

A DEFINITION OF SYMMETRY

Michel Petitjean

Address: CEA Saclay, DSV/iBiTec-S/SB2SM (CNRS URA 2096), 91191 Gif-sur-Yvette Cedex, France.

E-mail: michel.petitjean@cea.fr; <http://petitjeanmichel.free.fr/itoweb.petitjean.html>

Received: 5 May 2008; *Accepted:* 18 June 2008

Abstract: *A definition of symmetry is presented. It is based under few assumptions and is shown to be applicable to several situations where specific definitions were previously in vigor. The roles of groups and distances are investigated in this framework.*

2000 MSC code: 58D19; **PACS:** 02.20.Bb

1. INTRODUCTION

It is an incredible fact that a so ancient concept such as symmetry has not yet received a widely accepted general definition. Rather, several definitions are found in the literature and on the web. Most time, different terms and wording are used, although the underlying concept seems to be the same. Furthermore, practical definitions are often based on strong assumptions, such as the existence of the euclidean structure for geometric symmetries. In most cases, symmetry is exemplified rather than defined. It is not claimed here that all kind of symmetries are coverable by a single mathematical definition: books reviewing symmetry concepts over a broad spectrum of fields, going from Weyl (1952) to Darvas (2007), show us that there is much to learn and to explore before stating whether or not a unique definition is possible. Our purpose is rather to consider some situations involving different symmetry definitions, and show that a single common one suffices. Moreover, defining symmetry appears to be a hot topic, as

suggested in several recent international conferences (SymCon 2007, ISC 2007, Symmetry Festival 2006).

One of the previous attempts to define symmetry was done in an open access paper by Petitjean (2003). Originally, this latter was devoted to define symmetry and chirality measures rather than to define symmetry itself. The Wikipedia definition of symmetry, which appeared on the web in its most general form in 2005, does not fundamentally differ from the 2003 one. By no way it is claimed that that the 2003 definition was never published before (somebody has to look), and retrieving the first occurrence of the definition is outside the scope of this paper. Here, the deep role of a group structure is investigated and its need is demonstrated rather than being a priori imposed. The need of a metric space is also pointed out.

The mathematical terminology we use here is the set theory one, but we reintroduce some basic mathematical concepts in order to let the paper be self-contained for the broad spectrum of readers of the Journal. For convenience, special set theory symbols such as “for all” and “it exists” are avoided.

2. THE ASSUMPTIONS ABOUT OBJECTS

Intuitively, an object is symmetric when it is declared to be identical to a transform of itself: so, we must be able to declare when an object is identical to one of its transforms. Here, it is pointed out that not all kind of transforms should be allowed in a symmetry context. E.g., any string of at least three bits is such that the permutation of any two identical bits returns the same string. Declaring that we have found a symmetry here would lead to conclude that all strings of more than two bits are symmetric, an obviously false conclusion. Our main idea is thus to allow only distance-preserving transforms (it will be shown further that it solves the problem above). Now we need to define objects and transforms.

First we define a set E of which the elements may be called points, or symbols (bits, digits, letters, etc.), or may have some other name, depending on the practical symmetry study we would like to do.

Then we define objects:

Definition 1: An object is a function having its input argument in E .

E is not the set of objects and we do not care about the values returned by the function, which are not assumed to be in E . The objects are functions defined on E on a way which does not need to be specified now, provided that we are able to declare when an object is identical to another object. This kind of identity has sense under the three following assumptions:

- (a1) An object is identical to itself.
- (a2) If an object is identical to a second object, then the second one is identical to the first one.
- (a3) If an object is identical to a second object, and this latter is identical to a third object, then the first one is identical to the third one.

As known, the properties (a1), (a2) and (a3) of the identity relation are respectively called *reflexivity*, *symmetry*, and *transitivity*. It will be shown further that the word “*symmetry*”, as it is used in the preceding sentence, indeed corresponds to a situation covered by the general definition we are presenting.

The symbol of the equality “=” is used by the mathematicians when the three properties above stand, and these properties have obviously been defined to clarify what the equality symbol should mean. So, we retain it here to denote the identity between objects. It means that the debate occurring in the symmetry community about invariance, similarity, identity, and equality, is not crucial since it is obvious that the lack of any of the properties (a1), (a2), or (a3), would lead to an improper situation. So, the equality symbol is appropriate. Nevertheless, it is important to mention here the extremely interesting remark done by one of the reviewers of the present paper: “I prefer the stricter use of *identical*, where the state of a square rotated by 90 degrees can be equivalent to the original state, while only 360-degrees rotation gives an identical state.” It means that, despite that both the terms *identical* and *equivalent* are based on the properties (a1)-(a3), it could be safe to reserve the former to specific situations.

As known, each set of all identical objects defines a *class of equivalence*.

Our main assumption about E is H_0 :

H_0 : E is a metric space.

In other words, H0 means that we are able to compute the distance between any two elements of E . A distance is a function δ of (E,E) in R satisfying to the following properties.

For all x in E and y in E :

$$\delta(x,y) \geq 0$$

$$\delta(x,y) = \delta(y,x)$$

$$x=y \text{ implies } \delta(x,y)=0$$

$$\delta(x,y)=0 \text{ implies } x=y$$

For all x in E , y in E and z in E :

$$\delta(x,z) \leq \delta(x,y) + \delta(y,z) \text{ (triangle inequality)}$$

As known, the function δ is called a symmetric function, due to the property $\delta(x,y) = \delta(y,x)$. Here again, it will be shown further that the word “*symmetry*”, as it is used in the preceding sentence, indeed corresponds to a situation covered by the general definition we are presenting.

There is a huge of distances defined in various contexts (Deza and Deza 2006). So, as proposed in Petitjean’s 2003 paper, a practical way to define the equality between objects is to exhibit a distance between objects, this distance being not to be confused with the distance defined on E . Thus, identical objects are fully characterized by a null distance between them. It is emphasized that supplying a distance is not mandatory to characterize identical objects. However, it may help to clarify what is the set of objects.

3. THE ASSUMPTIONS ABOUT TRANSFORMS

3.1 BIJECTIVE TRANSFORMS

Having defined the metric space E and the equality over the set of objects (with or without using a distance between objects), we consider now the set F of transforms. Here comes a major assumption about how symmetry should be modeled:

**Objects defined on E are transformed
via transforms over the elements of E .**

Thus we consider the set F of transforms over E .

Let x be an element of E and U a transform, i.e. U is an element of F . The image y of x is an element of E denoted Ux , and we assume that there is at most one transformed element $y=Ux$ of x . Should it happen that there are two images of x through U , we would consider that we are dealing with two transforms. This assumption is denoted $H1$:

H1: Any element has at most one image through a given transform

The debate about alternate notations of the transformed element, such as $U(x)$ (i.e. U is denoted as a function), or xU (i.e. U operates on x on the right of x), is not crucial, and the foregoing conclusions would be the same. So, we retain the most compact notation Ux .

We consider that any element of E can be transformed by any element of F . Otherwise there would exist at least one x which could not be transformed by some U in F . In this latter situation, we consider that in fact U transforms x into x . Thus any element U of F maps E on E . This assumption is denoted $H2$:

H2: Any element has at least one image through a given transform

Assuming that $H1$ and $H2$ stand, the equality between transforms is defined below and we refer to it as the definition 2:

Definition 2: $U_1=U_2$ if and only if $U_1x=U_2x$ for all x in E .

The equality $U_1=U_2$ satisfies to the properties (a1), (a2) and (a3) evoked in section 2.

Looking for a symmetry in an object lead us to compare this object to its image through some transform of the set E on which the object is defined. In this context, we would not like to privilege the role of the object over the role of its image, so that we would like to consider the inverse transform associating each element of E to its image with the need to have this inversible transform also satisfying to $H1$ and $H2$. Let us denote by U^{-1} the inverse of U .

In order to have U^{-1} satisfying to $H1$, U must satisfy to $H3$:

H3: All transforms U of F are injections of E in E .

In other words, for any transform, distinct elements have distinct images, or, alternatively, x and y being two elements and U being a transform, $Ux=Uy$ implies $x=y$.

In order to have U^{-1} satisfying to H2, U must satisfy to H4:

H4: *All transforms U of F are surjections of E onto E .*

In other words, for any transform U , any element of E is an image of an element of E through U .

Summarizing the four assumptions H1-H4: all transforms U are bijections from E onto E . A bijection of a finite set onto itself is called a permutation.

3.2 COMPOSITION OF BIJECTIONS

We consider the set G of all bijections of E onto E . The set F is a subset of G but we do not assume that F is G . However, the definition 2 of equality between transforms is valid for all elements of G .

The composition (U_1U_2) of bijections U_1 and U_2 is defined below and we refer to it as the definition 3:

Definition 3: For all x in E and all U_1 and U_2 in F , $(U_1U_2)x=U_1(U_2x)$.

It is easy to check that the composition of bijections defines a group which operates on E : see definitions and properties of groups in the appendix. Since there is no ambiguity, parenthesis and operator symbols can be omitted in expressions. E.g., the product in the definition 3 is an element of E which can be noted U_1U_2x .

The existence of a group structure has been many times pointed out in the context of symmetry. However, there is some ambiguity about which group(s) are to be considered. The set G of all bijections of E onto E is a group for the composition of bijections, but it is not convenient to identify the set F of transforms to it, as mentioned at the beginning of section 2. An other example is achirality, for which the transforms to be considered could be the indirect isometries, but the composition of two indirect isometries is never an indirect isometry, so that no group is defined this way. Clearly, we need to define carefully which transforms are allowed, i.e. we need to define F . Now comes the assumption H5:

H5: for any x in E and y in E and for any U in F , $\delta(Ux, Uy) = \delta(x, y)$.

In other words, F is a set of bijections preserving the distance δ defined on E . Since there is no reason to discard any distance-preserving bijection, we define F as follows:

F is the set of all bijections of E onto E preserving δ .

F satisfies to the three following properties:

F is not empty because the neutral element I_F satisfies to H5

(I_F is the bijection such that $I_F x = x$ for all x in E).

For all U_1 in F and U_2 in F , $U_1 U_2$ is in F ,

i.e. F is stable for the composition of distance-preserving bijections.

For all U in F , U^{-1} is in F .

Obviously U^{-1} is a distance-preserving bijection if and only if U is a distance-preserving bijection. Thus F is stable for the inversion: the role of the object is not privileged over the role of its image.

The three properties above permits to prove that F is a subgroup of G (see appendix: the properties (A6)-(A8) are satisfied). Furthermore, it is a group action on E , i.e. a group operating on E : the six properties (A1)-(A4) and (B1), (B2) in the appendix are all satisfied.

3.3 TRANSFORMATIONS OF OBJECTS

Now we can define what is the transformation of an object, an object being a function defined on E (i.e. its input argument is in E). We associate one transformation T_U of an object Y to one transformation U in F :

Definition 4: The transform T_U of an object Y is an object $T_U(Y)$ such that for all x in E , $(T_U(Y))(x) = Y(U^{-1}x)$.

In other words, the object Y is a function on E which is transformed via distance-preserving bijections (permutations) of the elements of E .

Remark: The transforms T_U are applicable to all objects and the transformed objects are indeed objects.

For clarity, the index U will be omitted, the transforms $T_{U_1}, T_{U_2}, T_U, T^{-1,U}$, will be respectively noted T_1, T_2, T, T^{-1} , and the images of Y through these transforms will be respectively noted $T_1 Y, T_2 Y, TY, T^{-1}Y$, etc. The transform T_{I_F} , which obviously leaves Y invariant, will be noted I .

Let Θ be the set of transforms of the objects. The composition of distance-preserving bijections induces the composition of transforms of objects, which is defined as follows:

Definition 5: T_1 and T_2 being two elements of Θ , their composition $T_2 T_1$ is such that, for any object Y and for all x in E , $((T_{U_2} T_{U_1})(Y))(x) = Y((U_2 U_1)^{-1}x)$.

Because we have set a one to one correspondence (i.e. a bijection) between the elements U of the group F and the elements T of Θ , then Θ is a group for the composition defined above, i.e. it satisfies to the axioms (A1)-(A4) in the appendix. Moreover it obviously satisfies also to the axioms (B1) and (B2), so that Θ is a group acting (operating) on the set of objects.

Remark: The group Θ operates on the left, i.e. an object Y transformed by T_U is denoted by TY (index U omitted). Should we have defined $(T_U(Y))(x) = Y(Ux)$ rather than $(T_U(Y))(x) = Y(U^{-1}x)$ in the definition 4, we would have again Θ being a group, but operating on the right, i.e. an object Y transformed by T_U would be denoted by YT (index U omitted). In this latter situation, all further conclusions would be the same. This remark is coherent with the fact that we never privileged the role of the object over the role of its image, and conversely, of course.

4. DEFINITION OF SYMMETRY

We summarize here what we need to define symmetry.

Let E be a metric space, δ its associated distance function, and F the set of all bijections of E onto E preserving δ . F has been shown to be a group operating on E . The neutral element of F is I_F .

Let Y be an object defined on E , i.e. Y is a function having its input argument in E .

Definition 6: An object Y is symmetric if there is a bijection U of F , with $U \neq I_F$, such that for all element x of E , $Y(Ux) = Y(x)$.

We do not need anything more to define symmetry. What follows is useful to exhibit some immediate properties of symmetric objects.

A symmetric object Y is such that $Y(U^{-1}x) = Y(x)$, and it is proved iteratively that $Y(U^m x) = Y(x)$ for any signed integer m .

When Y is symmetric for both bijections U_1 and U_2 in F , then $Y(U_1 U_2 x) = Y(U_2 U_1 x) = Y(x)$.

For any object Y , we consider the subset S_{YF} of F containing all elements U of F such that for all element x of E , $Y(Ux) = Y(x)$. This set is not empty because it contains I_F .

S_{YF} is a subgroup of F for the composition of bijections of E onto E preserving the distance δ .

Y is symmetric if and only if S_{YF} contains at least two elements.

We had defined Θ as being the group of transforms operating on the space of objects, such that Θ is the image of the group F via the following bijection: one element T of Θ is associated to one element U of F such that $(TY)(x) = Y(U^{-1}x)$, or equivalently, $(TY)(x) = Y(Ux)$, both definitions leading to the same set Θ and the same properties. For convenience, we keep the notations coherent with the first one (i.e. definition 4).

Now we can define the set of symmetry operators associated to an object Y :

Definition 7: Let S_Y be the subset of Θ containing all elements T of Θ such that $Y = TY$. S_Y is the set of symmetry operators associated to Y .

S_Y is not empty because it contains at least the neutral element I of Θ . So we get an alternate definition of a symmetric object, which is obviously equivalent to the definition 6:

Definition 8: An object Y is symmetric if the set S_Y of its symmetry operators contains at least two elements.

Some immediate properties follow:

For all T in S_Y , any object Y is such that $T^{-1}Y=Y$, and it is proved iteratively that $T^mY=Y$ for any signed integer m .

T_1 and T_2 being two elements of S_Y , then $T_1T_2Y=T_2T_1Y=Y$, meaning that T_1 and T_2 operate commutatively on Y . Moreover, the symmetry operators themselves commute: $T_1T_2=T_2T_1$.

S_Y is a subgroup of Θ for the composition of transforms of objects, and this subgroup is commutative (or abelian).

Remark: Neither Θ nor S_{YF} offer this commutativity property.

5. SYMMETRY EXAMPLES

5.1 EUCLIDEAN SPACES

We first consider symmetrical figures defined by a set of points (e.g. the vertices of an isoscele triangle), a domain of the space, and so on. The set E is here the d -dimensional space, where the elements are points with d coordinates. The distance δ is the usual euclidean distance, and the set F of bijections preserving δ is the set of isometries, which contains all compositions of translations, rotations, and mirror inversions. The objects Y are the indicator functions of the symmetrical figures or domains.

E.g., an isoscele triangle in the euclidean plane can be modeled by the function taking the value 1 at the vertices of the triangle, and taking the value 0 elsewhere. If the sides are to be involved, the triangle is modeled by the function taking the value 1 at each point of the sides, and taking the value 0 elsewhere. If we would like to modelize the whole triangle with its interior, we should consider the function taking the value 1 at each point of the triangle (including sides and vertices), and taking the value 0 elsewhere. If we would like to modelize only the interior of the triangular domain, we should consider the function taking the value 1 at each point interior to the triangle (excluding sides and vertices), and taking the value 0 elsewhere. In all cases, the symmetry would be recognized.

In the unidimensional case ($E=R$), a function Y_1 such that $Y_1(x)=Y_1(-x)$ would be also recognized as a symmetric function. Remark: this kind of symmetry is in fact a mirror symmetry (the mirror is the point at the origin), and so this function should be called achiral. Considering the curve representing the function Y_1 in the plane is an other way to modelize symmetry. In this situation, $E=R^2$, and we would have to consider the indicator function Y_2 of the curve, taking the value 1 at each point of the curve, and taking the value 0 elsewhere: this curve is an achiral object, but the object Y_2 is a function with input arguments in R^2 rather than in R .

If we consider now a symmetric real function y of a real variable x such that $y(x)=-y(-x)$, its curve in the plane has a direct rotational symmetry, which can be recognized via the indicator function Y taking the value 1 at each point of the curve, and taking the value 0 elsewhere. This curve is symmetric, in the sense of a direct symmetry. However, this kind of symmetry cannot be recognized via y itself, since no rotation is possible in R except the identity. Furthermore, the sign inversion of the objects has not been defined.

In the d -dimensional euclidean spaces, the restriction above still applies, due to the fact that we did not define the sign inversion of the objects. Functions invariant upon sign inversion of one or several components of their argument in R^d or invariant upon some permutation of the components of the argument, are such that both ways to recognize symmetry operate: either via objects in R^d (i.e. the function themselves), or via objects in R^{d+1} (i.e. the indicator functions of the curves of the functions). For the functions needing a sign inversion, only the second way operates. For $d>1$, both types of functions may be either direct-symmetric, or achiral, or both.

Translation symmetries are recognized by both ways described above, in any dimension. E.g. a periodic function Y in R or in R^d (such as an helix), would be recognized as being symmetric.

It may be encountered the term *skew symmetry* in various contexts. E.g. the image of a symmetric object in the d -dimensional euclidean space through a full rank affine transform is sometimes called a skew symmetric object. From our definition of symmetry, a skew symmetric object is not symmetric unless the affine transform is an isometry, which is normally not the situation of skew symmetric objects. It should be pointed out that all non-degenerated triangles are skew symmetric, although the term "symmetric" should apply only to isosceles and equilateral triangles.

More generally, we do not recommend that the images of symmetric objects through simple mathematical transforms are called symmetric, unless they are indeed proved to be symmetric. If we would authorize that, there would be no limit on the sophistication of the mathematical transforms, and any object would be flagged as being symmetric, soon or later.

5.2 FUNCTIONS AND DISTRIBUTIONS

Attributing weights on points lead to modelize symmetry situations with distributions, i.e. probability distributions. E.g., the vertices of an equilateral triangle having their respective weights equal to $1/2$, $1/3$, and $1/6$, constitute a chiral object unless the vertices are in a space having more than two dimensions. E being a set assumed to be measurable, a distribution is a function having its input argument in the set of measurable subsets of E and returning an output value in the interval $0;1$. When E is \mathbb{R}^d , the distribution function (i.e. the cumulated distribution function) is a function having its input argument in $E=\mathbb{R}^d$ and which also returns an output value in the interval $0;1$. The distribution and the distribution function should not be confused: their input argument are not in the same set. To achieve the confusion, as a joke, the Dirac delta function has been called a function, although it is a distribution. It is known that the distribution function always completely describes the distribution. Thus it is possible to recognize a symmetric distribution in \mathbb{R}^d from its distribution function. E.g., the gaussian distribution has a symmetric distribution function which can be recognized through the indicator function of its curve.

As mentioned in the previous section, multivariate distribution functions which are invariant upon sign inversion of one or several components of their argument x in \mathbb{R}^d or invariant upon some permutation of the components of the argument, are such that there are both ways to recognize symmetry: either via objects in \mathbb{R}^d , i.e. the distribution functions themselves, or via objects in \mathbb{R}^{d+1} , i.e. the indicator functions of the curves of the distribution functions. For the distribution functions Y such that $Y(x)+Y(Ux)=1$, U being an isometry in \mathbb{R}^d , only the second way operates, because we have not defined operations in the set of objects. It should be pointed out that asymmetry coefficients such as Pearson's skewness and its multivariate analogs are inadequate to recognize symmetry because there are non symmetric distributions having a null centered third-order moment.

When a symmetric distribution function admits a density, the symmetry of the distribution can be recognized through its density function, this latter being interpreted as a weight function. However some situations cannot be modeled through distributions, such as infinite patterns in the euclidean space, which would assume an infinite mass. Here, it is better to consider a weight function in R^d , even if it is not integrable. If needed, the curve of the weight function in R^{d+1} can be considered. E.g., the symmetry properties of periodic functions, helices, etc., on which points are weighted can be analyzed through the symmetry of their weight function.

From the definitions 6 or 7, no symmetry exists when E contains only one element x . If we consider the set of subsets of E , it contains the void subset and E itself, so we turn on to the case of a set containing two elements.

5.3 EUCLIDEAN SPACES WITH COLOR CONSTRAINTS

The vertices of an equilateral triangle having different colors, such as red, green, and blue, constitute a chiral object (unless the vertices are in a space having more than two dimensions). The next example of the use of colors is encountered in chemistry. The fluoro-chloro-bromo-methane has five atoms: assuming them to be punctual and putting the carbon at the center of a regular tetrahedron with the four other atoms lying at its vertices would constitute an achiral model of the molecule if the nature of the atoms is ignored. This situation can be modeled through the attribution of colors to the atoms: when the four substituents of the carbon have all different colors, the model of the molecule is chiral.

The colors are values taken by a variable “color” in a non euclidean space. So, we have to consider a space E which is the cartesian product of two spaces: the euclidean space R^d , and the space of colors C , this latter being assumed to be measurable in order to define distributions on $E = (R^d, C)$. Then, we assume that the random variables have distributions such that to each color c is associated a distribution in R^d , its weight being the probability to get the color c . The full theory is outside the scope of this paper (Petitjean 2002, 2004), but it is clearly an extension of the concept of mixture of distributions (Everitt and Hand 1981).

For clarity, we restrict us to the case where C contains a finite number of colors c_1, c_2 , etc., with respective probabilities $Pr(c_1), Pr(c_2)$, etc. To each color c_i is associated a distribution in R^d , with a distribution function F_i . Thus, z being a vector of R^d and c

being a color of C , an element x of E is a couple (z,c) , and, using conditional probabilities, the objects Y are such that $Y(x) = \sum F_i(z) \mathbf{1}_{\{c=c_i\}} \Pr(c = c_i)$, where $\mathbf{1}_{\{\cdot\}}$ denotes an indicator function.

When there is only one color, i.e. this color has a probability equal to 1, then $Y(x)$ is just the distribution function associated to this unique color. In this situation, it is equivalent to say that there is no color at all and to turn on back to the case of an ordinary multivariate distribution.

Returning to the general case, the distance δ in E between two elements x_1 and x_2 of E , with $x_1 = (z_1, c_1)$ and $x_2 = (z_2, c_2)$, may be built from several ways, depending on the structure of C . If we denotes by δ_e a distance in R^d and by δ_c a distance in C , expressions such as $\delta(x_1, x_2) = \delta_e^e(z_1, z_2) + \delta_c(c_1, c_2)$ can work. Anyway, the isometries in E are the isometries in R^d preserving the distances in C .

In order to clarify what precedes, we examine the case of the chessboard. The distribution in the space of colors C is such that there are two equiprobable colors c_b (black) and c_w (white). To the black color is associated a distribution in the euclidean plane, which is a mixture of 32 equiprobable uniform laws over a square: the vertices of the reference square having for coordinates $(0,0),(1,0),(1,1),(0,1)$, 16 of the 32 squares are images of the reference square through the vectors $(2i,2j)$, the signed integers i and j varying in $-2,-1,0,+1$, and the 16 other squares are its images through the vectors $(1+2i,1+2j)$, i and j varying in $-2,-1,0,+1$. To the white color is associated a distribution in the euclidean plane, which is a mixture of 32 equiprobable uniform laws over a square: 16 of the 32 squares are images of the previous reference square through the vectors $(1+2i,2j)$, i and j varying in $-2,-1,0,+1$, and the 16 other squares are its images through the vectors $(2i,1+2j)$, i and j varying in $-2,-1,0,+1$.

If we neglect the colors, the symmetries of the chessboard reduce to those of the square $(-4,-4),(+4,-4),(+4,+4),(-4,+4)$. Apart the identity, the euclidean isometries in R^d are: the rotations of angles $\pi/2$, π , and $3\pi/2$, and the mirror reflections through the abscissas axis, the ordinates axis, and through the two diagonals of the square. Having two colors in C , the set of bijections in C contains only two permutations: the identity and the permutation of the two colors. If we allow for distance in C the strict conservation of colors, e.g. $\delta_c = 1_{c \neq c_i}$, where 1 denotes an indicator function, we retrieve the symmetries associated to the rotation of angle π , and the mirror reflections through the diagonals. If we allow also the exchange of colors in C , we are now able to retrieve, in addition to the symmetries associated to the identity in C , the following set of symmetries: those

associated to the rotations of angles $\pi/2$ and $3\pi/2$, and those associated to mirror reflections through the coordinates axis. This latter set of symmetries permits to recognize that the colors black and white play a symmetric role in the chessboard. Since there is no need to use a weight function to modelize the chessboard, we could have worked with the indicator function on the reference unit square rather than with the uniform distribution on this square. The conclusions would have been the same.

More generally, if some figure or pattern has more than two colors, we can investigate its symmetry properties at various levels in the space of colors, via adequate subsets of permutations of colors. E.g., allowing to permute only two colors, or two pairs of two colors, or all $k!$ permutations of a subset of k colors, or some combination of permutations of subsets of colors, or allowing all colors to be permuted, or none. However, the set of authorized permutations should have a group structure preserving some distance in C . The analysis of this group structure helps to understand the symmetry structure of the pattern. Also we recall that images of symmetric objects through simple mathematical transforms should not be confused with symmetric objects unless they are proved to indeed be symmetric.

The use of colors may cover special situations, such as when some real function in R^d is not defined elsewhere: the function is extended and takes values in the space $E = (R^d, C)$, one of its values in C being a color associated to the domain where the real function in R^d was primarily undefined, and its corresponding value in R^d being zero. An other situation arises in chemistry and physics, where we have two distributions of charges: the negative charges and the positive charges. Summing them algebraically would need to work with a signed distribution, for which many powerful theorems do not work. Furthermore, it is impossible to retrieve the magnitude of the charges from their summation. Working with two colors (one for each kind of charge), permits to analyze the symmetry via a colored mixture of distributions as proposed by Petitjean (2004): the charges are not added, and only their effects on some other entity are algebraically added when needed, and can be handled via a suitable operator. Infinite crystal lattices may be modelized via an appropriate set of colors and an appropriate weight function taking in account masses and isotopes when needed.

5.4 NON EUCLIDEAN SPACES; FINITE SPACES

In the previous section, we considered functions taking values in the space (R^d, C) , where C is the space of colors. Here we consider functions in finite spaces $E = (E_m, C)$

where E_n is a finite space containing n elements. The space C may be finite or not, as for the euclidean space, or may even be not necessary, but E_n is assumed to be finite. What follows may be extended to countable and denumerable spaces.

A simple example occurs in literature. A palindromic word such as "RADAR" can be modeled with a function which takes its input argument in the set of integers $E=1,2,3,4,5$ and returns a symbol in the set of capital letters A..Z. The distance $\delta(x_1, c_2)$ between the positions x_1 and x_2 of the letters is $\delta(x_1, x_2) = |x_2 - x_1|$. It means that there are only two bijections preserving δ : the identity, and the permutation reversing $(1,2,3,4,5)$ into $(5,4,3,2,1)$. The symmetry of the word "RADAR" arises from this latter. Anagrams such as "ARDRA" do not preserve δ . Extending this approach to texts, to finite or infinite strings of symbols (letters, bits, digits, etc.), to patterns of symbols in the plane or in the space, and so on, is obvious. If we would like to distinguish capital and non capital letters, introduce colors, etc, we use an appropriate space of colors.

A particular situation to be mentioned is matrices and tensors, having their elements in any space we like. E.g. a real matrix of size m is viewed as a function having as input argument a bicomponent vector with both coordinates in $1, \dots, m$, and returning a real. Remark: the term "vector" may be improperly used in this context, but we keep it for clarity. The first component of the vector is the line index, and the second component is the column index. The distance between two vectors x_1 and x_2 is the euclidean norm of their difference: $\delta(x_1, x_2) = \|x_2 - x_1\|$. The matrix transposition is a permutation preserving δ , so that a matrix equal to its transposed is symmetric. Other cases of symmetry in matrices related to the symmetries of a square, such as the symmetry through the antidiagonal, do not seem to have received a particular name. The case of an antisymmetric matrix needs to define the sign of the matrix, so it is handled via an indicator function, as mentioned several times.

A relation over two sets E_0 and E_1 may be modeled as an indicator function having its input argument in $E = (E_0, E_1)$, i.e. the first component of the argument is an element of E_0 and the second component of the argument is an element of E_1 . The function returns 1 when the second component of the argument is associated to the first one, and returns 0 if it is not. It could also return "true" when the second component is associated to the first one, and returns "false" if it is not. Now, assuming that $E_1 = E_0$, we consider the trivial distance $\delta(x,y)=0$ when the non ordered tuple of components of x and the non ordered tuple of components y are equal, and $\delta(x,y)=1$ when the non ordered tuples of x and y are different. In other words, δ is just able to tell us whether or not two elements

of E are identical up to some permutation of their components. The relation is symmetric when it is insensitive to the permutation of the components of its argument. The difference with the usual meaning of symmetry is meaningless: we consider a single argument with two components rather than two arguments. Thus, the property (a2) in section 2 is indeed a symmetry property. Generalizations to ternary, quaternary, etc., relations are obvious. Antisymmetric relations are handled as usual.

More generally, functions of several arguments are viewed as functions of only one argument. Using the trivial distance defined above, we can recognize symmetric functions. As a particular situation, we noticed in section 2 that any distance is symmetric.

5.5 GRAPHS

Graphs are nodes and edges structures, and should not be confused with the curves of real functions of real variables. Graphs are extremely important in many areas. Most chemical databanks store molecular formulas as non directed graphs: e.g. the water molecule H-O-H may be represented by a graph having three nodes (the atoms) and two edges (the bonds). Formally, a binary relation is defined over the set of nodes E_0 . When this binary relation is symmetric (see the end of the previous section), the graph is a non directed graph, else it is a directed graph. Moreover, colors may be associated to the nodes and to the edges. The colors of the nodes and the colors of the edges may be numerical or not. E.g. the water molecule has two nodes with the color "hydrogen" and one node with the color "oxygen", and the two edges have the color "simple bond". When needed, the isotopically labelled molecules are modeled by attributing numerical values to the nodes. For general graphs, when the colors are positive real values, they are most time called weights and may appear either on nodes or on edges on both. When no weights are defined, they can be assumed to take the same value for all nodes, or for all edges. Sophisticated situations require to define colors having several parts, numerical or not (the "or" is not exclusive). The numbering of the nodes and the numbering of the edges are discarded.

Symmetries in graphs arise from the number of graph automorphisms, i.e. the number of isomorphisms between the graph and itself. Usually, the graph is called symmetric when there is more than one automorphism. E.g., the graph of the water molecule is symmetric because it has two automorphisms, and the graph of the ethanol molecule CH₃-CH₂-OH is symmetric because it has twelve automorphisms. In order to modelize

graph symmetries, we define the space of nodes E_0 (each node may be colored), and then the space of edges $E = (E_0, E_0)$ (the edges may be directed or not). A graph is a function Y having its input argument in E and returning the color of the edge. In the simplest case, this color is either 0 or 1 so that Y is just an indicator function of the edges. In this model, there are always m^2 edges, m being the number of nodes. The edges which do not exist receive a special color, e.g. the value 0 of the indicator function. For multigraphs, Y would return an integer. The colors of the nodes and the colors of the edges may have a complex structure containing several components, numerical or not. For clarity, we consider the square matrix A of size m associated to Y , such that each line is associated to the first component of the input argument of Y and such that each column is associated to the second component of its input argument, the elements of the matrix being the colors of the edges. When Y is the indicator function of the edges, this matrix is known as the connectivity matrix of the graph. There are $(m^2)!$ permutations of the elements E because E has m^2 elements. As we did in the previous section, the line index and the column index are a bicomponent vector with both coordinates in $1, \dots, m$. However, the distance δ between two edges x_i and x_j is no more an euclidean distance between the vectors x_i and x_j . We set $\delta(x_i, x_j) = 0$ when $x_i = x_j$ and $\delta(x_i, x_j) = 1$ elsewhere. There are $m!$ isometries, which are biunivoqually associated to the $m!$ ways of numbering the nodes. Let P be a permutation matrix of order m , P' is its inverse (i.e. its transposed), and I is the identity permutation. When A is invariant upon some simultaneous renumbering of its lines and columns, i.e. there is $P \neq I$ such that $PAP' = A$ (meaning that $PA = AP$ and that A and P commute), a non trivial graph automorphism is detected and Y is a symmetric object.

The graph structure is a convenient model in many situations. E.g., the word "RADAR" may be viewed as a non directed graph containing a chain of five nodes and four edges, just as the hydrogen-suppressed structural formula of the dimethoxymethane $\text{CH}_3\text{-O-CH}_2\text{-O-CH}_3$. However, much complicated situations arise in chemistry. Rigid conformers of chemical compounds need to be modeled both through their structural formulas and through their geometry, and most molecules have time-dependant conformations. Symmetry in chemistry is an extremely vast field (Hargittai and Hargittai 1995), and is outside the scope of this paper.

6. CONCLUSION

We have attempted to show that symmetric functions are suitable models for a number of real situations where symmetry properties of objects are investigated. Our present

definition of symmetry is not proved to be universal, and only few physical situations have been examined: experts are welcome to do further investigations. Nevertheless, it is expected that most practical uses are covered through the present model. The community of symmetrists will decide whether or not the term symmetry should receive an official definition, and if any, which one it should be. Some associated topics are (i) the classification of symmetries, which should be done on the basis of the symmetry group structure of the object, and (ii) the measures of symmetries, i.e. symmetry is considered as a quantity varying continuously. In this latter situation, we need to define a distance in the space of objects, with an adequate normalization factor in order to be scale-independant. This distance is minimized for the isometries in the space in which the function have its input argument, or for some selected class of isometries in this space, depending on the structure of the group of isometries.

Acknowledgments

I thank very much György Darvas and Raymond Tennant for having managed the anonymous peer review process of this paper. I am grateful to the first reviewer for his important remark about identity/equivalence, and I thank very much both reviewers for their highly encouraging comments.

APPENDIX: GROUP, SUBGROUP, GROUP ACTION

A group is a set G with a binary operation noted here $*$, satisfying to the axioms (A1)-(A4):

(A1) *stability (or closure)*:

For all U_1 and U_2 in G , the result of $U_1 * U_2$ is also in G

(A2) *associativity*:

For all U_1, U_2 and U_3 in G , $(U_1 * U_2) * U_3 = U_1 * (U_2 * U_3)$
(the result can be noted $U_1 * U_2 * U_3$).

(A3) *neutral element (or identity element)*:

It exists an element I in G such that for all U in G , $I * U = U * I = U$.

(A4) *inverse element (sometimes called symmetric element)*:

For each U in G , it exists an element in G noted U^{-1} , such that $U*U^{-1}=U^{-1}*U=I$, where I is a neutral element.

When the property (A5) stands, the group is commutative (or Abelian):

(A5) *commutativity*: $U_1*U_2=U_2*U_1$.

There is only one neutral element. Proof: I being a neutral element, assume that there is an other neutral element I' . Thus for any U in G , $U*I'=U$, thus $U^{-1}*U*I'=U^{-1}*U$ and thus $I*I'=I$, then $I'=I$.

Any element has only one inverse. Proof: assume that some U in G has two inverses U^{-1} and U'^{-1} . Thus, $U^{-1}*U=I$, thus $U'^{-1}*U*U^{-1}=I*U^{-1}$, thus $U'^{-1}*I=U^{-1}$, then $U'^{-1}=U^{-1}$. The same conclusion comes when starting from $U*U'^{-1}=I$.

The inverse of the neutral element is the neutral element itself:

$I^{-1}*I=I*I^{-1}=I$, thus $I^{-1}=I$.

A subset H of G is called a subgroup for the operation $*$ if it is a group for this operation. It is easy to prove that the properties (A6), (A7), and (A8) together constitute a necessary and sufficient condition for H to be a subgroup of G :

(A6) H is not empty.

(A7) *stability* (or *closure*) for the product operation:

For all U_1 and U_2 in H , the result of U_1*U_2 is also in H .

(A8) *stability* (or *closure*) for the inversion:

For all U in H , U^{-1} is also in H .

Remark: for a finite subgroup, (A8) is not necessary. Proof: for all U in H , all successive powers of U are in H . When generating these successive powers U^i for increasing values of the integer i , and since there is a finite number of elements in H , the sequence necessarily repeats. Thus it exists two integers n_1 and $n_2 > n_1$ such that $U^{n_2}=U^{n_1}$. Setting $n=n_2-n_1$, we deduce that $U^n=I$ and thus that $U^{-1}=U^{n-1}$, which is in H .

A group action is a function of (G, E) on E , i.e. an operation (which we note here \cdot) between an element of G and an element of E which returns an element of E , satisfying to the axioms B1 and B2:

(B1) For all U_1 in G and U_2 in G and x in E , $(U_1 * U_2) \cdot x = U_1 \cdot (U_2 \cdot x)$

This associativity property involves two types of operations, and thus should not be confused with the associativity (A2) which involves only one type of operation.

(B2) For all x in E , $I \cdot x = x$, I being the neutral element of the group.

In other words, I leaves unchanged the elements of E . It is a neutral element for both types of operations defined above.

It is also said that the group G acts (or operates) on E . When no ambiguity occurs, both operation symbols may be omitted in expressions involving elements of G and elements of E .

REFERENCES

- Darvas G. (2007) *Symmetry*, Basel-Boston-Berlin: Birkhäuser, xi + 504 pp.
- Deza E., Deza M.M. (2006) *Dictionary of Distances*, Paris: Elsevier.
- Everitt B.S., Hand D.J. (1981) *Finite Mixture Distributions*, chap. 1, London: Chapman and Hall.
- Hargittai I., Hargittai M. (1995) *Symmetry through the Eyes of a Chemist*, London: Plenum Press.
- ISC (2007) *International Symmetry Conference*, Edinburgh, 14-17 January 2007, <http://isc.cs.st-and.ac.uk/>
- Petitjean M. (2002) Chiral mixtures, *J. Math. Phys.* 43 [8], 4147-4157.
- Petitjean M. (2003) Chirality and Symmetry Measures: A Transdisciplinary Review, *Entropy*, 5 [3], 271-312 (see section 6), <http://www.mdpi.org/entropy/papers/e5030271.pdf>
- Petitjean M. (2004) From Shape Similarity to Shape Complementarity: toward a Docking Theory, *J. Math. Chem.*, 35[3], 147-158.
- SymCon'07 (2007) *The Seventh International Workshop on Symmetry and Constraint Satisfaction Problems*, A Satellite Workshop of CP 2007, 23 September 2007, Providence, RI, USA, <http://www.cmi.univ-mrs.fr/%7Ebenhamou/symcon07/>
- Symmetry Festival (2006) Budapest, 12-18 August 2006, <http://symmetry.hu/festival2006.html>
- Weyl H. (1952) *Symmetry*, Princeton, N.J: Princeton University Press.
- Wikipedia (2005) <http://en.wikipedia.org/wiki/Symmetry>.