# Principle of Representation-Theoretic Self-Duality

**Shahn Majid**

**Abstract**

*We argue that the search for the ultimate theory of physics can be modeled mathematically as the search for an algebraic system with the property that the set of all its representations forms an algebraic system isomorphic to the original. The simplest examples of algebraic systems with this property are identified with corresponding theories of physics. The philosophical setting is developed in the form of a Kantian-like thesis.*

**Key words:** principle, representation theory, self-duality, Pontryagin, Kant, algebraic structure, Hopf algebra, category, quantum mechanics, gravity

## 1. INTRODUCTION AND PRELIMINARIES

Theoretical physics is the search for a complete consistent set of fundamental laws of physics. In this paper we take some steps towards developing a Kantian-like thesis that the ultimate structure or set of laws being uncovered by physicists is no more and no less than the structures allowed by constraints in thinking known as being a physicist.

This long-term goal is not a new idea. There is a long-established tradition in theoretical physics in which it is hoped that certain principles or "*a priori* conditions of physical thought" such as "locality," "causality," and "renormalizability" can serve to narrow down the range of allowed physical theories. See, for example, Ref. 1. The ideal goal in this tradition would be to find one or more such principles sufficiently powerful to constrain the allowed theories to just one, and one that is in agreement with experiment. In this paper we introduce a new and unusual principle of this type, the "principle of representation-theoretic self-duality." Unlike the more familiar principles of "locality," etc. above, this principle is a global one: it pertains to the overall algebraic structure of the theory from a very general point of view, and is grounded in a philosophical problem concerning the distinction between reality and representations of reality. After explaining the principle, we argue in turn *necessity, sufficiency*, and historical evidence. Naturally, we shall also have something to say about the status of this overall program within Kant's trancendental idealism. It should be stressed that the principle and the arguments that we present for it constitute no more than a crude first attempt to characterize physical thought; but one that is intended to survive in some form in more sophisticated formulations. It constitutes the long-term motivation underlying the technical works.[(2)–(8)]

An outline of the paper is the following. Section 2 introduces an interesting piece of mathematics on which the paper depends, a theorem of Pontryagin, and its generalizations. Let $S$ be an Abelian group and denote by $\hat{S}$ the set of all its representations. Then, first, under certain technical restrictions, $\hat{S}$ is also an Abelian group, the dual group, and $\hat{\hat{S}}$ is a group isomorphic to $S$. This is a remarkable property of the category of Abelian groups. The category is "self-dual." Second, within this category, some groups $S$ have the special property that $\hat{S}$ itself is isomorphic to $S$. Such groups are "self-dual." These same notions make sense in some considerable generality. It turns out that some of the few self-dual structures available are ones that arise in an algebraic approach to quantum mechanics combined with gravity[(2),(3),(5),(6)] and in more recent work connected with string theory and conformal field theory.[(7),(8)] The considerations there lead us to formulate the principle of (representation-theoretic) self-duality: *A complete consistent theory of physics must have a self-dual structure in the sense described.* In this paper we always mean "self-duality" in this representation-theoretic sense.

In Sec. 3 we argue the necessity of this principle: why should we accept this principle? A two-step Kantian trancendental argument is presented. The first analytic part argues that models arising from theoretical physics can always be viewed equivalently as self-dual ones. Hence if theoretical physics is to model reality successfully, the true theory must necessarily have the capability of a self-dual formulation. The argument is based on a model of theory and experiment, a model that plays the role that is played by Kant in his theory of judgment. A model such as this is needed in order to define what is meant by "physical thought." The second dynamical step of the argument argues from a model of the interaction of theorists and experimentalists that they would in practice be subject to an urge to evolve towards such self-dual formulations. Within the model we shall then argue that the advancement of physical theories is propelled by the tension arising from the distinction between representation and represented and an aesthetic urge for unification.

In Sec. 4 we argue towards sufficiency of the principle: the limitation to self-dual structures is a powerful one, since there are few such mathematical structures available. Thus with the aid of some subsidiary principles such as "locality," it could serve to pin down a theory uniquely. In familiar applications of "locality," etc., in theoretical physics, a great deal of algebraic infrastructure is presupposed. We argue that the principle of self-duality, being of a very general and algebraic nature, can force this overall algebraic structure in a way that the familiar principles cannot. We find that the structures allowed are necessarily organized in a hierarchical fashion, but that this hierarchy, far from being an infinite regression, is one that rapidly ends in universality.

In Sec. 5 we consider historical evidence for the principle at work. We argue that there is a historical sequence of self-contained physical theories built out of increasingly complex self-dual structures. They include the old quantum theory, recent work on quantum mechanics combined with gravity,[(2),(3),(5),(6)] and recent work on string theory and conformal field theory in the form of Refs. 7 and 8.

The paper concludes with an indication of further work that remains. Since physical thought is, after all, human thought, the principle should be contained in an overall thesis comparable in scope to that of Kant. We would like to be able to derive it from purely logical "intellectual conditions of the understanding." Also, the mechanism put forward in Sec. 3 suggests that it provides a possible framework within which it would be interesting to analyze the history of science.

**Preliminaries:** A convenient language in which to formulate the mathematical constructions is category theory.[(9)] For our nontechnical purposes a category is a collection of objects $S_1, S_2, \ldots$ of a certain structural type and for any two objects a set Mor $(S_1, S_2)$ of "morphisms" or "arrows." These are endowed with properties analogous to "maps from $S_1$ to $S_2$ preserving the structure." Except in Sec. 5 we shall try to keep the discussion as nontechnical as possible in order to concentrate on the philosophical picture.

Since comparisons with Kant will be inevitable, we give now some preliminary arguments on this. For references to Kant we adopt the sympathetic treatment of Ref. 10.

Recall that Kant distinguishes between two broad and mutually exhaustive modes of philosophy. In the trancendental realist mode, more common among theoretical physicists, reality or "things in themselves" are of the same type as appearances or "representations" being those that are independent of humans. In the trancendental idealist mode this conflation is absent. "Things in themselves" are the things being represented and are not to be confused with the outputs of representations. Since we shall make a clear distinction between the outputs of representations and the things being represented, the thesis of the paper is a trancendental idealist one. Note that the identification expressed in the principle of self-duality is of a different nature, namely between structures and the *collection* of all representations of the structure. Moreover, we shall see in Sec. 3.2 that this is an identification that Kant himself makes.

A note about the limited scope of the thesis. We shall be concerned with physical thought, rather than human thought in general. This makes the program above more amenable to a mathematical treatment. This is because theoretical physicists adopt a mathematical framework in formulating theories, so that their structure is more apparent. For a well-known example, in defining quantum mechanics, not only Schrödinger's equation is postulated, but the entire algebraic structure of the act of measurement, etc. It is this axiomatic tradition that makes the thesis tractable by comparison with that of Kant.

The limitation to physical thought rather than all human thought also introduces a complication. A key point in Kant's analysis is that it applies only to human thought and not to all conceivable thought systems, see, for example, Ref. 10, p. 83. The latter are not a real alternative for us as humans. By contrast, the corresponding claim to apply only to physical thought leaves open real choices of other forms of thought. Thus, whereas in Kant, experience is made possible by the conditions of human understanding, for us physical reality will be made possible by the *choice* to think by the conditions of physical understanding. Moreover, once we accept this we must accept the possibility of "subrealities" made possible by "subchoices" within physics as a whole. This will be justified in Sec. 4 where it is expressed in the form of a hierarchical structure of physics. For example, Newtonian mechanics, far from being obsolete, is alive and well but as an idealization or "subreality" of physical reality, which we enter when we choose to work in the limit that all velocities are arbitrarily much less than the speed of light. Thus as well as making claims about the structure of any "ultimate" theory of physics, we aim to include also the structure of its "subrealities" and their relations to each other. Thus we envisage not one Kantian-like thesis, but a hierarchy of them, the axioms of each subtheory being the "*a priori* conditions of understanding" in the context of the subtheory. Let us call the entire structure of any "ultimate theory" and all the subtheories and their relations, the "total theory of physics."

This analysis of the totality of physics is important, because it replaces Kant's analysis in the second half of the trancendental deduction and in the schematism by which experience is made possible as a consequence of the *a priori* conditions of human understanding. See Ref. 10, Chaps. 7 and 8. The difference is that because of the axiomatic tradition of physics, it is possible to make the analysis in more detailed rather than general terms.

## 2. THEOREM OF PONTRYAGIN OR PLATO'S CAVE REVISITED

Plato raised the question, How could some prisoners in a cave confined by chains to seeing only shadows on the cave wall know that they are the shadows of something (Ref. 11, Book 7, 514a)? If the shadows on the wall represent real objects, then how can we recover the reality from the representation of reality? Moreover, if some prisoner insisted that the shadows themselves were the reality, could we prove him or her wrong? When the structure being represented is sufficiently simple, one can study this question literally, as follows. We describe in detail a theorem of Pontryagin and illustrate its content by means of an everyday example. It will be applied more abstractly to the structure of physical theories in the next section.

Recall that a set $S$ has the structure of an *Abelian* (or *commutative*) *group* if $S$ is a set of objects $\{a, b, c, \ldots\}$ with a multiplication defined such that the product $ab$ is in $S$ and $a(bc) = (ab)c$, $ab = ba$, $a^{-1}a = 1$, $a1 = a$ for any typical elements $a, b, c$. Here, 1 is some element of $S$ called the identity, and $a^{-1}$ must be some element of $S$ and is called the inverse.

An example of an Abelian group is, of course, the set of points on a circle. These form a group if one thinks of each point as some clockwise angle from north – the group "multiplication" means adding these angles. We shall denote this "circle group" or "rotation group" by $S^1$. A map $\phi$ which assigns to each element of $S$ an element of $S^1$ in such a way as to respect the group structures is called a *representation*, that is,

$$S \xrightarrow{\phi} S^1, \qquad \phi(ab) = \phi(a)\phi(b)$$

for all $a, b$ in $S$. Thus $\phi$ represents the product of $a$ with $b$ as a product of elements $\phi(a)$, $\phi(b)$ in $S^1$. Note that like the shadows above, it may be that many points in $S$ map under $\phi$ to the same point in $S^1$. The word "representation" is used a little differently in philosophy – for us it is the map itself that is the representation. The purest form of the result that we shall be discussing in more human terms is the following.

**Theorem (Pontryagin)**[1]: The set of all such maps $\phi$, that is, the set of all representations $\phi, \psi, \ldots$ of $S$ is called $\hat{S}$, the *Pontryagin dual* of $S$ and itself has the structure of an Abelian group. The group law is

$$\phi\psi(a) = \phi(a)\psi(a),$$

which specifies what element of $S^1$ the map $\phi\psi$ assigns to any element such as $a$ in $S$. Furthermore, if the group $S$ is not too infinite,[2] then $S \cong \hat{\hat{S}}$. Here, "$\cong$" denotes equality as groups – the elements of $S$ and $\hat{\hat{S}}$ correspond under a suitable dictionary, and this correspondence respects the group structure. One says that such an equality is an *isomorphism* of structures.

This is a remarkable theorem, for it says firstly that the space of *all* representations of the Abelian group $S$ has a structure of the same type, that is, $\hat{S}$ is also in the category of Abelian groups, and then that all the representations of this group of representations of $S$ can be identified with $S$ itself. It means that if some iconoclast insisted that $\hat{S}$ was the reality, then what he would call the representation of his reality, $\hat{\hat{S}}$, would be equivalent to what we call reality, $S$.

To see what is this correspondence between $S$ and $\hat{\hat{S}}$ explicitly, let $a$ be any member of $S$. The corresponding element $A$, say, of $\hat{\hat{S}}$ is that map from $\hat{S}$ to $S^1$ which sends a typical element $\phi$ in $\hat{S}$ to

$$A(\phi) = \phi(a).$$

This map therefore literally turns the tables and says that the object $a$ being seen by $\phi$ can instead be viewed as an observer $A$ (corresponding directly to $a$) seeing $\phi$ with the same value $\phi(a) = A(\phi)$.

One may try to extrapolate from this to more complex structures, for example, "all the ways to represent the structure of the set of all the ways to represent the structure of Fred (meaning the relations between the particles that make up Fred) can be identified with Fred himself." If Fred were an Abelian group, then in addition the set of all the ways to represent Fred would also be an Abelian group, $\widehat{\text{Fred}}$, Fred's dual, maybe called Sally.

Of course, the relations that make up Fred are not the structure of an Abelian group – very few categories possess this property of the category of Abelian groups. To make this theorem and its generalizations a little more concrete, we shall now spell out an example.

Suppose there is a room containing an elliptical hoop, of some shape and some orientation. Let $S$ be the space of all possible hoops (the set of all shapes and orientations). Let various people walk into the room and view the hoop each from some location and some direction. Let us say that each person forms in his brain an impression or image of the hoop, that is, forms some *brain state*. To be perfectly concrete let us take a mechanistic view: the "person" here could for our purposes be a robot; by "brain state" we mean literally some state of his brain cells, which could for our purposes be the voltage distribution in a robot's electrical circuitry. Then the role of $S^1$ is played here by the set of all possible brain states of any one person. Let us assume they have identical brains and states before they first look upon the hoop, so that just the location and direction of the person determines a representation of $S$; any actual hoop $a$ is mapped to some brain state $\phi(a)$ when seen by the person at location and direction $\phi$. So the set $\hat{S}$ of the possible representations is just parametrized by location and direction of the observer.

Next, a psychologist interviews each of the observers. He builds up in his brain a picture of what each of the observers saw. Thus in his brain there is a map $A$, say, from the set $\hat{S}$ to the space of brain states; he builds up in his brain a representation of $\hat{S}$. The generalized theorem asserts that this representation $A$ (which for observer $\phi$ puts in the brain of the psychologist the image $A(\phi) = \phi(a)$, that is, equal to the image that $\phi$ had of $a$) corresponds to $a$ itself. While not literally $a$ (how could a hoop literally be in the mind of the psychologist?) the sequence of images $A$ corresponds to $a$ in the sense that if instead the hoop were actually of shape and orientation $b$, then the resulting sequence of images $B$ would bear a corresponding relation to $A$ as $b$ does to $a$. In other words, it is the hoop type of $a$, the structure that was variously observed, that has been communicated.[3] The element $a$ has been literally communicated to the mind of the psychologist in the sense that for any given sequence of images $A$ there is only one $a$ – deduced by a process of triangulation – such that $\phi(a) = A(\phi)$ for all observers $\phi$.

Clearly, the ability to recover the structure of reality from the representations of the structure by taking the representations of the structure of the space of all representations of the structure is a highly desirable property for a theory of physics. So the structures $S$ should be in some category equipped with a duality operation, that is, such that if $S$ is an object of the category, then $\hat{S}$, the space of all representations, is some dual mathematical structure and $\hat{\hat{S}} \cong S$.

In the case when the category was the category of all Abelian groups (with the slight technical restriction), we saw above that in addition $\hat{S}$ was also an object in the same category. One says that such a category is *self-dual*. This occurence leads to the possibility that some iconoclast could consider $\hat{S}$ as the reality – it is, after all, in the same category of objects – there would not be a logical problem as long as other people accepted that his reality is what we called representations of reality and his representations, $\hat{\hat{S}}$, can be identified with what we call reality.

The trouble comes if there is no way to clearly distinguish between such iconoclasts and other people. In terms of the cave scenario most people might regard the shadows as a representation or acting out of some abstract reality that the theoretical physicists present would try to infer. But others, let us call them the experimental physicists, would think of the evidence – the shadows – as the reality and the theoretical abstractions as nothing more than a way to tabulate or represent the observational data. Our aim in the next sections is to make this more precise and argue that the inherent conflict caused by these points of view requires for its resolution that the reality be self-dual, $\hat{S} \cong S$, so that those who regard $S$ real and those who regard $\hat{S}$ real are, in fact, saying the same thing modulo the dictionary between $S$ and $\hat{S}$ that constitutes their isomorphism.

### 3. NECESSITY OF THE PRINCIPLE OF SELF-DUALITY

In this section we shall refine the considerations at the end of the last section and argue that they lead to the following principle, which, we argue, governs the structure of physical thought.

**Principle of (representation-theoretic) self-duality:** The search for a complete consistent theory of physics is the search for a self-dual structure in a self-dual category.

After justifying why we should plausibly accept the principle, in Secs. 4 and 5 we adopt it as a postulate or hypothesis, and explore its consequences to see if they are in agreement with the actual structure of physical theories.

The strategy for the justification is as follows: in Sec. 3.1 we shall establish the existence of self-dual theories by means of a mathematical trick (so that at any stage there does exist a self-dual formulation to be searched for). We compare the situation with Kant in Sec. 3.2 and proceed in Sec. 3.3 to argue that physics does actually tend to consist in such a search. We work only within a crude model sufficient to establish the plausibility of these points.

#### 3.1 Analytic Argument

We denote by $\check{S}$ the collection of *possible physical states of affairs*, and we take $S$ to be the collection of *theoretical physical concepts* or "questions that could be asked" in a given theory of physics. We denote by $K$ the set of intuitively acceptable concepts in which an answer to a question must lie. By this we mean numbers, say integers, fractions, real numbers, or real numbers modulo $2\pi$ (i.e., the circle $S^1$, as in Sec. 2), or perhaps matrices of numbers. The collection of theoretical concepts $S$ is defined as a collection of maps from $\check{S}$ to $K$,

$$\check{S} = \{\text{states}\}, \qquad \check{S} \xrightarrow{\text{concept}} K. \tag{1}$$

Moreover, this collection $S$ is endowed by the theory with an "algebraic" structure consisting of the theoretical relations between the concepts. For example, if mass and momentum are two concepts, they could be related in the theory by

$$\text{mass}^2(\phi) = \text{momentum}^2(\phi), \qquad \text{for all } \phi \in \check{S}$$

(the Klein-Gordon equation). Here, $\text{mass}^2(\phi) = \text{mass}(\phi)\,\text{mass}(\phi)$ for all states $\phi \in \check{S}$, etc., are further relations. An *experiment* is the evaluation of a question or questions on a state. For example, the determination $\text{mass}(\phi) = 5$ is an experiment.

Note that naively, we would not think that $\check{S}$ also has a structure, since the "real world" would be just one point $\phi$ in $\check{S}$. In practice we refer only to those aspects of the "real world" under consideration in a laboratory. It is the responsibility of *theoretical physicists* to specify $S$, that is, which maps are in the collection and what relations hold. It is the responsibility of *experimental physicists* to decide, in consultation with theorists, exactly what the points in $\check{S}$ are and to arrange for them to come about in a laboratory.

Finally, it is the task of the theorists and experimenters together to arrange that $S$ is the "smallest" collection such that the knowledge of the numbers $a(\phi)$ for all $a \in S$ completely determines $\phi$. "Smallest" refers here to Occam's razor. There should not be redundant concepts, but there should be enough to fix the state of affairs, that is, so that the theoretical model is *complete* enough to encode through the value of all its concepts, the information contained in the actuality of a given state of affairs. This is the *reconstructability property*.

We can now follow the type of reasoning sketched in the preceding section. Thus, what does the reconstructability property mean? The input data is a family of numbers, one for each $a \in S$, that is, a map $\Phi : S \longrightarrow K$. From this family we require that we can uniquely determine a state $\phi \in \check{S}$ such that

$$\Phi(a) = a(\phi), \qquad \text{for all } a \in S.$$

This map $\Phi$ is not any map, but one that could arise in keeping with the theoretical relations. Thus a map that includes assignments $\Phi(\text{mass}) = 2$, $\Phi(\text{mass}^2) = 5$ would not be one that we would require to correspond to a state. We denote by $\hat{S}$ those maps $S \longrightarrow K$ that respect the theoretical relations in $S$. The reconstructability property is, therefore,

$$\hat{S} \leftrightarrow \check{S}, \tag{2}$$

a one-to-one correspondence. The correspondence asserts that we may equivalently view physical states of affairs as maps from $S$ to $K$,

$$S = \{\text{concepts}\}, \qquad S \xrightarrow{\text{state}} K, \tag{3}$$

which is the point of view taken in quantum theory; see, for example, Ref. 13. Given reconstructability, these two are two equivalent ways of thinking about states. For brevity, when we talk about a theory $S$, we mean theoretical concepts $S$ and states $\hat{S}$ or $\check{S}$.

Now, whatever the structure of $S$, $\hat{S}$ necessarily inherits in an analytic or tautological way some structure (*not necessarily* in the same category as that of $S$, although this was the case if $S$ was an Abelian group as in Sec. 2). Likewise, $\hat{\hat{S}}$ will inherit some structure. Moreover, again tautologically, the map $S \longrightarrow \hat{\hat{S}}$ given by $a$ mapping to that $A \in \hat{\hat{S}}$ defined by $A(\Phi) = \Phi(a)$, is necessarily an inclusion $S \subseteq \hat{\hat{S}}$, and one that respects the two algebraic structures. With suitable technical restrictions this will be an isomorphism. These are essentially tautological features that hold in wide generality (the remarkable part of Pontryagin's theorem was not this, but that $\hat{S}$ was also an Abelian group when $S$ was).

Hence these remarks and the above correspondence between $\check{S}$ and $\hat{S}$ endows $\check{S}$ with an algebraic structure, and one such that $\hat{\check{S}} \cong S$. Hence if we think of $S$ and $\check{S}$ together, that is, the set of pairs of the form $(a, \phi)$ where $a \in S$ and $\phi \in \check{S}$, we will have a self-dual object, $S \times \check{S}$. This is because $(S\times\check{S})^{\wedge}\cong \hat{S}\times\hat{\check{S}}\cong \check{S}\times S\cong S\times\check{S}$. The equalities here are the obvious isomorphisms and respect the algebraic structures.

This $S\times\check{S}$ is the collection of concepts of a theory that is equivalent to the original theory with concepts $S$. The states of affairs of the new theory are $\check{S}\times S$, and $K$ is replaced by $K\times K$, which we identify in $K$, since $K$ already includes pairs or matrices of numbers. In other words, a non-self-dual theory can be viewed equivalently as a self-dual one by considering the possible physical states of affairs as concepts, say *experimental concepts*, and including them among our theoretical concepts. Thus, without loss of generality, we need only consider models of the form

$$S=\{\text{concepts}\},\qquad S\ \overset{\text{concept}}{\longrightarrow}\ K. \tag{4}$$

This suffices to prove mathematical necessity, that the structure of a theory of physics is equivalent to a self-dual one. Since we also want to make historical comparisons in later sections, we will next argue that physicists themselves explicitly make this "shift" to a self-dual theory. Before turning to this, we pause to compare the model above to Kant's corresponding theory of judgment.

### 3.2 Kant's Theory of Judgment

Here, we recall Kant's theory of judgment, (cf. Ref. 10, Chap. 4), but in a form suitable for comparison with the above. This formulation is not intended to do justice to all the subtleties of Kant's theory, but to formulate the structure relevant to the paper. Let $O$ denote the collection of all "things in themselves" or "objects." An intuition (in the sense of the process of intuition) is a map from $O$ to a space $I$ of "images" or intuitions (in the sense of results of intuitions). Kant calls these maps or their images "immediate representations." Let $S$ denote the collection of all "concepts." A concept is defined (recursively) as a map from $I$ or $S$ to the set of two elements $\{0,1\}$, a "general representation." We have

$$S=\{\text{concepts}\},\qquad O\ \overset{\text{intuition}}{\longrightarrow}\ I\cup S\ \overset{\text{concept}}{\longrightarrow}\ \{0,1\}. \tag{5}$$

A *judgment* is the evaluation of a concept on an element of $S$ (i.e., on another concept) to obtain 1. Thus the judgment "all bodies are divisible" is expressed by us as "divisibility(body) = 1." Now, $S$ has some algebraic structure defined by logic. Thus the concepts "cup," "saucer," and "cup and saucer" are related by "and." The relations are reflected by any concepts acting on them. Thus "divisibility(cup and saucer) = divisibility(cup)divisibility(saucer)." Thus concepts (as maps) "represent" the structure of $S$. It would therefore be better to denote the collection of these "concepts as maps" as $\hat{S}$. The fundamental conflation of Kant's theory of judgment is then the conflation of $\hat{S}$ with $S$ by deliberate identification.

We see here that Kant's theory of judgment is, loosely speaking, a self-dual one. The role of $K$ is played by the set $\{0,1\}$ or $\{$"false," "true"$\}$. One difference is the notion of intuition, absent in the model for physical theories above. Intuitions are needed by Kant to serve as a ground for the determination of concepts, because, it is assumed, a closed circular system could never have any "objective validity." This does not seem to be necessary in our model of physical judgment. This is because in physics it seems to be sufficient to leave the determination of concepts to be grounded only in consistency: experiments can never prove anything absolutely, only verify consistency between the parts of the theory involved in setting up the state with the parts of the theory consisting of concepts being tested. This does leave a problem. Whereas for Kant, the collection $S$ can be validated objectively, in a completely circular system there would appear to be many satisfactory theories $S$, each consistent. We address this in Sec. 4 and indicate the resolution in Sec. 6.

### 3.3 Dynamical Argument

In this section we aim to show that if physicists initially agreed on a non-self-dual theory, then they would themselves experience an urge to shift to a self-dual theory along the lines above, as a consequence of the interaction between theorists and experimentalists. This is needed if the principle of self-duality is to be a (crude) historical guide as well as a merely mathematical prediction of structure.

We (crudely) model the interaction as follows. Suppose that theorists and experimentalists have been pursuing their tasks as above, but convene once a year at a special conference in Vienna. The goal of these talks is to review and agree on the structures $S$ and $\hat{S}$ of the overall theory of physics.

Now, since the last conference, experimenters have been busy creating the elements of $\hat{S}$ and exploring their relations with each other. In doing this, they are working with the elements of $\hat{S}$ as "experimental concepts." For example, they have just created an exotic state "of sinusoidally varying position." This also means that they have created the concept of a state of sinusoidally varying position. The concept has value 1 on $\phi$ if $\phi$ is such a state, and otherwise has value 0. This is the associated "experimental concept," but it is not yet an element of $S$ (a theoretical concept): it is up to theorists to determine the relations between this possibly new concept and existing theoretical concepts. They will try to show that the new concept can be fully understood or "explained" by these relations in terms of existing concepts, that is, that it is not a new concept in $S$.

Thus at the following talks the experimenters are going to present these new exotic states as possibly interesting new concepts for theorists to explain. If $\hat{S}\cong S$, then some theorist will be able to point out the old concept in $S$ corresponding to any possible new concept in $\hat{S}$. The output of the talks is then $S^{\text{new}}=S$. If, at the other extreme, all experimental concepts $\hat{S}$ are offered and do not have any obvious explanation, the theorists will have to accept $S^{\text{new}}=S\times\hat{S}$. In general, $S^{\text{new}}$ will lie in between these extremes.

A similar phenomenon goes on for the states of the theory. Theorists have been busy studying the elements of $S$ and exploring their relations. In doing this they privately think of their concepts as "theoretical states of affairs," for example, "it is the case that the Klein-Gordon equations hold." More precisely, they imagine that there is a state such that the value of $\text{mass}^2-\text{momentum}^2$ is, in this case, zero. This is not yet an element of $\check{S}$ (an experimental state): it is up to experimenters to determine if this proposed state can be created in their laboratories, that is, to construct it explicitly as a representation of $S$ (an element of $\hat{S}$ or, equivalently, an element of $\check{S}$). Of course, it is expected that such a possibly new state is a combination of existing states in $\hat{S}$. At the talks the theorists are going to present these concepts as possibly new states. If $S\cong\hat{S}$, it will be easy to identify the corresponding known states and the output will be $\hat{S}^{\text{new}}=\hat{S}$. At the other extreme it will be $\hat{S}^{\text{new}}=\hat{S}\times S$. In general, $\hat{S}^{\text{new}}$ will lie in between. What will have to be hammered out at the conference is to make sure that $\hat{S}^{\text{new}}=(S^{\text{new}})^{\wedge}$, that is, to check that the new concepts can reconstruct the full set of states, and the new states respect the relations in the full set of concepts.

Clearly, these two transference phenomena go hand in hand. A concept can be experimentally verified, so lead to the acceptance of a new element of $\hat{S}$ which then leads to further, perhaps new concepts. It might be thought that, in fact, no new elements could be created this way, for example, that any possibly new concept arising from an element in $\hat{S}$ could always be reduced to a combination of concepts already in $S$. This would be the claim that all experiments are analytic judgments. Even if this were true, it fails in practice because the structures $S$ and $\hat{S}$ are the structures of the entire theory while most physicists, as specialists, are only concerned with their part of the theory. It is easy for a concept to appear new by forgetting relations that would, in fact, render it equivalent to existing ones. Once the new concept is accepted and integrated by theorists into $S^{\text{new}}$ and representations $(S^{\text{new}})\hat{}$ have been constructed, then it is illogical to go back and correct any such error. This is because by this stage the relations in the old $S$ no longer hold exactly on some of the new states. One could still return to $S$ and $\hat{S}$ by renouncing *both* the new concepts and the new states, but this would be viewed as a matter of choice (like the Newtonian limit) rather than a reason not to accept the new theory.

Realistic examples of this transference phenomenon (in contrast to the figurative examples above) are necessarily technical. For example, consider the concept of *momentum*. In the Newtonian theory *position* is the main concept. We denote it $\mathbf{x}$. The concept of momentum, which we denote $\mathbf{p}$, is derivative (not really independent): $\mathbf{p}$ = mass $\times\ \dot{\mathbf{x}}$. Now, working within Newtonian mechanics, experimenters construct waves $\phi_p$ of given momentum $p \in K$, that is, such that $\mathbf{p}(\phi_p) = p$. These are Newtonian states, but exotic ones consisting of many particles. They are experimental states motivated by theoretical concepts (mathematically the wave states $\phi_p$ are representations of the additive structure of Newtonian space expressed equivalently in additional structure on the algebra $S$ of position coordinate functions $\mathbf{x}, \mathbf{x}^2, \mathbf{x}^3, \ldots$, etc., on the Newtonian space). Now, as the subject of continuum mechanics develops in its own right, it is easy to loose sight of its analytic origins in Newton's laws. In particular, the usefulness of the concept of momentum $\mathbf{p}$ (proved useful because it is conserved) encourages us to think of it as independently defined as a concept $\mathbf{p}$ such that $\mathbf{p}(\phi_p) = p$. This then encourages theorists in developing a new theory $S^{\text{new}}$ to include both $\mathbf{x}$ and $\mathbf{p}$ as independent concepts, or at least ones with weaker relations than in the Newtonian theory. Momentum $\mathbf{p}$ is now a new theoretical concept inspired by experiments with waves. The new $\mathbf{x}$ and $\mathbf{p}$ coincide with the old concepts when evaluated on classical waves, but theorists now anticipate possible exotic non-Newtonian (in this case, quantum) states on which the old relations do not hold. When such exotic representations of $S^{\text{new}}$ have been constructed, the new theory is considered necessary (to deal with these new states).

Returning now to our crude model, if the output of the talks $S^{\text{new}}$ is not self-dual, then it will be changed on the next meeting. Thus theorists will only agree over a period of time if the theory becomes self-dual. This concludes the necessity argument.

It should be mentioned that this, however, is not the only criterion. If the theory is of a form like $S \times \hat{S}$, then another principle will urge theorists to modify the theory. This is the principle of *unity*. A theory consisting of two-independent pieces is not satisfying for most theorists. Likewise, a collection of states of the form $\hat{S} \times S$ is unsatisfying for most experimenters, because it corresponds to two independent concurrent "universes." If the urge to unify is accepted (for whatever reason), then an output of the form $S \times \hat{S}$ for $S^{\text{new}}$, although self-dual, will continue to evolve. Theorists will tend to introduce relations between $S$ and $\hat{S}$ to form a new theory $S^{\text{new}} = S \bowtie \hat{S}$, say. Experimenters will be looking at $(S \bowtie \hat{S})\hat{}$ and seeking to modify $\hat{S} \times S$ to agree with this. A mathematical model of how such structural modifications might be made, the theory of bi-cross-products, is developed

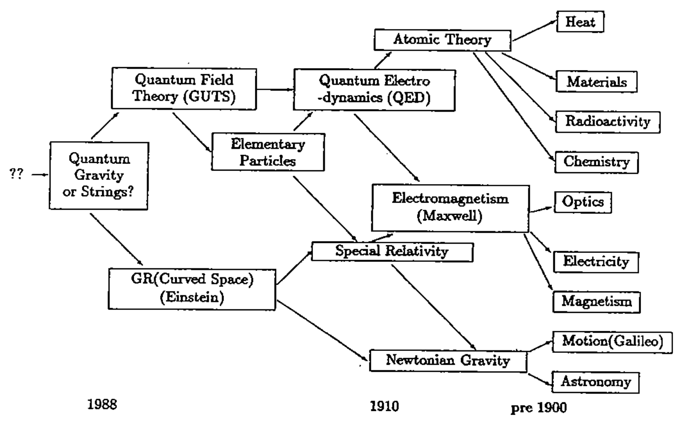


Figure 1. Historical breakdown of theoretical physics. Arrows indicate how the laws of one area of physics may be deduced as a limit of more abstract ones. Assessment of the author.

in Ref. 3, Sec. 1.1 in connection with ideas of Ref. 14, II.v3 and Ref. 15, p. 54. Typically these modifications will destroy self-duality, and so at the next talks the theory will have to keep evolving and growing.

This aesthetic urge to unify occurs along with the urge to eliminate redundant concepts already discussed. Thus we see that the principle of self-duality and the aesthetic urge to avoid $S \times \hat{S}$ form a two-stroke "engine" that propels the evolution of physics towards a self-dual indecomposable theory. In the first stroke $S$ expands to $S \times \hat{S}$. In the second stroke it is reduced and perhaps modified to $S^{\text{new}}$. On achieving a self-dual theory the engine "stalls." Such an occurrence in our model signifies a stagnant period, such as at the end of the last century, when physicists feel that their task is done: a complete self-consistent theory which there is no particular urge to change. This completes our crude model of the origin of the principle of self-duality in the methodology of physics.

## 4. SUFFICIENCY OF THE PRINCIPLE OF SELF-DUALITY

In this section we assume that the principle of self-duality holds and along the lines of the preceding section. We examine the broad implications of this for the structure of the totality of physics. Recall from the introduction that this means not just a theory $S$, but its various limiting subtheories and their relations. We aim to show that the principle is mathematically and sufficiently powerful enough to force the allowed pattern to be of the general type actually known. We are concerned in this section only with the nature of this constraint: the practical correspondence with physics is the subject of the next section.

The observed structure of the totality of physics is estimated in Fig. 1. Boxes indicate well-known theories of physics and an arrow maps to a theory that has axioms given as limit, that is, a special case of the axioms of the theory at the source of the arrow. It is not our intention to defend this assessment here. The diagram merely summarizes the overall picture to be found in, for example, the text of Ref. 16. Of course, what physicists actually do is more subtle (and more ad hoc) than any simple principle. We cannot expect agreement on anything but the most broad aspects of the global structure. These are, (1) Theories are nested in a tree of successively more accurate theories. (2) Different theories tend to become unified. Indeed, this unification in the present century has succeeded so well that, as is well known, there are now just two paradigms, quantum mechanics and general relativity. If these can be combined into a single successful theory (possibly string theory) such that they correctly appear as limits, then physics as we know it will be finished. As stated in Ref. 17, the end of physics is in sight.

These two essential features are implied by the principle of self-duality (in the model of Sec. 3) as follows. (1) Shifts of the form $S^{\text{new}} \subseteq S \times \hat{S}$ (or a modification $S^{\text{new}} \subseteq S \bowtie \hat{S}$) are forced upon us by the urge for self-duality. Therefore, theories keep improving, each new theory containing the old theory $S$ and other aspects $\hat{S}$, that is, a tree structure. Note that in $S^{\text{new}}$, $\hat{S}$ appears as a subtheory, albeit one that originally arose from experimental concepts associated to the subtheory $S$. Apart from these unifications of duals, there may also ad be hoc unifications of similar structure between subtheories $S_1, S_2$ to give $S_1 \bowtie S_2$, say. These are not implied by the principle of self-duality, but might occur anyway and add to the tree structure. (2) Those unifications that arise from the principle of self-duality involve a radical shift in language to much more general or abstract structures. This is because $S$ and $\hat{S}$ are (if $S$ is not self-dual) of very different categories; any category in which $S \times \hat{S}$ (or a deformation $S \bowtie \hat{S}$) lie is rather more general than that of either $S$ or $\hat{S}$. Hence each advancement of this type involves a radical generalization of the category, that is, such unifications are "quantum leaps." This mechanism can be contrasted with, say, Ref. 18. Finally, because of this, after just a few steps the categories are so general that the evolution stops because we become unable to conceive of more general structures.

The important point about the above argument is the claim that it follows only from the principle of self-duality. The best way to see this is to ignore any physical considerations and simply ask after the mathematical consequences of the principle of self-duality. Thus, to try to prove the result mathematically, let us take the property of self-duality as a definition of a category $\mathcal{P}$. What does $\mathcal{P}$ look like? We have a relatively precise mathematical problem to solve. It turns out to be both highly restrictive – most structures are not self-dual – and yet admits a rich variety of solutions. The structure of $\mathcal{P}$ is estimated in Fig. 2.

Known self-dual structures are shown along the central axis. An arrow indicates that the category at the source arose naturally as a generalization of the category at the head. The way that these structures arise follows the pattern above: structures $S$ in a self-dual category are deformed and inspire the definition of a non-self-dual category. This is then unified in a more general category, general enough to include both the deformed category and its dual. In addition, these more general categories inspire various offshoots, of which a few relevant ones are shown. Since some mathematicians work very abstractly, but without examples (i.e., ahead of their time), time in the diagram is not intended to be exactly the historical order of first definition. Rather, the diagram is a structural one depicting the logical relations between the axioms.

The diagram expresses the mathematical consequences of the principle of self-duality, as it is understood in mathematics as we know it. It is to be compared in general terms with Fig. 1. The treelike structure is clearly exhibited, as too is the claim that categories rapidly become very general: the left-most category is the category of (small) functored monoidal categories. We conclude that as far as is known, the totality of structures allowed by the principle of self-duality is of a similar nature to the apparent structure of the totality of physics. There might still be other self-dual structures not yet known: if mathematicians found a chain of such self-dual structures very different from Fig. 2, the principle of self-duality would assert that

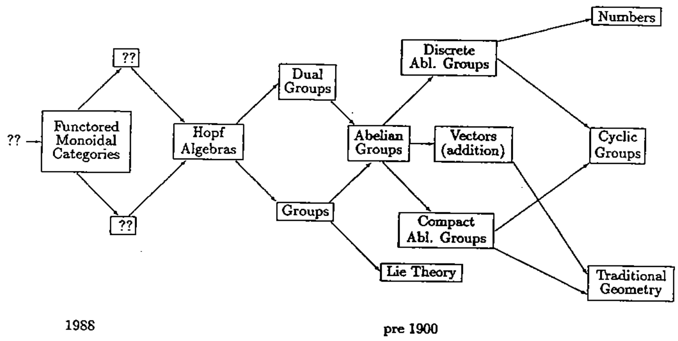


Figure 2. Known self-dual mathematical structures. Self-dual categories are on the central axis. Off center are non-self-dual deformations of a self-dual category unified by a more general self-dual category. Further out are some relevant spin-offs. Arrows indicate that axioms of one category are viewed as a restriction of more general ones. Assessment of the author.

it corresponds to a development of physics. This would necessarily be very different from physics as we know it (because it would have to be based on very different structures from those familiar in physics).

Note that in each self-dual category there may be many objects. Some of these will be self-dual or of self-dual type. According to the principle of self-duality, it is these that should more specifically correspond to self-contained (i.e., complete) theories of physics. These should be stagnant, having no particular reason to evolve according to our model. For aesthetic reasons they could slowly evolve anyway into non-self-dual deformations which would then be unified into another self-dual theory.

## 5. DETAILED HISTORICAL EVIDENCE

In this section we show that there is a historical sequence of successively more complex theories of physics built on successively more complex self-dual structures.

### 5.1. The Simplest Self-Dual Structures

The simplest self-dual category known is of Abelian groups as we showed in Sec. 2. What are all the self-dual structures $S$ in this category? The answer is simply the following: (1) the "clock groups" (2) the addition of vectors, and (3) pairs consisting of an Abelian group $S$ and its dual $\hat{S}$ independently (Ref. 12, Sec. 12.1).

According to the principle of self-duality, these should correspond to the simplest of the complete theories of physics or paradigms. Certainly, (1) clocks are among the earliest structures identified in natural philosophy, and, certainly, (2) it may be argued that the main structure of physics in the last century was that of a vector space. The special case $\boldsymbol{R} \times \hat{\boldsymbol{R}}$ was particularly important and corresponds to the (position, momentum) of a free particle in one dimension in mechanics. The so-called "old quantum theory" of Planck and Bohr was the next innovation and featured *quantization*, that is, discretization of momentum. An example is a particle in a crystal. For such a particle the position in space is effectively the circle group $S^1$. Its dual is the group of whole numbers, $\hat{S}^1 = \boldsymbol{Z}$ and the physical (position, momentum) is an element of $S^1 \times \boldsymbol{Z}$, which is the simplest example of type (3).

Whether or not this is a fair picture of the central structures of the early complete theories of physics is a matter of debate. The word "complete" here meaning "theoretically self-contained" is essential, for the method of example (3) shows that any structure at all can be *part* of a complete theory, so long as its dual structure is also included. In truth, physics before 1900 was in practice to a large extent descriptive, that is, explaining data without much emphasis on considerations of mathematical structure. Consequently, one must turn to the physics of this century to see the principle of self-duality more definitively at work. We now turn to this.

## 5.2 The Next Simplest Self-Dual Structures – Hopf Algebras as the Structure of Quantum Mechanics and Gravity

The next simplest self-dual category known is *Hopf algebras*. For details, see Ref. 19 and Ref. 20, Sec. 1, for example. According to our principle of self-duality, the self-dual structures in this category should be the central structures of the next-most-complicated complete theories of physics after, say, the "old quantum theory." The next-most-complex physics that we might hope to understand this way is modern quantum theory and then gravity. A simple form of modern quantum theory is modern quantum (statistical) mechanics. We shall need the abstract-algebra formulation of Ref. 13, Chap. 2, which is equivalent to the usual concrete one as operators on a Hilbert space, once the vacuum state has been chosen.

Indeed, abstract algebraic structure is well known as the foundation of quantum mechanics. The observables are the (self-adjoint) elements of an algebra $S$. Now, a Hopf algebra is an algebra $S$ with some further structure $\Delta$ called the *coproduct*. This coproduct gives $S$ a *coalgebra* structure in addition to its existing algebra structure, and a Hopf algebra requires these to be compatible in a natural way.

Indeed, it turns out that this additional structure $\Delta$ encodes space-time geometry, and the compatibility equations turn out to be somewhat comparable to Einstein's equations for the gravitational field. Ordinarily this geometrical information is expressed in terms of Hamiltonians and Schrödinger's equation, but it has been shown by the author, at least in some simple toy models, that these amount to the structure $\Delta$. Although it is not well known to physicists, this structure $\Delta$ belongs to a well-established tradition in mathematics known as algebraic geometry, dating perhaps from works of Grothendiek, Zariski, and a theorem of Gelf'and and Naimark.

In the simplest toy model investigated in Ref. 6, $S$ is the algebra of quantum mechanics of a particle in one-dimensional flat space. This well-known algebra, "the canonical commutation relations algebra," is generated by the well-known canonical commutation relations $[\mathbf{p}, \mathbf{x}] = i\hbar$. It can be shown that this algebra admits no compatible $\Delta$ of similar but dual structure. This means that for $S$ to be a Hopf algebra of self-dual type as the principle of self-duality suggests, then ordinary flat-space quantum mechanics must be modified. By modifying the canonical commutation relations, the motion of a freely moving particle is also modified, and so it is found that gravitylike forces are induced by the self-duality requirement.

In summary, ordinary flat space quantum mechanics as a description of matter can never satisfy the principle of self-duality. Instead, gravitylike forces are inevitably required to balance off the quantum structure and thereby maintain self-duality.

This Hopf-algebraic approach to quantum mechanics combined with gravity provides simple toy models in which quantum mechanics and gravity are unified into one mathematical structure. We know that our more usual intuitive geometrical notions of space-time break down on a microscopic scale where quantum effects are dominant, leading to various conjectures about the "foamlike structure of space-time" at such small scales. So, more abstract language is genuinely needed by theoretical physicists.

In this Hopf-algebraic approach to quantum mechanics combined with gravity, the duality properties of Hopf algebras take on a definite physical interpretation that one might reasonably expect to hold also in more realistic models. So, as we did for Abelian groups in Sec. 2, we now demonstrate in detail what the principle of self-duality amounts to in this context. To do this we need to describe the duality in more detail.

The $S$ denotes the quantum mechanics algebra, the "algebra of observables." The normalized state vector $|\phi\rangle$, say, determines a map $\phi$ from $S$ to ordinary numbers. It assigns to any observable $a$ in $S$, its average expectation value $\phi(a) = \langle\phi|a|\phi\rangle$. Since only such expectation values are, effectively, all that we actually measure in quantum mechanics, it is the map $\phi$ itself which is the *state*. One calls all linear maps[4] $\phi$ from $S$ to ordinary numbers, "states." The $\phi$ shown above is a "pure state," but more general states are "mixed states" of the form

$$\phi(a) = \langle\phi_1|a|\phi_1\rangle t_1 + \langle\phi_2|a|\phi_2\rangle t_2 \ldots + \langle\phi_n|a|\phi_n\rangle t_n$$

for some $n \geq 1$ and $t_1, \ldots, t_n$ positive and adding up to 1. Thus *states* represent the algebra $S$ in terms of ordinary numbers, the expectation values. In fact, given such an abstract state $\phi$ one can reconstruct an entire algebra of operators on a Hilbert space, representing $S$ and such that $\phi$ is the vacuum state. This is known as the Gelf'and Naimark Segal (GNS) construction (Ref. 13, Chap. 3). Thus mathematically, if $S$ is the algebra of observables of the quantum system, then $\hat{S}$ is the set of states.

Now, because $S$ is an algebra, $\hat{S}$ is automatically a *coalgebra*. This means that is has a map $\Delta$ that "shares out" or "comultiplies" elements of $\hat{S}$. If $\phi$ is an element of $\hat{S}$, then $\Delta\phi$ is a string of pairs of elements of $\hat{S}$. More precisely, it is an element of the tensor product $S \otimes S$, conveniently written $\Delta\phi = \sum \phi_{(1)} \otimes \phi_{(2)}$. Explicitly, the coproduct of $\phi$ is defined by

$$\sum \phi_{(1)}(a)\phi_{(2)}(b) = \phi(ab).$$

Our principle of self-duality says that in a complete theory of physics the structure of $\hat{S}$ must be the same as that of $S$ (i.e., up to a dictionary). So, for this to be possible, $S$ must itself have a coalgebra structure in addition to and compatible with the existing algebra structure. This is why $S$ should be a Hopf algebra.

In a similar way, this coalgebra structure that our principle requires on $S$ now defines automatically an algebra structure on $\hat{S}$. Explicitly, the product of $\phi$, $\psi$ is defined by

$$(\phi\psi)(a) = \sum \phi(a_{(1)})\psi(a_{(2)}),$$

where $\Delta a = \sum a_{(1)} \otimes a_{(2)}$. So $\hat{S}$ is also a Hopf algebra as we wanted, called the dual Hopf algebra of $S$.[5]

Now we said that the coalgebra structure on $S$ corresponds to geometry. Thus if some iconoclast were to insist that $\phi$ in $\hat{S}$ should be thought of as the observable and $a$ in $S \cong \hat{\hat{S}}$ should be thought of as the state, writing $a(\phi)$ for the same number that we would write $\phi(a)$, then what he calls the algebra structure of his quantum mechanics would correspond to what we call the coalgebra structure of our geometry, and vice versa. Moreover, if $S$ is self-dual, $S \cong \hat{S}$, then we and the iconoclast are merely using very different words to describe equivalent mathematical algebra and coalgebra structures.

Thus in Fig. 3 the left slope where quantum effects are dominant – where

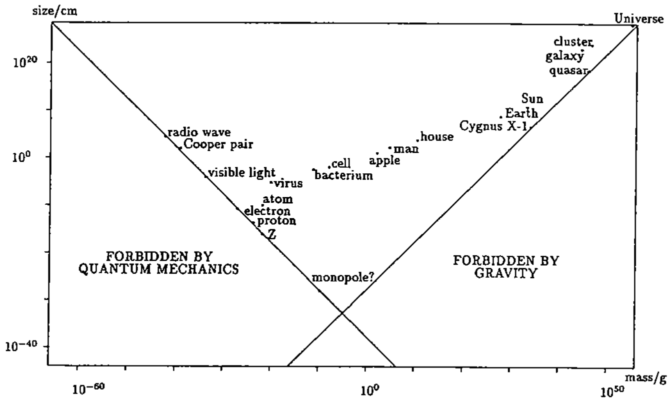


Figure 3. The range of physical phenomena. Plotted is typical mass-energy vs size of objects in the universe.

one finds *elementary particles* – would be for our iconoclast the right slope where he would say that in his view gravitational effects are dominant, and vice versa. It is interesting that on both slopes one finds particlelike objects; on the right slope one finds, according to most astronomers, black holes, for example, the object called Cygnus X-1. A number of theorists have occasionally speculated[6] that one should be able to transform our view of the world such that black holes literally become elementary particles, and indeed this is precisely what our iconoclast would be doing.[7]

In summary, the presence of gravity was dictated by the principle of self-duality in such a way as to maintain the quantum-gravity symmetry depicted by left-right symmetry in Fig. 3. In the present context this principle amounts to the requirement of maintaining complete symmetry between observables and states in quantum mechanics. Note, also, that in Fig. 3 the upper boundary asserts that the universe is closed, that is, has a finite size. As a result, all physical objects lie in the triangle shown.

### 5.3 More General Self-Dual Structures – Monoidal Categories as the Structure of String Theories

The next most complex self-dual category currently known is that of the category of (small) *functored monoidal categories*. A monoidal category $\mathcal{C}$ is one equipped with a "tensor product" $\otimes$ between objects. Classic examples are the categories of vector spaces, representations of groups, and representations of Hopf algebras. See Ref. 20, Sec. 7 for a review. A functor between monoidal categories $F : \mathcal{C} \to \mathcal{V}$ is, roughly speaking, a map of objects and morphisms in $\mathcal{C}$ to objects and morphisms in $\mathcal{V}$ respecting the category structures and $\otimes$. By a functored monoidal category we mean such a triple $S = (\mathcal{C}, F, \mathcal{V})$.

Motivated precisely by the principle of representation-theoretic self-duality described in the present paper, we defined in Ref. 8 the notion of a representation of a functored monoidal category $S$ and showed that the collection of representations, $\hat{S}$, forms again a functoral monoidal category $\hat{S}$. It takes the form $\hat{S} = (\mathcal{C}^{\circ}, F^{\circ}, \mathcal{V})$ (Ref. 8, Theorem 3.3). So the category of (small) functored monoidal categories is self-dual[(8)] as was depicted in Fig. 2.

All the categories and dualities that we have talked about previously in this section are contained in this very general duality of functored monoidal categories for the special case $\mathcal{V}$ = Vec. (Here, Vec is the category of vector spaces.)

Independently of these representation-theoretic self-duality considerations, it has recently been found that the problem of the construction of a rational conformal field theory (a class that includes string theories) appears to be equivalent to the problem of construction of a monoidal category of a certain type. This is recent work of Moore and Seiberg. See Ref. 20, Sec. 7.6 for a relevant introduction. In these theories there is a structure, the *chiral algebra*, which plays the role of a maximal "group of symmetries." We have recently argued in Ref. 7 that such a chiral algebra can be viewed essentially as an example of a certain kind of generalized Hopf algebra. Moreover, these generalized Hopf algebras correspond to examples of functored monoidal categories of the form $(\mathcal{C}$, identity, $\mathcal{C})$. These generalized Hopf algebras are therefore a candidate for the lower unfilled box in Fig. 2 (and their duals in the upper box). This is an example of interesting mathematics arising out of physics and conversely may provide a precise mathematical formulation of chiral algebras. This is a topic of current research. Other physics also possibly included here is Ref. 22 and its generalizations.

The principle also predicts that among rational conformal field theories (of which there are a plethora), physicists should look for representation-theoretic self-dual ones in the form of ones corresponding to self-dual functored monoidal categories. We note that physicists have considered various duality and self-duality restrictions in conformal field theories, but so far not one of the representation-theoretic type suggested by the principle.

The physics covered by string theories and some other conformal field theories also aims to include quantum theory and gravity, but in rather more general form than the simple models covered by Hopf algebras in the preceding subsection. Namely, they should unify quantum field theory, not merely quantum mechanics, and more general Riemannian geometry, not merely the simplest curved spaces. It is hard to conceive of radically more general self-dual categories than the category of functored monoidal categories. To be radically more general than such categories, we may expect to need a notion radically more general than that of category itself. Such a notion would be hard to formulate in mathematics as we most often think of it (where structures are most often formulated implicitly in terms of category theory). Such an eventuality corresponds then to the belief among physicists[(17)] that the end of physics as we currently think of it is in sight.

## 6. CONCLUDING REMARKS

In this section we conclude with a brief discussion of the ontological status of the totality of physics implied by the thesis of the paper. Recall that unlike Kant, we did not use in Sec. 3 any notion of "things-in-themselves" $\mathcal{O}$. In Kant these are required in order, via intuitions, to objectively validate concepts. Instead, in physics, concepts in $S$ are validated via experiments by $\check{S}$ while elements of $\check{S}$ in the form of $\hat{S}$ are validated by $S$. This is already a circular system. Further, in a self-dual theory $\check{S}$ and $S$, that is, experimental and theoretical concepts, are explicitly identified.

It might be thought, then, that in a circular situation anything would be possible. We have argued in Sec. 4 that this is not so precisely because of the additional self-duality constraint, that is, that consistency of the self-dual system is such a powerful constraint that it forces a rigid structure for the allowed theories, and one that resembles the totality of physics. Hence the position on realism implied by the sufficiency argument is far from the conventionalism of, say, Ref. 23. Rather, there is a definite physical reality being uncovered by physicists, in the form of a definite underlying structure for the totality of physics, but this definite structure is, we have argued, no more and no less than the mathematical implication of the principle of self-duality. The principle of self-duality is therefore like an "axiom of physics," and one that is an *a priori* condition in the conception of physics as we know it. From this point of view the "reality of physics" experienced by physicists is *created* by this axiom, in a similar way, for example, that the reality of low-dimensional topology is created by the axioms of a manifold. Once we take on the axioms of a manifold, all the rich structures of low-dimensional topology (the various types of surfaces, knots, etc.) follow mathematically. The main difference is that in studying low-dimensional topology the axioms are, at least initially, stated explicitly. By contrast, in physics the "axiom" in the form of the principle of self-duality is taken on unconsciously in the framework of physical thought. Both the axioms of a manifold and the framework of theory/experiment in the form of the principle of self-duality are matters of choice. If accepted, they "create" the reality of their respective subjects. This does not of course mean that either low-dimensional topology or physics is empty or arbitrary. In each case there are many volumes to be written studying and cataloguing the allowed structures and their relations.

Likewise, we can conceive of other forms of thought, not resembling fundamental physics, but rather created mathematically by other very different "axioms." These various choices constitute the structure of the reality not of physics, but of mathematics as a whole. It might be that in a similar fashion, this mathematical reality is itself created by yet more general *a priori* conditions, and so on. Such a thesis is beyond the scope of the mathematical treatment of the paper, but is of a scope comparable to that of the thesis of Kant. It may also be possible to adopt here the work of Ref. 24 in resolving this issue. This is one direction for further study.

Returning now to physical reality, Fig. 2 (in so far as it applies to Fig. 1) expresses the same argument applied *within* physical reality. Thus within the assumption of the principle of self-duality is a further choice of self-dual category. Say we choose functored monoidal categories/string theory. Within this there are various choices of subtheories. These are part of the subreality of the theory of functored monoidal categories/string theory. Adding more and more axioms, that is, progressing to the right in Fig. 2 or Fig. 1, creates further subrealities. All this follows once we accept that physical reality is created mathematically from an axiom. If this is the mechanism by which physical reality is created, then there is not one reality, but many, one for each of the subtheories $S$, one for each of its subtheories, its sub-sub-theories, etc., until the point where we choose to perceive a fact in a given situation. It seems appropriate to call this picture of reality *relative realism*. In it reality is created as the structure of all the well-defined consequences of a choice, but at each stage the catalogue of allowed choices is part of a more general reality. Moreover, far from an infinite regression, this does not indefinitely defer the question of what is reality. Rather, as in Fig. 2, the nature of the choice becomes in a few steps so general that the choice becomes difficult to conceive otherwise. Notice that we do not attempt to disprove a notion of an absolute reality of "things in themselves," only that such a notion is not necessary.

It should be admitted that the observations in this section are not particularly new. The difference is that we have attempted to prove the position by identifying the choice to be made for the reality of physics (within mathematics) and attempted to show that it does create a reality resembling the reality of physicists.

To summarize, returning to Fig. 3 we note that the objects that make up everyday life, from viruses to houses, are almost exactly in the center of the allowed area. From the usual point of view the reason for this remarkable fact is that as we approach either of the two forbidden zones, the structures must get simpler, so close are they to not being allowed at all. Therefore, the most complicated structures, where life would be most likely to develop, are in the center as far away as possible from the two forbidden regions. By contrast, we have derived these fundamental laws of quantum mechanics and a toy kind of gravity in Sec. 5 from the principle of self-duality. Thus assuming this can be carried through to more realistic models, it would be better to say, *We are in the center in Fig. 3 because we created physics around ourselves*. To start with we wanted concepts that were useful in everyday life (such as clocks and vectors). Useful here means repeatable or, equivalently, verifiable. However, as we progressed to invent and incorporate more and more abstract structures while subject to a constraint of theory/experiment verifiability, we soon found that we had boxed ourselves in as in Fig. 3.

Clearly, much more work is needed in refining the principle of self-duality and the associated analyses of Secs. 3, 4, and 5 if these general conclusions are to be fully justified. In the meantime, the principle of self-duality and the philosophical point of view described here may be taken as a position within which such an analysis, both structural and historical, can be undertaken.

**Acknowledgment**
There are many whom I would like to thank for encouragement and advice during the various drafts of the manuscript. Among them are E. Seymour, K. Arnaud, A. Miller, and S. Odiari. I was supported in part by a Herchel Smith scholarship. I also thank the University College of Swansea and SERC for support during revisions.

Received 21 June 1990.

**Résumé**
*Nous avançon que la recherche d'une théorie ultime de la physique peut être modelée mathématiquement comme la recherche d'un système algébrique ayant la propriété que l'ensemble de toutes ses représentations forme une algèbre isomorphe à l'original. Les exemples plus simples d'algèbres douées de cette propriété sont identifiés ici avec les théories physiques correspondantes. L'environnement philosophique est développé dans la forme d'une thèse à la Kant.*

## Endnotes

[1] Compare Ref. 12, Chap. 12.1.

[2] The group should be what is called locally compact.

[3] In the case when $S$ was a group – it is no longer – the precise statement is $AB(\phi) = A(\phi)B(\phi) = \phi(a)\phi(b) = \phi(ab)$ where the first equality is the definition of multiplication in $\hat{\hat{S}}$, the second is the definition of the correspondence $\hat{\hat{S}} \cong S$, and the third is the property that $\phi$ is a representation of $S$. More generally, what is represented is not a group structure, but more complicated relations between the elements of $S$.

[4] That is, all linear maps obeying the further technical condition of *positivity*. Similarly, only the *self-adjoint* elements $a$ of $S$ are considered observables. Although physically important, both these technical restrictions are not at all relevant to the present discussion, so we shall gloss over them for now.

[5] This Hopf algebra duality in fact correctly generalizes the notion of Pontryagin duality when one looks at a certain subcategory of Hopf algebras corresponding to groups or their duals. See, for example, Ref. 3, Sec. 1. Hence models of quantum mechanics combined with gravity using Hopf algebras generalize models based on groups discussed in the previous subsection.

[6] For a recent example cf. concluding remarks in Ref. 21.

[7] As a by-product we remark that this also establishes an interesting correlation between Boolean algebra self-duality and the respresentation-theoretic self-duality of the present paper. Namely, logic as the crudest model of reality has a well-known symmetry in which the use of the words "exist – not exist," "and – or," "everything – nothing" etc. are interchanged. Such a dual picture would be logically equivalent to our picture (by De Morgan's theorem). It can be argued that although at first sight this self-duality seems to be lost in physics (e.g., "apples" curve space but "not-apples" do not), it does, in fact, appear to persist in the combined theory of both quantum mechanics and gravity, interchanging the left-right slopes in Fig. 3. What for us is the statement, "the space is as full as Einstein's theory will allow" (right slope of Fig. 3) would read equivalently in the dual theory as "the space is as empty as Heisenberg's uncertainty principle will allow" (left slope of Fig. 3), and vice versa. Details will be developed elsewhere. The remark may be relevant to understanding the principle within a larger Kantian thesis.

**Shahn Majid**
Lyman Laboratory of Physics
Harvard University
Cambridge, Massachusetts 02138 U.S.A.

Present address:
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Cambridge CB3 9EW United Kingdom

## References

1. R. Streater and A. Wightman, *PCT, Spin, Statistics and All That* (Benjamin, Reading, 1964).
2. S. Majid, Ph.D. thesis, Harvard University, 1988.
3. *Idem*, J. Algebra **130**, 17 (1990).
4. *Idem*, Pac. J. Math. **141**, 311 (1990).
5. *Idem*, J. Funct. Analysis **95**, 291 (1991).
6. *Idem*, J. Classical Quantum Gravity **5**, 1587 (1988).
7. *Idem*, Lect. Notes Phys. **375**, 131 (Springer, 1991).
8. *Idem*, Supl. Rend. Circ. Mat. Palermo, to be published.
9. S. Maclane, *Categories for the Working Mathematician* (Springer, 1974), GTM Vol. 5.
10. H. Allison, *Kant's Trancendental Idealism* (Yale University, 1983).
11. Plato, *Plato's Republic* (Hackett, 1974), translation of G.M.A. Grube.
12. A.A. Kirillov, *Elements of the Theory of Representations* (Springer-Verlag, Heidelberg, 1976).
13. I. Segal, *Mathematical Problems of Relativistic Physics* (AMS, 1963).
14. E. Mach, *The Science of Mechanics* (Open Court, 1893), translated by T.J. McCormack, 3rd paperback ed. (1974).
15. A. Einstein, *The Meaning of Relativity* (Chapman and Hall, London, 1922). Reprinted 1980.
16. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1963), Vols. I-III.
17. S. Hawking, *Lucasian inaugural address*, Cambridge University, 1982.
18. T.S. Kuhn, *The Structure of Scientific Revolutions*, 2nd ed., with postscript (Chicago, 1969).
19. M.E. Sweedler, *Hopf Algebras* (Benjamin, 1969).
20. S. Majid, Int. J. Mod. Phys. A **5**, 1 (1990).
21. G.'t Hooft, *Proc. XVIth GIFT Int. Seminar on Theoretical Physics, Spain* (World Scientific, 1986).
22. S. Doplicher and J.E. Roberts, Inv. Math **98**, 157 (1989).
23. P. Feyerabend, *Against Method* (London, 1977).
24. I. Lakatos, *Proofs and Refutations: The Logic of Mathematical Discovery* (Cambridge, 1976).