Once psychophysical least-actions are different, psychophysical conservations must be different, too. Specifically, Psychophysical conservations are veridical representations of the outside world. They are used to plan and execute purposive actions in the physical environment.

In this context, the empirical results obtained in the area of Cognitive Psychology called “intuitive Physics” are not surprising (Kellman & Spelke, 1983; Spelke, 1998). Once we realize that visual Perception, as well as other types of perception such as tactual and auditory percepts, are conservations derived from the symmetries of the Natural Laws, it is natural to expect that the perceptual systems have access to the symmetries and conservations that characterize Physics. We already know that "intuitive Physics” is not perfect in the sense that some Natural Laws are not intuitive, but a number of the Natural Laws described by Spelke, are intuitive. So, intuitive Physics may actually reside in the perceptual systems, rather than in higher levels of cognition. Similarly, now that we know that intuitive Physics is present quite early in the human's development, the widespread use of "thought experiments" in Physics appears to be quite natural, too. Shepard (2008) described several examples where physicists were able to derive the Laws of Physics without performing a single physical experiment. This was true with Archimedes’s Law of the Lever, Galileo’s Law of Falling Bodies, Newton’s Law of Action and Reaction, as well as Einstein’s Special and General Relativity Theory. In all of these cases, the assumption that Natural Laws obey symmetry was sufficient to derive each of these particular laws. Physical experiments were performed after the thought experiment was carried out, confirming what the thought experiments had already established. Again, the scientist’s intuition, which is based on symmetry, seems natural simply because our perceptual representations rely on symmetry, just as the Laws of Physics do. This explains, at least partially, “the unreasonable effectiveness of mathematics in the natural sciences”, a fact that Wigner (1960) found intriguing. Symmetry is a mathematical concept, so once we realize, with Rosen (2008), that “symmetry rules” in Natural Science, the effectiveness of mathematics in natural sciences cannot be surprising, especially once we recognize that mathematics is a product of the human mind. These comments bring us to the point where an attempt will be made to generalize Wigner’s hierarchy from Psychophysics to the rest of Cognitive Psychology.

Extending the Science of Perception to the Science of Cognition

All Cognition is concerned with inferences. This includes linguistic communication, learning, decision making, problem solving, motor control, concept formation and categorization, social cognition, as well as formulating scientific theories. The Theory of Inverse Problems tells us that veridical (correct) inferences must combine data with *a priori* constraints. This takes the form of a cost function, whose minimum corresponds to the correct interpretation. This approach to Cognitive Psychology was adopted, among others, by Feldman (2000, 2016), Chater (1999) and Chater & Vitanyi (2003), but in their hands, the *cost functions operated in isolation.* This is the same shortcoming that characterized the early work of Physicists in the 18th and 19th centuries, as well as the early work of Psychophysicists in the second half of the 20th century. Recall that these researchers did see a need to use a least-action principle and cost functions early on, but they either did not see any reason to connect them to symmetries and conservations, or they were simply not aware of them when they did this work. Similarly, Cognitive Scientists, who started using cost functions to explain how the human mind finds the simplest, or the most likely, interpretations of the sensory data, did not explore the nature of the regularities and redundancies (symmetries) in the system that had generated the data. Their emphasis was placed on *compressing* the data, under an implicit assumption that the simplest (most compressed) interpretation will be the correct interpretation. Note that this kind of search over all possible interpretations is neither possible nor needed. Such a blind search may only work with very simple inferences. In all realistic cases, the inference process must begin with restoring the one-to-one mapping between the distal stimuli characterizing the cognitive problem at hand and the sensory data available. Restoring a one-to-one mapping must be based on the redundancies that are present in the distal stimuli and making use of these redundancies requires identifying the symmetry that is the source of the redundancy. If incorrect, or if no symmetry is assumed, a one-to-one mapping will not be restored and the inference will fail. The redundancies in the distal stimuli and the resulting invariants in the sensory data, as well as the nature of the cognitive task at hand, inform the observer about the relevant symmetries and these symmetries should drive the inferential process. Simply put, there is no reason to assume that a visual or cognitive system has a *single* cost function that is applied to all stimuli and all tasks. For example, when am I looking at a mirror-symmetrical object, my visual system should use the redundancy inherent in the mirror-symmetry, rather than in the rotational symmetry, and the cost function should measure the departure from the former, not from the latter. This fact was pointed out and used by the present author in his early paper on inverse problems (e.g., see Pizlo, 2001; footnote 2, p. 3151). A similar approach seems to have been employed by Tenenbaum et al. (2006) in their hierarchical Bayesian models of cognitive inferences.

Tenenbaum et al. (2006) started their paper by saying that “traditional accounts of induction emphasize either the power of statistical learning, or the importance of strong constraints from structured domain knowledge” (p. 309). Tenenbaum et al. go on to state that both components are necessary, and that they should be combined by using the Bayesian formalism. Not surprisingly, in their approach, the sensory data are incorporated in the likelihood function, and the constraints, representing *a priori* domain knowledge, are Bayesian priors. The product is a posterior that is used for making optimal inferences. What is new in Tenenbaum et al.’s approach, when compared to prior approaches, is that they have a number of different *principles* that represent a collection of *a priori* constraints taken from structured domain knowledge. These principles include, but are not necessarily limited to: taxonomic trees, directed networks, linear orders, clusters, and causal laws. Each of these abstract *a priori* principles applies to a wide range of inferences, which represents what is called an "invariance" in Mathematics and Physics. By identifying these principles, Tenenbaum et al. effectively grouped cognitive inferences into subsets that look like equivalence classes. If these subsets are actually equivalence classes, the individual principles can be viewed as defining symmetry groups. No one analyzed Tenenbaum et al.'s hierarchical Bayesian approach using the formalism of symmetry groups and the underlying invariants, but I am convinced that this is well worth doing. Note that using symmetries proved absolutely instrumental in explaining visual inferences, so the chances are that the same will be true with cognitive inferences. After all, we know that one can produce new and meaningful results, that is, *veridical* perceptions in Psychophysics, and *conservations* in Physics, *only* when cost functions and least-actions are applied to symmetries. But note that in order to apply the approach advocated here to cognitive inferences, it is essential not only to specify the equivalence classes implied by symmetries. It is also necessary to identify the redundancies that characterize the symmetries in the distal stimuli. Without redundancies, the inferences from the sensory data will be impossible simply because of the many-to-one mapping from the real state of affairs "out there" to the sensory data. These redundancies surely exist in the kinds of inferences that Tenenbaum et al. discussed in their paper. But note that these redundancies must also be made explicit in the computational models that are formulated to account for the inferences. If this is not done, one will be limited in cognitive experiments and theories to only the simplest cases that can be handled by a one-to-one mapping from the distal to the proximal stimulus, the kind of mapping that limited psychophysicists in their perceptual experiments for more than a century.

To summarize, Cognitive Science, following in the footsteps of Psychophysics, can, and probably should use the same mathematical structure that Physics uses, namely, symmetry, least-actions and conservations and connect them in the same way. Specifically, we should treat cognitive inferences as conservation laws and derive them from symmetries characterizing the physical and social environment through the application of a minimum, or simplicity principle. Doing this would give us what I consider to be a minimal program. If this were adopted, Cognitive Scientists could benefit from the enormous set of mathematical tools that Physicists have been using for the last 100 years. Psychophysicists are already benefiting from doing this. But, if one chooses to be more ambitious, one can go much further by assuming that the neurobiological mechanisms, operating in the brain, actually implement the least-action principle that connects the mental representations with the outside world. The resulting theoretical structure could provide the often-sought approach to a Unified Natural Science. Will all aspects of this newly unified Cognitive Science and Neuroscience be a *hard* science based on symmetries, least-actions and conservations? Almost surely not, but whatever remains outside this scientific structure, probably runs the risk of being called "stamp collecting", at least by Physicists.

Acknowledgment: This work was supported by a grant from NIH (1R01EY024666-01). The author thanks Louis Narens, Robert M. Steinman and two anonymous reviewers for comments and suggestions, and Isaac Weiss for comments and corrections.

References

Alhazen (1083/1989) *The optics. Books 1-3.* Translated by A.I. Sabra, London: The Warburg Institute.

Ashby, W.R. (1956) *An introduction to cybernetics*. NY: Wiley.

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

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. *Psychological Review* 94, 115-147.

Binford, T.O. (1971) Visual perception by computer. *IEEE Conference on Systems and Control*. Miami.

Burns, J.B., Weiss, R. & Riseman, E.M. (1990) View variation of point set and line segment features. *Proceedings of DARPA Image Understanding Workshop*, 650-659.

Cassirer, E. (1944). The concept of group and the theory of perception. *Philosophy and Phenomenological Research* 5(1), 1-36.

Chan, M.W., Stevenson, A.K., Li, Y. & Pizlo, Z. (2006) Binocular shape constancy from novel views: the role of a priori constraints. *Perception & Psychophysics* 68, 1124-1139.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review* 103(3), 566-581.

Chater, N. (1999) The search for simplicity. *Quarterly Journal of Experimental Psychology* 52A, 273-302.

Chater, N. and Vitanyi, P. (2003) Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Science* 7, 19-22.

Conant, R.C. & Ashby, W.R. (1970) Every good regulator of a system must be a model of that system. *International Journal of Systems Science* 1, 89-97.

Curie, P. (1894/1982) On symmetry in physical phenomena, symmetry of an electric field and of a magnetic field. In: J. Rosen (Ed.), *Symmetry in Physics: Selected Reprints*. Stony Brook, NY: American Association of Physics Teachers (pp. 17-25).

Cutting, J. E. (1986) *Perception with an eye for motion*. Cambridge, MA, USA: MIT Press.

Ellis, W.D. (1938/1955) *A source book of gestalt Psychology*. NY: Routlege.

Falmagne, J.-C., and Doble, C. (2015) *On meaningful scientific laws.* Springer, Berlin.

Fechner, G. (1860/1966) *Elements of psychophysics*. New York: Holt, Rinehart & Winston.

Feldman, J. (2000) Minimization of Boolean complexity in human concept learning. *Nature* 407, 630-633.

Feldman, J. (2016) The simplicity principle in perception and cognition. *WIREs Cogn. Sci.* **7**, 330-340.

Feynman, R.P., Leighton, R.B. & Sands, M. (1963) *The Feynman lectures on physics.* Addison-Wesley, Reading, MA.

Foster, D. H. (1975) Visual apparent motion and some preferred paths in the rotation group SO (3). *Biological Cybernetics* 18.2: 81-89.

Foster, D. H. (1978) Visual apparent motion and the calculus of variations. In: Leeuwenberg, E.L.J. & Buffart, H.F.J.M. (Eds.), *Formal theories of visual perception*, pp. 67-82. NY: Wiley.

Garner, W. R. (1970) Good patterns have few alternatives: Information theory’s concept of redundancy helps in understanding the Gestalt concept of goodness. *American Scientist* 58 (1), 34–42.

Gepshtein, S., & Kubovy, M. (2000) The emergence of visual objects in space–time. *Proceedings of the National Academy of Sciences* 97(14), 8186-8191.

Gibson, J.J. (1950) *The perception of the visual world.* Boston: Houghton Mifflin.

Hochberg, J. & McAlister, E. (1953) A quantitative approach to figural “goodness”. *Journal of Experimental Psychology*, 46, 361-364.

Hoffman, D.D. & Bennett, B.M. (1986) The computation of structure from fixed-axis motion: rigid structures. *Biological Cybernetics* 54, 71-83.

Hoffman, D. D., Singh, M. & Prakash, C. (2015) The interface theory of perception. *Psychonomic Bulletin & Review* 22.6: 1480-1506.

Huh, D., & Sejnowski, T. J. (2016). Conservation law for self-paced movements. *Proceedings of the National Academy of Sciences*, 201608724.

Kanade, T. (1981) Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence* 17, 409-460.

Kellman, P.J. & Spelke, E.S. (1983) Perception of partially occluded objects in infancy. *Cognitive Psychology* 15, 483-524.

Klein, F. (1939). *Elementary mathematics from an advanced standpoint: Geometry*. New York: Dover.

Knill, D.C. & Richards, W. (1996) *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Köhler, W. (1920) *Die physischen Gestalten in Ruhe und im stationären Zustand*. Braunschweig, Druck und Verlag.

Kubovy, M. & Wagemans, J. (1995) Grouping by proximity and multistability in dot lattices: a quantitative Gestalt theory. *Psychological Science* 6, 225-234.

Leeuwenberg, E.L.J. (1971) A perceptual coding language for visual and auditory patterns. *American Journal of Psychology* 84, 307-349.

Leeuwenberg, E. L. J., & van der Helm, P. A. (2013) *Structural information theory: The simplicity of visual form*. Cambridge, UK: Cambridge University Press.

Li, Y., Pizlo, Z. & Steinman, R.M. (2009) A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vision Research* 49, 979-991.

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, Y., Sawada, T., Shi, Y., Steinman, R.M. & Pizlo, Z. (2013) Symmetry is the sine qua non of shape. In: S. Dickinson & Z. Pizlo (Eds.), *Shape perception in human and computer vision* (pp. 21-40) Springer, London.

Mach, E. (1893/1919) *The science of mechanics*. Chicago: Open Court.

Mach, E. (1906/1959) *The analysis of sensations*. New York: Dover.

Michaux, V., Jayadevan, V., Delp, E. & Pizlo, Z. (2016) Figure-ground organization based on 3D symmetry. *Journal of Electronic Imaging* 25(6).

Michaux, V., Kumar, V., Jayadevan, V., Delp, E. & Pizlo, Z. (2017) Binocular 3D object recovery using a symmetry prior. *Symmetry* 9, 64: Special issue: Symmetry in Vision (doi:10.3390/sym9050064).

Nagel, E. (1953) Teleological explanations and teleological systems. In: Ratner S. (Ed.), *Vision & Action: Essays in honor of Horace M. Kallen in his 70th birthday.* (pp. 192-222). New Brunswick, NJ: Rutgers University Press.

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

Narens, L. (2007) *Introduction to the theories of measurement and meaningfulness and the use of symmetry in science.* Lawrence Erlbaum, Mahwah, NJ.

Noether, E. (1918) Invariante variationsprobleme. Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-Phys. Klasse, 235-257.

Peterson, M. A. (2002) Galileo’s discovery of scaling laws. *American Journal of Physics* 70.6: 575-580.

Philip, J. A. (1996) Non-Iterative Algorithm for Determining All Essential Matrices Corresponding to Five Point Pairs. *The Photogrammetric Record* 15.88: 589-599.

Pizlo, Z. (1994) A theory of shape constancy based on perspective invariants. *Vision Research* 34, 1637-1658.

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

Pizlo, Z. (2008) 3D Shape: *Its unique place in visual perception.* MIT Press, Cambridge, MA.

Pizlo, Z., Li, Y. & Steinman, R.M. (2006) A new paradigm for 3D shape perception. *Perception*, 35, Suppl. p.182.

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

Pizlo, Z. & Rosenfeld, A. (1992) Recognition of planar shapes from perspective images using contour-based invariants. *Computer Vision, Graphics & Image Processing: Image Understanding* 56, 330-350.

Pizlo, Z. Rosenfeld, A. & Weiss, I. (1997a) The geometry of visual space: about the incompatibility between science and mathematics. Dialogue. *Computer Vision & Image Understanding* 65, 425-433.

Pizlo, Z. Rosenfeld, A. & Weiss, I. (1997b) Visual space: mathematics, engineering, and science. Response. *Computer Vision & Image Understanding* 65, 450-454.

Pizlo, Z. & Stevenson, A. (1999). Shape constancy from novel views. *Perception & Psychophysics* 61, 1299-1307.

Poggio, T. and Koch, C. (1985) Ill-posed problems in early vision: from computational theory to analogue networks. *Proceedings of the Royal Society of London*, B 226, 303-323.

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

Rosen, J. (2008) *Symmetry rules.* Springer, Heidelberg.

Rosen, J. and Freundlich, Y. (1978) Symmetry and conservation. *American Journal of Physics* 46.10: 1030-1041.

Sawada, T. (2010) Visual detection of symmetry in 3D shapes. *Journal of Vision* 10, 4, 1–22.

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

Shepard R.N. (1994) Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review* 1, 2-28.

Shepard, R.N. (2008) The step to rationality: the efficacy of thought experiments in science, ethics and free will. *Cognitive Science* 32, 3-35.

Spelke, E.S. (1998) Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development* 21, 181-200.

Strother, L., & Kubovy, M. (2003). Perceived complexity and the grouping effect in band patterns. *Acta Psychologica* 114 (3), 229–244.

Tavel, M.A. (1971) Milestones in mathematical physics. Noether's theorem. *Transport Theory and Statistical Physics* 1(3): 183-207.

Tenenbaum, J.B, Griffiths, T.L. & Kemp, C. (2006) Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science* 10, 309-318.

Tikhonov, A.N. and Arsenin, V.Y. (1977) *Solutions of ill-posed problems.* John Wiley & Sons, NY.

Tolman, E.C. (1932) *Purposive behavior in animals and men*. NY: Century.

Ullman, S. (1979) *The Interpretation of Visual Motion*. Cambridge, M.A.: MIT Press.

Vigo, R. (2015) *Mathematical principles of human conceptual behavior: the structural nature of conceptual representation and processing*. NY: Routledge.

Wallach, H. & O'Connell, D.N. (1953) The kinetic depth effect. *Journal of Experimental Psychology* 45, 205-217.

Warren, H.C. (1916) A study of purpose. I. *The Journal of Philosophy, Psychology and Scientific Methods* 13, 5-26.

Weiss, I. (1997) A perspective 3D formalism for shape from shading. In *Proceedings of DARPA Image Understanding Workshop* (Vol. 2, pp. 1393-1402).

Wiener, N. (1948) *Cybernetics*. Cambridge: MIT Press.

Wigner, E. P. (1960) The unreasonable effectiveness of mathematics in the natural sciences. *Communications on pure and applied mathematics* 13.1: 1-14.

Wigner, E. P. (1964) Symmetry and conservation laws. *Proceedings of the National Academy of Sciences* 51.5: 956-965.

Wigner, E.P. (1967) *Symmetries and reflections.* Indiana University Press.